Retinoic acid signaling in development and evolution

Michael Schubert

▶ **To cite this version:**

Michael Schubert. Retinoic acid signaling in development and evolution. Embryology and Organogenesis. Ecole Normale Supérieure de Lyon, 2011. tel-03408901

HAL Id: tel-03408901
https://hal.science/tel-03408901

Submitted on 29 Oct 2021

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

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Ecole Normale Supérieure de Lyon

Habilitation à Diriger les Recherches

presented by

Michael Schubert

entitled

Retinoic Acid Signaling in Development and Evolution

defended on

November 25, 2011

Jury Members:	**Jean-Nicolas Volff**	**President**
	Guillaume Balavoine	**Examiner**
	Sylvie Mazan	**Examiner**
	Filippo Rijli	**Examiner**
	Vincent Laudet	**Scientific Advisor**

Retinoic Acid Signaling in Development and Evolution

Michael Schubert

Habiliation à Diriger des Recherches
Ecole Normale Supérieure de Lyon

November 25, 2011

Table of Contents

Acknowledgements

I would like to thank everybody that, in the course of the last fifteen years (i.e. since the beginning of my scientific career) has supported my research efforts either directly or indirectly. Moreover, I am also grateful to the members of the jury that have agreed to serve on this committee hence accepting all the repercussions this might have. Most importantly, I am particularly indebted to my family, which is regularly faced with almost insurmountable challenges due to my continued chase of scientific glory.

Summary

Extensive research carried out over the last 100 years has established that retinoids, which constitute a group of fat-soluble morphogens related to retinol (vitamin A), play crucial roles in development, organogenesis, tissue homeostasis, cell proliferation, differentiation and apoptosis. In vertebrates, retinoic acid (RA), the main mediator of retinoid signaling, is synthesized from retinol in two steps, the second being the irreversible oxidation of retinal to RA by RALDH enzymes. In addition to RA synthesis, the availability of endogenous RA is also controlled through its degradation by CYP26 enzymes. Most RA functions in vertebrates are mediated by binding of RA to heterodimers of two nuclear receptors: the retinoic acid receptor (RAR) and the retinoid X receptor (RXR). Retinoid signaling was long thought to be vertebrate-specific, but developmental studies in invertebrate chordates have revealed roles for retinoids that are conserved in all chordates. These studies have also established that, of all invertebrates, the cephalochordate amphioxus has the most vertebrate-like retinoid signaling system, both in terms of molecular composition and of biological functions, with the important added value of a lack of significant genetic redundancy. Outside the chordate lineage, evidence for functional roles of retinoids and of the RAR/RXR heterodimer is scarce. Although genes encoding orthologs of vertebrate RALDH and CYP26 as well as of vertebrate RAR and RXR are present in ambulacrarian deuterostomes (such as hemichordates and echinoderms) and lophotrochozoan protostomes (such as annelids and mollusks), extensive experimental evidence for retinoid and/or RAR/RXR functions in these lineages is still lacking. I am thus proposing to study retinoid signaling in five invertebrate taxa, whose genomes encode orthologs of RALDH, CYP26, RAR and RXR (the annelid *Platynereis dumerilii*, the mollusk *Nucella lapillus*, the echinoderm *Paracentrotus lividus*, the hemichordate *Saccoglossus kowalevskii*, and the cephalochordate *Branchiostoma lanceolatum*). Experiments have been designed to independently characterize, in these five taxa, the roles of retinoids and of RAR and RXR, which will allow a detailed assessment of the functions of retinoids versus the functions of RAR and RXR in each lineage. Comparisons of these respective functions will reveal roles for retinoids that are dependent on RAR and RXR and those that are independent of the receptors. Moreover, cross-comparisons of these results between the five species will identify conserved and divergent elements of retinoid signaling, which in turn will allow the reconstruction of the evolutionary diversification of this morphogen-dependent signaling cascade in bilaterian animals. Altogether, these data, obtained in animals without obvious ties to humans, will instruct the complexity of the vertebrate (including, of course, also the human) retinoid system, with its crucial functions in development and tissue homeostasis.

Curriculum Vitae

Personal Information

Last Name:	**Schubert**
First Name:	**Michael**
Middle Name:	**Stefan**
Academic Title:	Doctor of Philosophy (Ph.D.)
Place and Date of Birth:	Stuttgart (Germany), 20/02/1974
Nationality:	German
Marital Status:	Married, Two Children
Current Position:	Research Scientist – Tenured (CR1 – Associate Research Professor level)
Employer:	Centre National de la Recherche Scientifique (CNRS), France
Work Address:	Institut de Génomique Fonctionnelle de Lyon (UCBL, CNRS UMR5242, ENS, INRA 1288) Ecole Normale Supérieure de Lyon 46, allée d'Italie 69364 Lyon Cedex 07 France
Work Telephone Number:	++33 (0) 4 72 72 86 85
Work Fax Number:	++33 (0) 4 72 72 86 74
Work E-mail Address:	Michael.Schubert@ens-lyon.fr
Private Address:	260, rue Paul Bert 69003 Lyon France
Private Telephone Number:	++33 (0) 4 82 31 63 88
Mobile Telephone Number:	++33 (0) 6 12 52 33 19
Private E-mail Address:	mschubert1974@hotmail.com

High School Education

Institution:	**Jörg Ratgeb – Schule Stuttgart-Neugereut, Stuttgart, Germany**
Major:	Mathematics/Chemistry
Graduation Date:	12/05/1993
Degree:	Abitur
Result:	Very good

Higher Education

First Degree

Institution:	**University of Stuttgart, Stuttgart, Germany**
Major:	Technical Biology
Graduation Date:	12/12/1996
Degree:	Vordiplom
Result:	Very good

Advanced Degree

Institution:	**Scripps Institution of Oceanography, University of California San Diego, La Jolla, USA**
Major:	Marine Biology
Graduation Date:	08/09/2001
Degree:	Doctor of Philosophy (Ph.D.)
Result:	N/A
Ph. D. Supervisor:	Dr. Nicholas D. Holland and Dr. Linda Z. Holland
Thesis Title:	*Wnt* family genes in the cephalochordate amphioxus: structure, phylogenetic analysis, and developmental expression

Postdoctoral Experience

Institution:	**Laboratoire de Biologie Moléculaire de la Cellule, Ecole Normale Supérieure de Lyon, Lyon, France**
Principal Investigator:	Vincent Laudet
Research Topic:	Evolution of metamorphosis and endocrine functions with special focus on diversification of the nuclear hormone receptor superfamily within the chordate lineage
Time Period:	01/11/2001 to 03/11/2004

Military/Social Service

Employer:	**AWO Ludwigsburg – Mobiler Sozialer Dienst, Kornwestheim, Germany**
Position:	Community service employee
Responsibilities, Tasks:	Support of elderly people at their homes and in a charitable home
Time Period:	01/07/1993 to 30/09/1994

Responsibility for Students and Research Scientists

Licence 3 student (June 2009 to July 2009): Laetitia Jimenez
Master 1 student (February 2005 to June 2005): Claire Lecroisey
Master 2 student (September 2005 to July 2006): Ferdinand Marlétaz
Diploma student (November 2008 to October 2009): Jasmin Schulz
Master Student (September 2011 to present): João Emanuel Marques Carvalho
Ph.D. exchange student (June 2011 to September 2011): Daniela da Silva Lima
Ph.D. student (January 2007 to May 2011): Jie Chen
Master 2 and Ph.D. student (March 2007 to present): Florent Campo Paysaa
Master 2 and Ph.D. student (September 2009 to present): Juliana Gutierrez Mazariegos
Postdoctoral researcher (November 2006 to April 2011): Maria Theodosiou

Teaching Experience

Teaching Assistantship, University of California San Diego, La Jolla, USA (Fall Quarter 2000), Course: “Introduction to Oceanography”
Reviewer for the Master Program “Biosciences” at the University of Lyon, Lyon, France (Academic Years 2008/2009 and 2009/2010)
Thesis defense committee member for the Master Program “Biosciences” at the University of Lyon, Lyon, France (Academic Year 2010/2011)

Fellowships and Stipends

Fulbright Fellowship for graduate studies at the Scripps Institution of Oceanography, University of California San Diego, La Jolla, USA (1997 and 1998)
United States National Science Foundation (NSF) Research Assistantship for graduate research in the Holland laboratory at the Scripps Institution of Oceanography, University of California San Diego, La Jolla, USA (1998 to 2001)
German Academic Exchange Service (DAAD) Postdoctoral Fellowship for work in the Vincent Laudet laboratory at the Ecole Normale Supérieure de Lyon, Lyon, France (2001 and 2002)
European Commission Marie Curie Individual Fellowship for work in the Vincent Laudet laboratory at the Ecole Normale Supérieure de Lyon, Lyon, France (2002 to 2004)

Research Funding and Grants

ACI-BCMS Grant “Evolution des Voies de Signalisation des Rétinoids chez les Chordés” (France), 2004 to 2007 (€ 90.000): Co-Investigator
Integrated Project “CRESCENDO” (Europe), 2006 to 2011 (€ 164.500): Co-Investigator
ANR Blanc Grant “RAnterior*Hox*” (France), 2007 to 2011 (€ 300.000): Principle Investigator
FACCTS Project “Evolution of Retinoic Acid Signaling in Deuterostomes” (USA-France), 2009 to 2010 (€ 5.000): Co-Investigator
ANR Blanc Grant “EvolAx” (France), 2009 to 2012 (€ 78.000): Co-Investigator
CNRS-PEPS Grant (France), 2010 (€ 15.000): Co-Investigator
FCT Grant (Portugal-France), 2011 (€ 12.150): Co-Investigator

ANR Jeune Chercheur Grant “RAvolution” (France), 2011 to 2014 (€ 250.000): Principle Investigator

Current Sub-Group Members

Master Student (September 2011 to present): João Emanuel Marques Carvalho
Master 2 and Ph.D. student (March 2007 to present): Florent Campo Paysaa
Master 2 and Ph.D. student (September 2009 to present): Juliana Gutierrez Mazariegos

Invited Talks and Oral Conference Presentations

Czech Academy of Sciences, Prague, Czech Republic (01/12/1999)
Swarthmore College, Swarthmore, Pennsylvania, USA (21/11/2000)
14th Scientific Meeting of the German Society of Developmental Biology (GfE), Ulm/Donau, Germany (23/03/2001)
Ecole Normale Supérieure de Lyon, Lyon, France (27/11/2001)
6th Evolutionary Biology Meeting, Marseille, France (20/06/2002)
7th Evolutionary Biology Meeting, Marseille, France (25/06/2003)
ELSO 2004 Conference, Nice, France (05/09/2004)
Colloque Annuel de la Société Française de Biologie du Développement, Batz-sur-Mer, France (04/10/2004)
Annual Meeting of the Société Française de Biologie du Développement, Obernai, France (03/10/2005)
Université Paris XI, Orsay, France (07/03/2006) (guest lecture in the course module "Evo-Devo")
The First and Founding Meeting of the European Society for Evolutionary Developmental Biology, Prague, Czech Republic (18/08/2006)
Ecole Normale Supérieure de Lyon, Lyon, France (25/09/2006) (Two guest lectures in the course module "Evolution et Développement")
MRC Centre for Developmental Neurobiology, Guy's Campus, King's College London, London, United Kingdom (02/11/2006)
9th Meeting of the UNESCO Chair of Developmental Biology, Rio de Janeiro, Brazil (01/12/2006)
1st Annual Meeting of the EU Integrated Project CRESCENDO, Paris, France (26/04/2007)
Unité de Biologie Moléculaire du Développement, Institut Pasteur, Paris, France (17/01/2008)
2nd Meeting of the European Society for Evolutionary Developmental Biology, Ghent, Belgium (31/07/2008)
Unité de Biologie du Développement, Observatoire Océanologique de Villefranche-sur-Mer, Villefranche-sur-Mer, France (03/04/2009)
Unité Mer et Santé, Station Biologique de Roscoff, Roscoff, France (08/06/2009)
Institut de Neurobiologie Alfred Fessard, Centre National de la Recherche Scientifique, Gif-sur-Yvette, France (29/01/2010)
Biology Department, Universität Konstanz, Konstanz, Germany (26/05/2010)
3rd Meeting of the European Society for Evolutionary Developmental Biology, Paris, France (07/07/2010)
Developmental Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany (25/10/2010)
Centre de Génétique et de Physiologie Moléculaire et Cellulaire, Université Claude Bernard Lyon 1, Lyon, France (16/09/2011)
15th Evolutionary Biology Meeting, Marseille, France (30/09/2011)

Other Experience and Professional Memberships

Member of the “Society for Developmental Biology” (1998 to 2004)
Member of the “Gesellschaft für Biochemie und Molekularbiologie” (2003 to 2004)
Member of the “Société Française de Biologie du Développement” (2003 to present)
Member of the “Gesellschaft für Entwicklungsbiologie” (2004 to present)
Member of the “European Society for Evolutionary Developmental Biology” (2006 to present)
Member of the Executive Committee and Treasurer of the “European Society for Evolutionary Developmental Biology” (2006 to present)
Member of the Scientific Committee and Organizer of the Biannual Meetings of the “European Society for Evolutionary Developmental Biology” (2006 to present)
Member of Section 24 (“Interactions Cellulaires”) of the “Comité National de la Recherche Scientifique” (2008 to present)
Member of the Habilitation Jury of Andrea Pasini (15/09/2011)
Grant reviewer for the “Agence Nationale de la Recherche” (2011 to present)
Member of the Editorial Board and Academic Editor of *PLoS ONE* (2011 to present)
Member of the Editorial Board of *The Open Evolution Journal* (2007 to present) and of *TheScientificWorldJournal* (2011 to present)
Reviewer for *Biochimica et Biophysica Acta - Gene Structure and Expression*, *BMC Evolutionary Biology*, *Current Biology*, *Evolution & Development*, *Gene*, *Genesis*, *Indian Journal of Marine Sciences*, *International Journal of Biological Sciences*, *Italian Journal of Zoology*, *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, *Journal of Molecular Biology*, *Journal of Molecular Endocrinology*, *Marine Biology*, *Marine Ecology*, *Molecular Biology and Evolution*, *Nature Communications*, *PLoS ONE*, *Proceedings of the Royal Society B: Biological Sciences*
Person in charge of the amphioxus animal facility at the Ecole Normale Supérieure de Lyon, Lyon, France (2008 to present)

Scientific Production (as of 01/10/2011)

42 research papers, 7 review articles, 1 book chapter
1146 citations
h-index: 21

List of Publications

Research Articles

1. Albalat, R.[*], Brunet, F.[*], Laudet, V., and **Schubert, M.**, 2011. Evolution of retinoid and steroid signaling: vertebrate diversification from an amphioxus perspective. **Genome Biol. Evol.** 3, 985-1005.
   [*]These authors contributed equally to this work.
2. Campo-Paysaa, F.[*], Semon, M.[*], Cameron, R. A., Peterson, K. J., and **Schubert, M.**, 2011. microRNA complements in deuterostomes: origin and evolution of microRNAs. **Evol. Dev.** 13, 15-27.
   [*]These authors contributed equally to this work.
3. Koop, D., Holland, L. Z., Setiamarga, D., **Schubert, M.**, and Holland, N. D., 2011. Tail regression induced by elevated retinoic acid signaling in amphioxus larvae occurs by tissue remodeling, not cell death. **Evol. Dev.** 13, 427-435.
4. Sobreira, T. J. P.[*], Marletaz, F.[*], Simoes-Costa, M.[*], Schechtman, D., Pereira, A. C., Brunet, F., Sweeney, S., Pani, A., Aronowicz, J., Lowe, C. J., Davidson, B., Laudet, V., Bronner-Fraser, M., de Oliveira, P. S. L., **Schubert, M.**[*], and Xavier-Neto, J.[*], 2011. Structural shifts of aldehyde dehydrogenase enzymes were instrumental for the early evolution of retinoid-dependent axial patterning in metazoans. **Proc. Natl. Acad. Sci. U.S.A.** 108, 226-231.
   [*]These authors contributed equally to this work.
5. Theodosiou, M., Colin, A., Schulz, J., Laudet, V., Peyrieras, N., Nicolas, J. F., **Schubert, M.**, and Hirsinger, E., 2011. Amphioxus spawning behavior in an artificial seawater facility. **J. Exp. Zool. B** 316, 263-275.
6. Koop, D., Holland, N. D., Semon, M., Alvarez, S., de Lera, A. R., Laudet, V., Holland, L. Z., and **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.
7. Kucera, T., Strilic, B., Regener, K., **Schubert, M.**, Laudet, V., and Lammert, E., 2009. Ancestral vascular lumen formation via basal cell surfaces. **PLoS ONE** 4, e4132.
8. 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.
9. Osborne, P. W., Benoit, G., Laudet, V., **Schubert, M.**[*], and Ferrier, D. E. K.[*], 2009. Differential regulation of ParaHox genes by retinoic acid in the invertebrate chordate amphioxus (*Branchiostoma floridae*). **Dev. Biol.** 327, 252-262.
   [*]These authors contributed equally to this work.
10. Zhong, J., Zhang, Q. J., Xu, Q. S., **Schubert, M.**, Laudet, V., and Wang, Y. Q., 2009. Complete mitochondrial genomes defining two distinct lancelet species in the West Pacific Ocean. **Mar. Biol. Res.** 5, 278-285.
11. Holland, L. Z., Albalat, R., Azumi, K., Benito-Gutierrez, E., Blow, M. J., Bronner-Fraser, M., Brunet, F., Butts, T., Candiani, S., Dishaw, L. J., Ferrier, D. E. K., Garcia-Fernandez, J., Gibson-Brown, J. J., Gissi, C., Godzik, A., Hallbook, F., Hirose, D., Hosomichi, K., Ikuta, T., Inoko, H., Kasahara, M., Kasamatsu, J., Kawashima, T., Kimura, A., Kobayashi, M., Kozmik, Z., Kubokawa, K., Laudet, V., Litman, G. W., McHardy, A. C., Meulemans, D., Nonaka, M.,

Olinski, R. P., Pancer, Z., Pennacchio, L. A., Pestarino, M., Rast, J. P., Rigoutsos, I., Robinson-Rechavi, M., Roch, G., Saiga, H., Sasakura, Y., Satake, M., Satou, Y., **Schubert, M.**, Sherwood, N., Shiina, T., Takatori, N., Tello, J., Vopalensky, P., Wada, S., Xu, A., Ye, Y., Yoshida, K., Yoshizaki, F., Yu, J. K., Zhang, Q., Zmasek, C. M., de Jong, P. J., Osoegawa, K., Putnam, N. H., Rokhsar, D. S., Satoh, N., and Holland, P. W. H., 2008. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. **Genome Res.** 18, 1100-1111.

12. Paris, M., Brunet, F., Markov, G. V., **Schubert, M.**, and Laudet, V., 2008. The amphioxus genome enlightens the evolution of the thyroid hormone signaling pathway. **Dev. Genes Evol.** 218, 667-680.
13. Paris, M., Escriva, H., **Schubert, M.**, Brunet, F., Brtko, J., Ciesielski, F., Roecklin, D., Vivat-Hannah, V., Jamin, E. L., Cravedi, J. P., Scanlan, T. S., Renaud, J. P., Holland, N. D., and Laudet, V., 2008. Amphioxus postembryonic development reveals the homology of chordate metamorphosis. **Curr. Biol.** 18, 825-830.
14. Paris, M., Pettersson, K., **Schubert, M.**, Bertrand, S., Pongratz, I., Escriva, H., and Laudet, V., 2008. An amphioxus orthologue of the estrogen receptor that does not bind estradiol: insights into estrogen receptor evolution. **BMC Evol. Biol.** 8, 219.
15. **Schubert, M.**[*], Brunet, F.[*], Paris, M., Bertrand, S., Benoit, G., and Laudet, V., 2008. Nuclear hormone receptor signaling in amphioxus. **Dev. Genes Evol.** 218, 651-665.
    [*]These authors contributed equally to this work.
16. Beaster-Jones, L., **Schubert, M.**, and Holland, L. Z., 2007. *Cis*-regulation of the amphioxus engrailed gene: insights into evolution of a muscle-specific enhancer. **Mech. Dev.** 124, 532-542.
17. Fuentes, M., Benito, E., Bertrand, S., Paris, M., Mignardot, A., Godoy, L., Jimenez Delgado, S., Oliveri, D., Candiani, S., Hirsinger, E., D'Aniello, S., Pascual-Anaya, J., Maeso, I., Pestarino, M., Vernier, P., Nicolas, J. F., **Schubert, M.**, Laudet, V., Geneviere, A. M., Albalat, R., Garcia Fernandez, J., Holland, N. D., and Escriva, H., 2007. Insights into spawning behavior and development of the European amphioxus (*Branchiostoma lanceolatum*). **J. Exp. Zool. B** 308, 484-493.
18. Kozmik, Z., Holland, N. D., Kreslova, J., Oliveri, D., **Schubert, M.**, Jonasova, K., Holland, L. Z., Pestarino, M., Benes, V., and Candiani, S., 2007. *Pax-Six-Eya-Dach* network during amphioxus development: conservation *in vitro* but context-specificity *in vivo*. **Dev. Biol.** 306, 143-159.
19. Rasmussen, S. L. K., Holland, L. Z., **Schubert, M.**, Beaster-Jones, L., and Holland, N. D., 2007. Amphioxus *AmphiDelta*: evolution of Delta protein structure, segmentation, and neurogenesis. **Genesis** 45, 113-122.
20. Glavin, D. P., Cleaves, H. J., Buch, A., **Schubert, M.**, Aubrey, A., Bada, J. L., and Mahaffy, P. R., 2006. Sublimation extraction coupled with gas chromatography-mass spectrometry: a new technique for future in situ analyses of purines and pyrimidines on Mars. **Planet. Space Sci.** 54, 1584-1591.
21. **Schubert, M.**, Holland, N. D., Laudet, V., and Holland, L. Z., 2006. A retinoic acid-*Hox* hierarchy controls both anterior/posterior patterning and neuronal specification in the developing central nervous system of the cephalochordate amphioxus. **Dev. Biol.** 296, 190-202.

22. Bardet, P. L., **Schubert, M.**, Horard, B., Holland, L. Z., Laudet, V., Holland, N. D., and Vanacker, J. M., 2005. Expression of estrogen-receptor related receptors in amphioxus and zebrafish: implications for the evolution of posterior brain segmentation at the invertebrate-to-vertebrate transition. **Evol. Dev.** 7, 223-233.
23. Holland, L. Z., Panfilio, K. A., Chastain, R., **Schubert, M.**, and Holland, N. D., 2005. Nuclear β-catenin promotes non-neural ectoderm and posterior cell fates in amphioxus embryos. **Dev. Dyn.** 233, 1430-1443.
24. **Schubert, M.**, Yu, J. K., Holland, N. D., Escriva, H., Laudet, V., and Holland, L. Z., 2005. Retinoic acid signaling acts via Hox1 to establish the posterior limit of the pharynx in the chordate amphioxus. **Development** 132, 61-73.
25. Fuentes, M., **Schubert, M.**, Dalfo, D., Candiani, S., Benito, E., Gardenyes, J., Godoy, L., Moret, F., Illas, M., Patten, I., Permanyer, J., Oliveri, D., Boeuf, G., Falcon, J., Pestarino, M., Garcia Fernandez, J., Albalat, R., Laudet, V., Vernier, P., and Escriva, H., 2004. Preliminary observations on the spawning conditions of the European amphioxus (*Branchiostoma lanceolatum*) in captivity. **J. Exp. Zool. B** 302, 384-391.
26. Glavin, D. P., Cleaves, H. J., **Schubert, M.**, Aubrey, A., and Bada, J. L., 2004. New method for estimating bacterial cell abundances in natural samples by use of sublimation. **Appl. Environ. Microbiol.** 70, 5923-5928.
27. **Schubert, M.**, Holland, N. D., Escriva, H., Holland, L. Z., and Laudet, V., 2004. Retinoic acid influences anteroposterior positioning of epidermal sensory neurons and their gene expression in a developing chordate (amphioxus). **Proc. Natl. Acad. Sci. U.S.A.** 101, 10320-10325.
28. **Schubert, M.**, Meulemans, D., Bronner-Fraser, M., Holland. L. Z., and Holland, N. D., 2003. Differential mesodermal expression of two amphioxus *MyoD* family members (*AmphiMRF1* and *AmphiMRF2*). **Gene Expr. Patterns** 3, 199-202.
29. Glavin, D. P., **Schubert, M.**, and Bada, J. L., 2002. Isolation of purines and pyrimidines from nucleic acids using sublimation. **Anal. Chem.** 74, 6408-6412.
30. Kreslova, J., Holland, L. Z., **Schubert, M.**, Burgtorf, C., Benes, V., and Kozmik, Z., 2002. Functional equivalency of amphioxus and vertebrate Pax258 transcription factors suggests that the activation of mid-hindbrain specific genes in vertebrates occurs via the recruitment of Pax regulatory elements. **Gene** 282, 143-150.
31. Glavin, D. P., **Schubert, M.**, Botta, O., Kminek, G., and Bada, J. L., 2001. Detecting pyrolysis products from bacteria on Mars. **Earth Planet. Sci. Lett.** 185, 1-5.
32. Kozmik, Z., Holland, L. Z., **Schubert, M.**, Lacalli, T. C., Kreslova, J., Vlcek, C., and Holland, N. D., 2001. Characterization of amphioxus *AmphiVent*, an evolutionarily conserved marker for chordate ventral mesoderm. **Genesis** 29, 172-179.
33. **Schubert, M.**, Holland, L. Z., Stokes, M. D., and Holland, N. D., 2001. Three amphioxus *Wnt* genes (*AmphiWnt3*, *AmphiWnt5*, and *AmphiWnt6*) associated with the tail bud: the evolution of somitogenesis in chordates. **Dev. Biol.** 240, 262-273.
34. Holland, L. Z., Holland, N. D., and **Schubert, M.**, 2000. Developmental expression of *AmphiWnt1*, an amphioxus gene in the *Wnt1*/*wingless* subfamily. **Dev. Genes Evol.** 210, 522-524.

35. Holland, L. Z., **Schubert, M.**, Holland, N. D., and Neuman, T., 2000. Evolutionary conservation of the presumptive neural plate markers *AmphiSox1/2/3* and *AmphiNeurogenin* in the invertebrate chordate amphioxus. **Dev. Biol.** 226, 18-33.
36. **Schubert, M.**, Holland, L. Z., and Holland, N. D., 2000. Characterization of two amphioxus *Wnt* genes (*AmphiWnt4* and *AmphiWnt7b*) with early expression in the developing central nervous system. **Dev. Dyn.** 217, 205-215.
37. **Schubert, M.**, Holland, L. Z., and Holland, N. D., 2000. Characterization of an amphioxus *Wnt* gene, *AmphiWnt11*, with possible roles in myogenesis and tail outgrowth. **Genesis** 27, 1-5.
38. **Schubert, M.**, Holland, L. Z., Holland, N. D., and Jacobs, D. K., 2000. A phylogenetic tree of the *Wnt* genes based on all available full-length sequences, including five from the cephalochordate amphioxus. **Mol. Biol. Evol.** 17, 1896-1903.
39. **Schubert, M.**, Holland, L. Z., Panopoulou, G. D., Lehrach, H., and Holland, N. D., 2000. Characterization of amphioxus *AmphiWnt8*: insights into the evolution of patterning of the embryonic dorsoventral axis. **Evol. Dev.** 2, 85-92.
40. Holland, L. Z., **Schubert, M.**, Kozmik, Z., and Holland, N. D., 1999. *AmphiPax3/7*, an amphioxus paired box gene: insights into chordate myogenesis, neurogenesis, and the possible evolutionary precursor of definitive vertebrate neural crest. **Evol. Dev.** 1, 153-165.
41. Kozmik, Z., Holland, N. D., Kalousova, A., Paces, J., **Schubert, M.**, and Holland, L. Z., 1999. Characterization of an amphioxus paired box gene, *AmphiPax2/5/8*: developmental expression patterns in optic support cells, nephridium, thyroid-like structures and pharyngeal gill slits, but not in the midbrain-hindbrain boundary region. **Development** 126, 1295-1304.
42. **Schubert, M.**, Holland, N. D., and Holland, L. Z., 1998. Amphioxus *AmphiDRAL* encoding a LIM-domain protein: expression in the epidermis but not in the presumptive neuroectoderm. **Mech. Dev.** 76, 203-205.

Review Articles and Book Chapters

1. Gutierrez-Mazariegos, J.*, Theodosiou, M.*, Campo-Paysaa, F.*, and **Schubert, M.**, 2011. Vitamin A: a multifunctional tool for development. **Semin. Cell Dev. Biol.**, in press.
   *These authors contributed equally to this work.
2. Theodosiou, M., Laudet, V., and **Schubert, M.**, 2010. From carrot to clinic: an overview of the retinoic acid signaling pathway. **Cell. Mol. Life Sci.** 67, 1423-1445.
3. Campo-Paysaa, F., Marletaz, F., Laudet, V., and **Schubert, M.**, 2008. Retinoic acid signaling in development: tissue-specific functions and evolutionary origins. **Genesis** 46, 640-656.
4. Marletaz, 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.
5. **Schubert, M.**, Escriva, H., Xavier-Neto, J., and Laudet, V., 2006. Amphioxus and tunicates as evolutionary model systems. **Trends Ecol. Evol.** 21, 269-277.
6. Holland, L. Z., Laudet, V., and **Schubert, M.**, 2004. The chordate amphioxus: an emerging model organism for developmental biology. **Cell. Mol. Life Sci.** 61, 2290-2308.

7. **Schubert, M.**, Escriva, H., and Laudet, V., 2004. La pensée à fleur de peau (Thinking with a sensitive skin). **Méd. Sci. (Paris)** 20, 273-274.
8. **Schubert, M.**, and Holland, L. Z., 2003. The *Wnt* gene family and the evolutionary conservation of *Wnt* expression. **Wnt Signaling in Development** (M. Kühl, Editor), 210-239. Landes Bioscience, Georgetown, Texas, USA.

Selected Conference Abstracts

1. Gutierrez-Mazariegos, J., Laudet, V., and **Schubert, M.**, 2011. Origin and diversification of the retinoic acid signaling pathway. **Programme and Abstract Book of the EMBO Conference on Nuclear Receptors in Sitges, Barcelona (Spain)**, 126.
2. **Schubert, M.**, 2011. Structural shifts of aldehyde dehydrogenase enzymes were instrumental for the early evolution of retinoid-dependent axial patterning in metazoans. **Abstracts of the 15th Evolutionary Biology Meeting at Marseille (France)**, 107.
3. **Schubert, M.**, 2010. Structural shifts of ALDH enzymes were instrumental for the early evolution of retinoid-dependent axial patterning in metazoans. **Abstracts of the 3rd Meeting of the European Society for Evolutionary Developmental Biology in Paris (France)**, 128.
4. Campo-Paysaa, F., Semon, M., Peterson, K. J., and **Schubert, M.**, 2009. Complement and genomic distribution of miRNAs in deuterostomes: conservation and diversification of the miRNA regulatory network. **Mech. Dev.** 126, S55.
5. **Schubert, M.**, Simoes-Costa, M., Marletaz, F., Sobreira, T. J. P., de Oliveira, P. S. L., Schechtman, D., Brunet, F., Albalat, R., Bronner-Fraser, M., Laudet, V., and Xavier-Neto, J., 2009. Evolutionary shifts in ALDH structure suggest transitions between pleiotropic and patterning functions. **Mech. Dev.** 126, S250.
6. **Schubert, M.**, Campo-Paysaa, F., Koop, D., Marletaz, F., Xavier-Neto, J., Holland, L. Z., and Laudet, V., 2008. The evolution of retinoic acid signaling in development. **Abstracts of the 2nd Meeting of the European Society for Evolutionary Developmental Biology in Ghent (Belgium)**, 180-181.
7. Koop, D., **Schubert, M.**, and Holland, L. Z., 2007. Retinoic acid and *Hox* genes in the patterning of amphioxus. **Dev. Biol.** 306, 289.
8. **Schubert, M.**, Marletaz, F., Holland, N. D., Holland, L. Z., and Laudet, V., 2006. Retinoic acid, *Hox* and the evolution of ectodermal patterning in chordates. **Abstracts of the First and Founding Meeting of the European Society for Evolutionary Developmental Biology in Prague (Czech Republic)**, 201.
9. **Schubert, M.**, Holland, N. D., Holland, L. Z., and Laudet, V., 2005. RA-signaling and *Hox* genes control both anteroposterior patterning and neuronal specification in the central nervous system of a simple chordate model (amphioxus). **Abstracts of the Annual Meeting of the Société Française de Biologie du Développement in Obernai (France)**, 51.
10. Holland, L. Z., **Schubert, M.**, Laudet, V., and Holland, N. D., 2004. Retinoic acid, *Hox* genes, and rostrocaudal patterning of amphioxus embryos. **Dev. Biol.** 271, 559.

11. **Schubert, M.**, Holland, N. D., Escriva, H., Holland, L. Z., and Laudet, V., 2004. Retinoic acid influences anteroposterior positioning of epidermal sensory neurons and their gene expression in a developing chordate (amphioxus). **ELSO 2004 Conference Proceedings**, 16.
12. **Schubert, M.**, Escriva, H., Holland, N. D., Holland, L. Z., and Laudet, V., 2003. The effects of exogenous retinoic acid (RA) and retinoic acid antagonist (BMS009) on developmental gene expression of amphioxus: insights into the evolution of chordate anteroposterior patterning. **Abstracts of the 7th Evolutionary Biology Meeting at Marseille (France)**, 13.
13. **Schubert, M.**, Escriva, H., and Laudet V., 2002. The role of thyroid hormones in the metamorphosis of the cephalochordate amphioxus. **Programme and Abstract Book of the BSCB/BSDB/Genetics Society Joint Spring Meeting in York (England)**, 76.
14. **Schubert, M.**, Meulemans, D. K., Holland, N. D., and Holland, L. Z., 2001. An amphioxus *engrailed cis*-regulatory region directs expression to zebrafish head mesenchyme: an example for evolutionary conservation of regulation within chordates. **Abstracts of the 14th Scientific Meeting of the Society of Developmental Biology (GfE) in Ulm/Donau (Germany)**, 113.
15. **Schubert, M.**, Holland, L. Z., Jacobs, D. K., and Holland, N. D., 2000. Phylogenetic analysis and expression of amphioxus *Wnt* genes: a possible ancient function for *Wnt1* during gastrulation. **Dev. Biol.** 222, 238.

Introduction

Vitamin A: a Multifunctional Tool for Development

Juliana Gutierrez-Mazariegos[†], Maria Theodosiou[†], Florent Campo-Paysaa[†], Michael Schubert
[†]These authors contributed equally to this work.

Abstract

Extensive research carried out over the last 100 years has established that the fat-soluble organic compound vitamin A plays crucial roles in early development, organogenesis, cell proliferation, differentiation and apoptosis as well as in tissue homeostasis. Given its importance during development, the delivery of vitamin A to the embryo is very tightly regulated with perturbations leading to severe malformations. This review discusses the roles of vitamin A during human development and the molecular mechanisms controlling its biological effects, hence bridging the gap between human development and molecular genetic work carried out in animal models. Vitamin A delivery during pregnancy and its developmental teratology in humans are thus discussed alongside work on model organisms, such as chicken or mice, revealing the molecular layout and functions of vitamin A metabolism and signaling. We conclude that, during development, vitamin A-derived signals are very tightly controlled in time and space and that this complex regulation is achieved by elaborate autoregulatory loops and by sophisticated interactions with other signaling cascades.

Introduction

Vitamin A is responsible for the formation and maintenance of many body tissues as well as for the promotion of healthy vision and immune functions. Vitamin A is obtained from the diet either directly, with the richest source being animal liver, or in the form of retinyl esters and carotenoids. For example, β-carotene is converted to vitamin A in the body through two successive oxidation steps. While absorption of preformed vitamin A in the diet is very efficient regardless of nutritional state, absorption of β-carotene is not as efficient, since, in the body, 12μg β-carotene are converted to only 1μg vitamin A [1].

The powerful effects of vitamin A on the promotion of healthy vision were already known in ancient Egypt, where night blindness was cured by ingestion of liver. Moreover, even before the pioneering experiments of Hale in the 1930s demonstrating the teratogenic potential of vitamin A deficiency (VAD), it was established that absence of vitamin A from the diet results in xerophthalmia. In 1933, Hale showed that a VAD sow gave birth to piglets without eyeballs, whereas deficient sows fed with cod-liver oil had normal offspring thus demonstrating that the observed phenotype was diet related [2]. Other reported birth defects were microphthalmia, accessory ears, cleft lip and palate as well as misplaced kidneys [3]. In 1950, Wilson and Warkany reported malformations of the eye, urogenital tract, heart and lung in the offspring of rats fed with VAD diets prior to and during gestation [4].

Soon thereafter, Cochlan was the first to show that excess vitamin A during pregnancy is teratogenic, inducing skeletal, craniofacial and central nervous system (CNS) defects [5, 6]. Thus, excess dietary vitamin A is also toxic to the organism. In animal experiments, hypervitaminosis A was shown to result in congenital abnormalities collectively termed retinoic acid embryopathy (RAE), which consists of malformations in the CNS (such as hydrocephaly, anencephaly or spina bifida), microtia/anotia, micrognathia, cleft palate, cardiac defects, thymic abnormalities and eye malformations [7, 8]. In addition, at lesser doses, neurogenesis can be affected resulting in learning disabilities [8].

Since these pioneering experiments, the importance of vitamin A for proper animal development has been firmly established in vertebrates as well as in certain invertebrate species. One of the most intriguing aspects of vitamin A-dependent signaling is the intricate nature of how this pathway is employed repeatedly during development in various tissues. This article is reviewing the impact of vitamin A and of its derivatives on human development and discusses these health-related issues in the context of our current understanding of the complex regulation of this multifaceted signaling cascade.

Vitamin A and Pregnancy

Maternal vitamin A status is important for implantation and later normal development of the fetus and neonate. Human studies have established that low or excess dietary levels of vitamin A during gestation result in teratogenesis [9]. During pregnancy, the nutritional requirement for vitamin A is increased, especially in the third trimester when fetal growth is most rapid. Vitamin A is transferred across the placenta to the embryo, even at maternal deficiency states. The needs of the embryo are met first, as suggested in a 1962 study where serum vitamin A levels were normal in cord blood but deficient in the mother [10]. The circulating levels of vitamin

A in neonates (1μmol/l) are lower than in maternal serum and levels below 0.7μmol/l in neonates are indicative of deficiency [11].

Delivery of vitamin A to the fetus is tightly regulated, which limits body stores at birth. Accumulation of vitamin A stores begins during the third trimester, and several months of sufficient intake after birth are required to build hepatic stores. Moreover, vitamin A is important for lung maturation *in utero*, and because of this premature infants are at high risk for bronchopulmonary dysplasia and require vitamin A supplementation [12]. Breast milk is the only source of vitamin A during the neonatal period for the exclusively breast-fed infant. Therefore, the vitamin A content of breast milk (which is dependent on the vitamin A status and serum levels of the mother during the last trimester of pregnancy) defines the levels of this vitamin in neonates [13]. The vitamin A content of colostrum and early milk is extremely high and neonate needs are met even with the milk of a mildly undernourished woman [12]. The vitamin A content of breast milk declines over 4-8 weeks of parturition [12]. In the case of VAD mothers, the needs of the infant cannot be met long term and supplementation is required to avoid detrimental health effects. Surprisingly, mothers of twins or short-interval births are also at risk for VAD, even in countries where vitamin A supplementation is not an issue [14].

Deficiency of vitamin A has been classified as a moderate health problem in developing countries. According to the World Health Organization (WHO), VAD is the leading cause of childhood blindness, and subclinical VAD elevates mortality risk from common childhood infections [12]. Vitamin A supplementation guides and programs have thus been put into place by the WHO: women should receive a dose of 200,000 International Units (IU) at parturition and prior to 8 weeks postpartum, infants aged 0-5 months 3 doses of 15,000 IU at least one month apart, infants aged 6-11 months a single dose of 100,000 IU, and infants of 12-59 months a dose of 200,000 IU every 6 months [15]. These doses of vitamin A are generally well tolerated, with few side effects reported. Moreover, food supplementation (for example of sugar, margarine, flour or rice) and low dose supplements over a longer period of time represent promising alternatives to the high dose administration of vitamin A, as they reduce the risk of formation of significant amounts of vitamin A-derived metabolites in the mother that could be passed on through the milk to the infant [15].

Developmental Teratology of Vitamin A and its Derivatives

Analogs of vitamin A, which is also called retinol, are generally referred to as retinoids. Other bioactive retinoids include, for example, different isomers of vitamin A acid, more commonly known as retinoic acid (RA), which is the most potent active metabolite of vitamin A (Fig. 1A). Since both retinoid deficiency and excess are harmful for the developing embryo, the endogenous retinoid supply during development must be very tightly regulated. Given the teratogenic potential of retinoids, the influence of various therapeutically employed retinoids on this fine regulatory balance has extensively been studied.

Vitamin A

Medical applications of vitamin A include the treatment of skin and eye disorders and the prevention of VAD in geographical regions, where VAD is considered a public health problem. The presence of adequate vitamin A stores is of critical importance during gestation and lactation and pregnant women should be

receiving 2,700-8,000 IU/day of vitamin A [16]. Toxicity due to vitamin A intake can occur when large doses (30,000 IU/day or higher) are ingested for prolonged periods of time. The Teratology Society considers doses of 8,000 IU/day as safe during pregnancy and doses over 25,000 IU/day as teratogenic [16].

Due to the tendency of vitamin A to bioaccumulate, consumption of large doses in the months before conception may lead to increased teratogenic risk. Moreover, high dietary intake of vitamin A before the 7th gestational week results in malformations attributable to vitamin A toxicity [17]. Vitamin A-dependent abnormalities reported in infants are reminiscent of those observed in animal models, with ear, limb and craniofacial malformations being the most abundant [8, 18]. Recently, a number of studies were published that examined the effects of vitamin A intake generally considered safe and nonteratogenic [19]. Not surprisingly, these analyses concluded that, if pregnant women are exposed to these innocuous vitamin A doses, teratogenicity appeared to be very unlikely [19].

Isotretinoin/Tretinoin

Whereas, isotretinoin (13-*cis* RA) is used in the treatment of cystic acne, tretinoin (all-*trans* RA) is utilized topically for treatment of skin disorders and orally for acute promyelocytic leukemia. Oral administration of both isotretinoin and tretinoin is teratogenic in every animal studied, including rodents, rabbits, pigs and primates [19]. The induced defects following isotretinoin and tretinoin treatment are RAE-like and similar to those induced by other retinoids [19]. In addition, topical treatment of rabbits with tretinoin results in fetal growth retardation and death [20].

Isotretinoin is a potent human teratogen [21] and is the first known teratogen [22, 23] to be approved as a drug (under the commercial name Accutane®). Shortly after isotretinoin was approved for medical use, there were reports of malformations due to inadvertent exposure during pregnancy [21, 24-26]. The most common malformations observed in this isotretinoin-induced RAE include hydrocephalus, microtia or anotia, maldevelopment of the facial bones and a flat, depressed nasal bridge [27].

Another common outcome of intrauterine exposure to isotretinoin is a lower IQ [28], which could be associated with, or caused by, the CNS malformations associated with RAE [19, 29]. A high spontaneous abortion rate is also associated with exposure to isotretinoin with the majority occurring 2-4 months after conception making death the most likely outcome of isotretinoin teratogenicity [25]. An estimated 700-1000 women had spontaneous abortions in the initial marketing period of 1982-1986 [19]. Another 5000-7000 women terminated their pregnancies in the same period for fear of congenital defects associated with isotretinoin exposure [19].

The critical period of exposure is the first trimester, more specifically the first 3-5 weeks following conception, well before organogenesis has been completed. However, due to its relatively short half-life (16-20h), isotretinoin is undetectable in the serum after 4-5 days and is therefore not believed to be of concern for conception that occurs after termination of maternal use [30, 31]. Interestingly, contrary to results obtained with animal models, observed effects in humans do not appear to be dose-dependent [27].

In contrast, the teratogenic potential of topically administered tretinoin has been questioned, because of a marked lack of significant percutaneous tretinoin absorption [32, 33]. Although cases of congenital malformations consistent with RAE following topical tretinoin treatment during pregnancy have been reported [34-37], several epidemiologic and experimental animal studies contradicted these findings

[38-43]. It should however be added that the epidemiologic and experimental animal analyses have not addressed subtle effects on brain development, which can occur even in the absence of typical RAE-associated malformations [28]. Taken together, the teratogenic potential of topically administered tretinoin cannot be excluded and exposure during pregnancy should thus be avoided [44].

Etretinate/Acitretin

Etretinate and its active metabolite, acitretin, are used orally for the treatment of psoriasis. Etretinate and acitretin induce malformations during organogenesis in all species studied, including rodents and rabbits. While for etretinate these include, among others, meningomyelocele, facial dysmorphia as well as skeletal and cardiovascular defects, acitretin treatments result in malformations of the limb and cleft palate [19].

In humans, a critical element of etretinate teratogenicity is exposure during the first 10 weeks of pregnancy. Developmental toxicity of both etretinate and acitretin usually results in spontaneous abortion or stillbirth with only some of the observed developmental malformations being indicative of RAE [19]. Interestingly, normal births following both etretinate and acitretin exposure have also been reported [45, 46].

Given that etretinate is very slowly released over a prolonged period of time, a number of etretinate toxicity cases have occurred 4 months to 4 years after the treatment ceased. Despite the half-life of acitretin being much shorter than that of etretinate (2-4 days for acitretin as opposed to 120 or more days for etretinate), acitretin can be converted to etretinate in the body [47], for example as a result of concurrent use of alcohol [48]. Thus, pregnancies are contraindicated for at least 3 years after discontinuation of either etretinate or acitretin treatment.

Vitamin A and its Derivatives have Pleiotropic Functions during Development

Apart from clinical investigations carried out on patients, many functions of vitamin A and its derivatives in humans have been extrapolated from studies in animal models. Initial pharmacological investigations into developmental processes controlled by vitamin A have shown that most biological functions controlled by this compound are actually mediated by RA [49]. These teratology-based experiments have subsequently been complemented with gene targeting analyses in classical vertebrate model species, such as the mouse, as well as with work on non-model organisms, such as invertebrate chordates [50-52]. Taken together, these studies strongly suggest that vitamin A and its biologically active derivatives (i.e. RA) have pleiotropic roles during development, which are very tightly regulated, both in time and space.

RA-dependent functions have actually been described for a wide range of developmental stages. Thus, in both vertebrates and invertebrate chordates, RA has been shown to regulate biological processes from early embryogenesis [53] to very late events in organ differentiation [51, 52]. For example, RA is implicated in controlling the early establishment of anteroposterior (AP) polarity in the gastrulating chordate embryo [54-56] and is also required at much later stages for organogenesis, for example of pancreas, lung, eyes, ears and digits [52, 57-59]. One outstanding example for time-dependent pleiotropic roles of RA during development is the differentiation of the heart, where RA is sequentially involved in cardiac field

specification [60], AP regionalization and patterning [61], cardiomyocyte differentiation [62], septation [63] and heart looping [64].

The multifunctional nature of RA-dependent signaling in development is also evident, when considering the tissues requiring this molecule for proper development. For example, studies carried out in both vertebrates and invertebrate chordates have shown that RA plays crucial roles in the differentiation of derivatives of all three embryonic tissue layers (ectoderm, mesoderm and endoderm), mainly by controlling regional patterning and tissue maturation [51, 52]. Moreover, RA is also required for proper differentiation of post-migratory neural crest cells [65], which some consider as a fourth embryonic tissue layer, and plays an instrumental role in germ cells differentiation [66].

The pleiotropy of RA-dependent functions is actually not limited to time and space, but includes also the biological processes controlled by this vitamin A derivative. Hence, RA, for example, is known to be involved in the regulation of cell differentiation, cell proliferation, cell survival, cell migration and apoptosis in vertebrate embryos [67]. This diversity of RA-controlled processes is exemplified by the action of this vitamin A derivative on neural cells: RA can induce neurite outgrowth [68], trigger the migration of neural precursors [69] and promote the specification of different neuronal populations, thus leading to the emergence of distinct neural territories [70]. Recent studies also indicate that RA signaling is involved in tissue regeneration in both vertebrates and invertebrates [71] suggesting that, apart from its general prodifferentiation capacities, vitamin A and its derivatives can also trigger transdifferentiation processes in certain biological contexts.

Vitamin A and RA Availability is Crucial and Highly Regulated during Development

Vitamin A metabolism and signaling are tightly controlled processes involving multiple levels of regulation (Fig. 1B). Ingested vitamin A is transported to target tissues in a complex with retinol binding protein (RBP) and transthyretin (TTR) [72], where cellular uptake is mediated by the transmembrane receptor STRA6 [73]. Once inside the cell, vitamin A is converted to RA in two separate oxidation steps. Vitamin A is first reversibly oxidized into retinaldehyde by alcohol dehydrogenase (ADH) or short-chain dehydrogenase/reductase (SDR) enzymes, such as RDH10 [74]. Retinaldehyde is subsequently irreversibly oxidized into RA by retinaldehyde dehydrogenase (RALDH) enzymes**,** with RALDH2 being the main RA synthesizing enzyme during early embryogenesis [53]. The combinatorial gene expression and concerted action of *rdh10* and *raldh2* have been suggested to constitute a so-called biosynthetic enzyme code required for axis formation and AP patterning of the vertebrate embryo [75, 76].

Elimination of RA is primarily catalyzed by members of the cytochrome P450 family [77]. In particular, the CYP26 enzymes, which are associated to the endoplasmic reticulum, are highly specific for RA and catalyze its hydroxylation into a wide variety of metabolites, such as 4-oxo RA, 4-OH RA and 18-OH RA. It is still controversial, though, whether these metabolites are biologically active [78, 79]. Furthermore, RA levels can be negatively regulated by endogenous removal of retinaldehyde. This can be achieved by reduction of retinaldehyde to vitamin A, a reaction carried out, for example, by DHRS3, a member of the SDR family [80].

Intracellular retinoid metabolism is further influenced by cellular retinol binding protein (CRBP) and cellular retinoic acid binding protein (CRABP) [81]. However, the

exact functions of these proteins remain elusive. While it has been suggested that CRBP act as a chaperone for vitamin A and may determine the cellular levels of vitamin A accumulation and esterification [82], CRABP might facilitate the translocation of RA from the cytoplasm to the nucleus thus acting as a coregulator of RA signaling [83].

The biological functions of vitamin A are chiefly mediated by the association of RA with heterodimers of two nuclear hormone receptors, the retinoic acid receptor (RAR) and the retinoid X receptor (RXR) [84], which together regulate the transcription of target genes in response to ligand binding. While RAR binds all-*trans* RA and 9-*cis* RA, RXR has been reported to only bind 9-*cis* RA *in vitro*. It is however still unclear, whether 9-*cis* RA is actually present in the embryo and thus whether this compound and its binding to RXR have a biological function during development [85, 86]. RAR/RXR heterodimers bind to specific DNA elements in the regulatory regions of target genes, called RA response elements (RAREs). Most RAREs consist of two direct repeats (DRs) with the canonical nucleotide sequence (A/G)G(G/T)TCA separated by a variable number of nucleotide spacers (usually either 1, 2 or 5 nucleotides) [87]. Fixation of RA leads to the binding of a coactivator complex to the heterodimer and activates the ligand-dependent transcription factor function of the RAR/RXR heterodimer [84].

One key element in tightly controlling the activity of RA signaling during development in time and space is the autoregulation of RA signaling components by RA. For example, while RA directly activates transcription of *rar* and *cyp26* [67, 88, 89], *raldh2* expression is repressed [90]. In combination with the differential expression domains of the three vertebrate *rar* paralogs (*rar*α, *rar*β and *rar*γ) [91] and the largely complementary domains of *cyp26* and *raldh* transcription in different tissues [92], the RA regulatory loop hence defines very specific time windows and tissue domains of RA sensitivity [93].

Developmental Functions of RA Signaling are Highly Dependent on Interactions with Other Signaling Pathways

A plethora of studies in vertebrate model systems have established that the functions of RA are tightly linked to the action of other signaling pathways during vertebrate development. Thus, RA signaling can functionally interact with the Fibroblast Growth Factor (FGF), Wnt, Nodal, Bone Morphogenetic Protein (BMP) and Sonic hedgehog (Shh) signaling cascades [94-97]. Crosstalks between RA and other signaling pathways can either be synergetic or antagonistic and can involve direct or indirect interactions (Fig. 2). Thus, RA-dependent regulatory loops implicating two or more signaling cascades and exhibiting very different architectures have been identified in many developmental processes in vertebrates. Such functional interactions between RA signaling and other cascades are involved in the control of somitogenesis [98], patterning of the CNS [99], otic and optic differentiation [59, 100], heart development [62], limb development [101] and pharyngeal differentiation [102].

Different modes of crosstalk of RA with other signaling cascades have been described. A typical situation observable in many developmental processes is that the action of RA signaling is upstream of other pathways (Fig. 2). For example, the retinoid signaling cascade activates the formation of skeletal muscle progenitors by repressing BMP signaling and by activating canonical Wnt signaling in the mesoderm [103]. In this case, RA controls developmental events upstream of the differentiation cascade. However, recent studies have indicated that RA signaling may sometimes

exhibit only a permissive role for developmental events. Thus, in the vertebrate limb, RA is required only transiently to inhibit axial FGF activity during the early steps of limb bud emergence [101].

Conversely, the RA signaling pathway can also act downstream of other signaling cascades, thus being tightly regulated as an effector regulatory system (Fig. 2). This situation is observable at different instances during vertebrate development, particularly during organogenesis. For example, during eye development proper RA signaling is achieved by activation of *raldh3* and concomitant repression of *cyp26a1* and *cyp26c1* by Vax2 [104].

RA signaling activity can further be inhibited by other signaling pathways, which leads to the creation of very well defined RA-free zones in the developing embryo (Fig. 2). This control of RA signaling activity is often achieved by downregulating expression of RA synthesizing (RALDH) and degrading (CYP26) enzymes. For example, during hindbrain development, repression of RA activity through activation of *cyp26b1* by FGF signaling is required for inhibition of neurogenesis in the center of the rhombomeres [105]. Similarly, during neural development, Pax6 proteins have been shown to repress *rarβ* and to activate *cyp26b1*, thus restricting the domains of RA signaling activity within the nascent CNS [106].

Another fundamental interaction of RA signaling and other pathways involves parallel and complementary action of different signaling cascades on the same target genes (Fig. 2). In this case, the coregulation of target genes by two or more pathways fine-tunes the output levels of the downstream genes. For example, it is now well established that RA and FGF signaling can act on the same target genes depending on the developmental context. Thus, during somitogenesis and hindbrain patterning these two signaling pathways act on the expression of *hox* genes, which results in the establishment of a *hox* code conveying regional identities to the body segments along the AP axis [107, 108]. This control of *hox* expression is mediated by *cdx* genes [109], which are direct convergence points (i.e. targets) of the RA, FGF and Wnt signaling pathways [110-112].

Taken together, the data obtained in vertebrate model organisms over the last few decades have succeeded in at least partially elucidating the complex functions of RA signaling during development. It is now obvious that RA signaling interacts at various levels with other signaling cascades and that the activity of RA is dependent on the developmental stage, the target tissue and the biological context. The biological functions of RA signaling are thus strongly interdependent with those of other developmental pathways. A prime example for this notion is the crosstalk between the RA and FGF signaling pathways, which together play instrumental roles in vertebrate development [98, 99, 101].

Acknowledgements

The authors are indebted to Vincent Laudet at the Ecole Normale Supérieure de Lyon, France, for critical reading of the manuscript. Our research is funded by ANR (ANR-07-BLAN-0038 and ANR-09-BLAN-0262-02), CNRS and CRESCENDO, a European Union Integrated Project of FP6. We would like to apologize to all colleagues, whose original work could not be cited due to space restrictions.

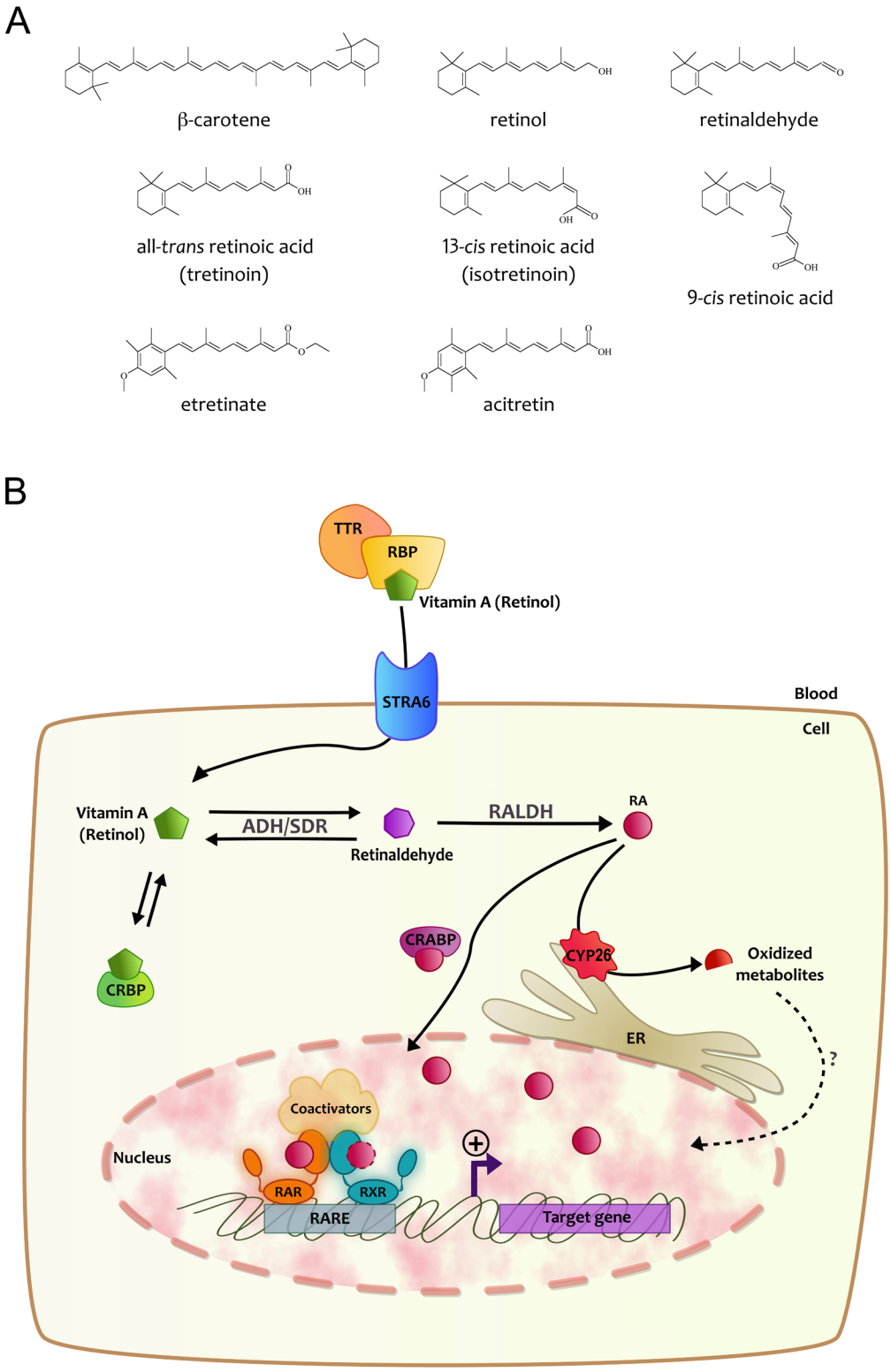


Fig. 1. Overview of vitamin A metabolism and signaling. (A) Chemical structures of retinoids. (B) Conversion of vitamin A to its major active metabolite, retinoic acid, and activation of retinoid-dependent signaling. ADH, alcohol dehydrogenase; CRABP, cellular retinoic acid binding protein; CRBP, cellular retinol binding protein; CYP26, cytochrome P450 family 26; ER, endoplasmic reticulum; RA, retinoic acid; RALDH, retinaldehyde dehydrogenase; RAR, retinoic acid receptor; RARE, retinoic acid response element; RBP, retinol binding protein; RXR: retinoid X receptor; SDR, short chain dehydrogenase/reductase; STRA6, stimulated by retinoic acid gene 6; TTR: transthyretin.

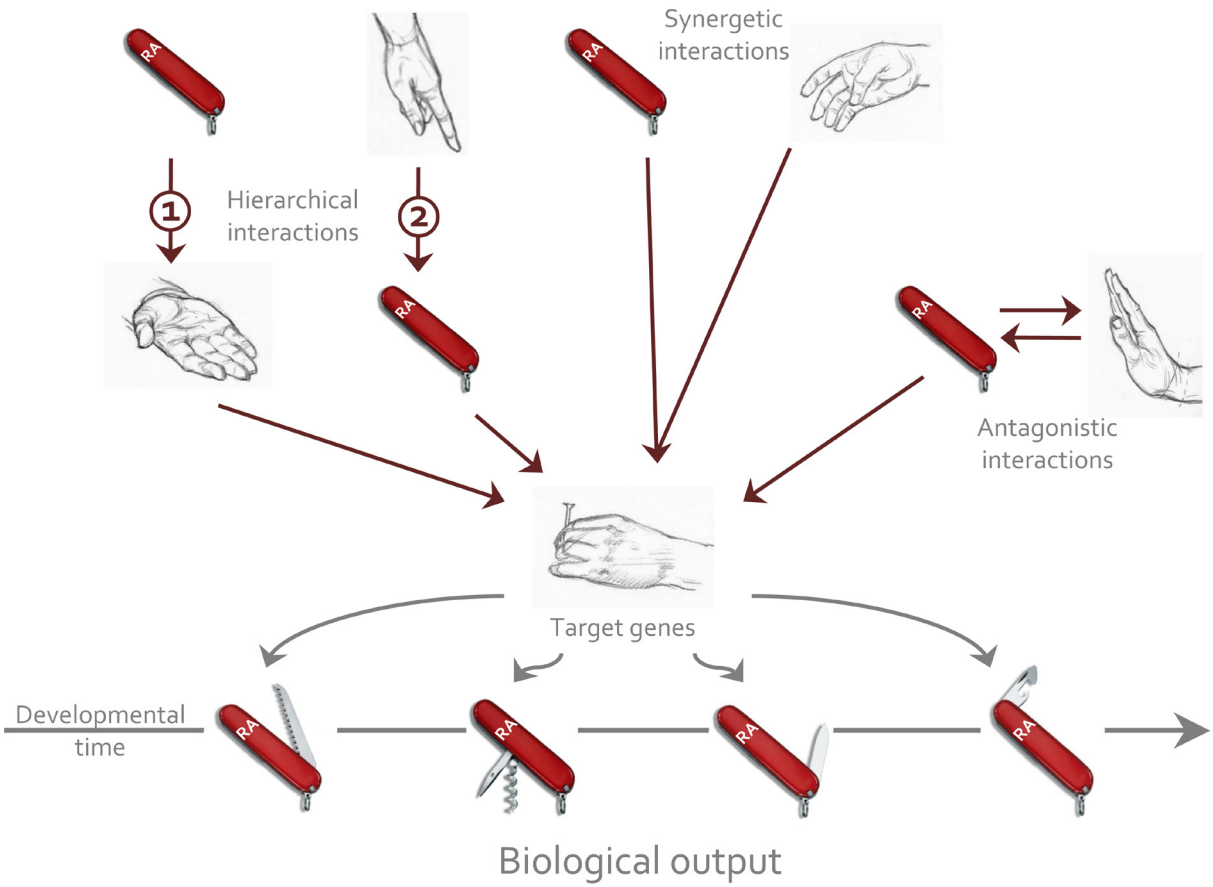


Fig. 2. The pleiotropic roles of retinoic acid (RA) signaling are mediated by sophisticated interplay with other signaling cascades. Hierarchical interactions are characterized by RA acting either upstream (1) or downstream (2) of other pathways. RA can also act synergetically or antagonistically with other signaling cascades.

References

[1] Allen LH, Haskell M. Estimating the potential for vitamin A toxicity in women and young children. J Nutr 2002;132:2907S-19S.

[2] Hale F. Pigs born without eye balls. J Hered 1933;24:105-6.

[3] Hale F. The relation of vitamin A to the eye development in the pig. J Anim Sci 1935:126-8.

[4] Wilson JG, Warkany J. Cardiac and aortic arch anomalies in the offspring of vitamin A deficient rats correlated with similar human anomalies. Pediatrics 1950;5:708-25.

[5] Cohlan SQ. Excessive intake of vitamin A as a cause of congenital anomalies in the rat. Science 1953;117:535-6.

[6] Cohlan SQ. Congenital anomalies in the rat produced by excessive intake of vitamin A during pregnancy. Pediatrics 1954;13:556-67.

[7] Shenefelt RE. Morphogenesis of malformations in hamsters caused by retinoic acid: relation to dose and stage at treatment. Teratology 1972;5:103-18.

[8] Hathcock JN, Hattan DG, Jenkins MY, McDonald JT, Sundaresan PR, Wilkening VL. Evaluation of vitamin A toxicity. Am J Clin Nutr 1990;52:183-202.

[9] Parkinson CE, Tan JC, Gal I. Vitamin A concentration in amniotic fluid and maternal serum related to neural-tube defects. Br J Obstet Gynaecol 1982;89:935-9.

[10] Venkatachalam PS, Belavady B, Gopalan C. Studies on vitamin A nutritional status of mothers and infants in poor communities of India. J Pediatr 1962;61:262-8.

[11] Godel JC, Basu TK, Pabst HF, Hodges RS, Hodges PE, Ng ML. Perinatal vitamin A (retinol) status of northern Canadian mothers and their infants. Biol Neonate 1996;69:133-9.

[12] Underwood BA. Maternal vitamin A status and its importance in infancy and early childhood. Am J Clin Nutr 1994;59:517S-24S.

[13] Ortega RM, Andrés PP, Martínez RM, López-Sobaler AM. Vitamin A status during the third trimester of pregnancy in Spanish women: influence on concentrations of vitamin A in breast milk. Am J Clin Nutr 1997;66:564-8.

[14] Schulz C, Engel U, Kreienberg R, Biesalski HK. Vitamin A and β-carotene supply of women with gemini or short birth intervals: a pilot study. Eur J Nutr 2007;46:12-20.

[15] Ross DA. Recommendations for vitamin A supplementation. J Nutr 2002;132:2902S-6S.

[16] US Teratology Society. Vitamin A during pregnancy. Teratology 1987;35:267-8.

[17] Rothman KJ, Moore LL, Singer MR, Nguyen US, Mannino S, Milunsky A. Teratogenicity of high vitamin A intake. N Engl J Med 1995;333:1369-73.

[18] Rosa FW. Retinoid embryopathy in humans. In: Koren G, editor. Retinoids in clinical practice. New York: Informa Healthcare; 1993. p. 77-109.

[19] Schardein JL, Macina OT. Human developmental toxicants: aspects of toxicology and chemistry. New York: Informa Healthcare; 2006.

[20] Christian MS, Mitala JJ, Powers WJ, Jr., McKenzie BE, Latriano L. A developmental toxicity study of tretinoin emollient cream (Renova) applied topically to New Zealand white rabbits. J Am Acad Dermatol 1997;36:S67-76.

[21] Rosa FW. Teratogenicity of isotretinoin. Lancet 1983;2:513.

[22] Geelen JA. Hypervitaminosis A induced teratogenesis. Crit Rev Toxicol 1979;6:351-75.

[23] Kamm JJ. Toxicology, carcinogenicity, and teratogenicity of some orally administered retinoids. J Am Acad Dermatol 1982;6:652-9.

[24] Benke PJ. The isotretinoin teratogen syndrome. J Am Med Assoc 1984;251:3267-9.

[25] Stern RS, Rosa F, Baum C. Isotretinoin and pregnancy. J Am Acad Dermatol 1984;10:851-4.

[26] Hersh JH, Danhauer DE, Hand ME, Weisskopf B. Retinoic acid embryopathy: timing of exposure and effects on fetal development. J Am Med Assoc 1985;254:909-10.

[27] Lammer EJ, Chen DT, Hoar RM, Agnish ND, Benke PJ, Braun JT, et al. Retinoic acid embryopathy. N Engl J Med 1985;313:837-41.

[28] Adams J, Lammer EJ. Neurobehavioral teratology of isotretinoin. Reprod Toxicol 1993;7:175-7.

[29] Lynberg MC, Khoury MJ, Lammer EJ, Waller KO, Cordero JF, Erickson JD. Sensitivity, specificity, and positive predictive value of multiple malformations in isotretinoin embryopathy surveillance. Teratology 1990;42:513-9.

[30] Brazzell RK, Vane FM, Ehmann CW, Colburn WA. Pharmacokinetics of isotretinoin during repetitive dosing to patients. Eur J Clin Pharmacol 1983;24:695-702.

[31] Colburn WA, Vane FM, Shorter HJ. Pharmacokinetics of isotretinoin and its major blood metabolite following a single oral dose to man. Eur J Clin Pharmacol 1983;24:689-94.

[32] Schaefer H, Zesch A. Penetration of vitamin A acid into human skin. Acta Derm Venereol 1975;74:50-5.

[33] Franz TJ, Lehman PA, Franz SF. Topical use of retinoic acid gel is not teratogenic. J Invest Dermatol 1993;100:490.

[34] Camera G, Pregliasco P. Ear malformation in baby born to mother using tretinoin cream. Lancet 1992;339:687.

[35] Lipson AH, Collins F, Webster WS. Multiple congenital defects associated with maternal use of topical tretinoin. Lancet 1993;341:1352-3.

[36] Navarre-Belhassen C, Blanchet P, Hillaire-Buys D, Sarda P, Blayac JP. Multiple congenital malformations associated with topical tretinoin. Ann Pharmacother 1998;32:505-6.

[37] Selcen D, Seidman S, Nigro MA. Otocerebral anomalies associated with topical tretinoin use. Brain Dev 2000;22:218-20.

[38] Willhite CC, Sharma RP, Allen PV, Berry DL. Percutaneous retinoid absorption and embryotoxicity. J Invest Dermatol 1990;95:523-9.

[39] Seegmiller RE, Carter MW, Ford WH, White RD. Induction of maternal toxicity in the rat by dermal application of retinoic acid and its effect on fetal outcome. Reprod Toxicol 1990;4:277-81.

[40] Jick SS, Terris BZ, Jick H. First trimester topical tretinoin and congenital disorders. Lancet 1993;341:1181-2.

[41] Seegmiller RE, Ford WH, Carter MW, Mitala JJ, Powers WJ, Jr. A developmental toxicity study of tretinoin administered topically and orally to pregnant Wistar rats. J Am Acad Dermatol 1997;36:S60-6.

[42] Shapiro L, Pastuszak A, Curto G, Koren G. Safety of first-trimester exposure to topical tretinoin: prospective cohort study. Lancet 1997;350:1143-4.

[43] Loureiro KD, Kao KK, Jones KL, Alvarado S, Chavez C, Dick L, et al. Minor malformations characteristic of the retinoic acid embryopathy and other birth outcomes in children of women exposed to topical tretinoin during early pregnancy. Am J Med Genet 2005;136A:117-21.

[44] Martinez-Frias ML, Rodriguez-Pinilla E. First-trimester exposure to topical tretinoin: its safety is not warranted. Teratology 1999;60:5.

[45] Vahlquist A, Rollman O. Etretinate and the risk for teratogenicity: drug monitoring in a pregnant woman for 9 months after stopping treatment. Br J Dermatol 1990;123:131.

[46] Maradit H, Geiger JM. Potential risk of birth defects after acitretin discontinuation. Dermatology 1999;198:3-4.

[47] Katz WI, Waalen J, Leach JJ. Acitretin in psoriasis: An overview of adverse effects. J Am Acad Dermatol 1999;41:S7-S12.

[48] Gronhoj Larsen F, Steinkjer B, Jakobsen P, Hjorter A, Brockhoff PB, Nielsen-Kudsk F. Acitretin is converted to etretinate only during concomitant alcohol intake. Br J Dermatol 2000;143:1164-9.

[49] Van Dorp DA, Arens JF. Biological activity of vitamin A acid. Nature 1946;158:60.

[50] Fujiwara S. Retinoids and nonvertebrate chordate development. J Neurobiol 2006;66:645-52.

[51] Campo-Paysaa F, Marlétaz F, Laudet V, Schubert M. Retinoic acid signaling in development: tissue-specific functions and evolutionary origins. Genesis 2008;46:640-56.

[52] Duester G. Retinoic acid synthesis and signaling during early organogenesis. Cell 2008;134:921-31.

[53] Niederreither K, Subbarayan V, Dollé P, Chambon P. Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. Nat Genet 1999;21:444-8.

[54] Durston AJ, Timmermans JP, Hage WJ, Hendriks HF, de Vries NJ, Heideveld M, et al. Retinoic acid causes an anteroposterior transformation in the developing central nervous system. Nature 1989;340:140-4.

[55] Sive HL, Draper BW, Harland RM, Weintraub H. Identification of a retinoic acid-sensitive period during primary axis formation in *Xenopus laevis*. Genes Dev 1990;4:932-42.

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

[57] Alexa K, Choe SK, Hirsch N, Etheridge L, Laver E, Sagerström CG. Maternal and zygotic *aldh1a2* activity is required for pancreas development in zebrafish. PLoS One 2009;4:e8261.

[58] Zhao X, Brade T, Cunningham TJ, Duester G. Retinoic acid controls expression of tissue remodeling genes *Hmgn1* and *Fgf18* at the digit-interdigit junction. Dev Dyn 2010;239:665-71.

[59] Bok J, Raft S, Kong KA, Koo SK, Dräger UC, Wu DK. Transient retinoic acid signaling confers anterior-posterior polarity to the inner ear. Proc Natl Acad Sci USA 2011;108:161-6.

[60] Xavier-Neto J, Rosenthal N, Silva FA, Matos TG, Hochgreb T, Linhares VL. Retinoid signaling and cardiac anteroposterior segmentation. Genesis 2001;31:97-104.

[61] Rosenthal N, Xavier-Neto J. From the bottom of the heart: anteroposterior decisions in cardiac muscle differentiation. Curr Opin Cell Biol 2000;12:742-6.

[62] Lin SC, Dollé P, Ryckebüsch L, Noseda M, Zaffran S, Schneider MD, et al. Endogenous retinoic acid regulates cardiac progenitor differentiation. Proc Natl Acad Sci USA 2010;107:9234-9.

[63] Li P, Pashmforoush M, Sucov HM. Retinoic acid regulates differentiation of the secondary heart field and TGFβ-mediated outflow tract septation. Dev Cell 2010;18:480-5.

[64] Chazaud C, Chambon P, Dollé P. Retinoic acid is required in the mouse embryo for left-right asymmetry determination and heart morphogenesis. Development 1999;126:2589-96.

[65] Dupé V, Pellerin I. Retinoic acid receptors exhibit cell-autonomous functions in cranial neural crest cells. Dev Dyn 2009;238:2701-11.

[66] Bowles J, Koopman P. Retinoic acid, meiosis and germ cell fate in mammals. Development 2007;134:3401-11.

[67] Blomhoff R, Blomhoff HK. Overview of retinoid metabolism and function. J Neurobiol 2006;66:606-30.

[68] Corcoran J, Shroot B, Pizzey J, Maden M. The role of retinoic acid receptors in neurite outgrowth from different populations of embryonic mouse dorsal root ganglia. J Cell Sci 2000;113:2567-74.

[69] Fu M, Sato Y, Lyons-Warren A, Zhang B, Kane MA, Napoli JL, et al. Vitamin A facilitates enteric nervous system precursor migration by reducing Pten accumulation. Development 2010;137:631-40.

[70] Novitch BG, Wichterle H, Jessell TM, Sockanathan S. A requirement for retinoic acid-mediated transcriptional activation in ventral neural patterning and motor neuron specification. Neuron 2003;40:81-95.

[71] Kaneko N, Katsuyama Y, Kawamura K, Fujiwara S. Regeneration of the gut requires retinoic acid in the budding ascidian *Polyandrocarpa misakiensis*. Dev Growth Differ 2010;52:457-68.

[72] Peterson PA. Studies on the interaction between prealbumin, retinol-binding protein, and vitamin A. J Biol Chem 1971;246:44-9.

[73] Kawaguchi R, Yu J, Honda J, Hu J, Whitelegge J, Ping P, et al. A membrane receptor for retinol binding protein mediates cellular uptake of vitamin A. Science 2007;315:820-5.

[74] Wu BX, Moiseyev G, Chen Y, Rohrer B, Crouch RK, Ma JX. Identification of RDH10, an all-*trans* retinol dehydrogenase, in retinal Müller cells. Invest Ophthalmol Vis Sci 2004;45:3857-62.

[75] Sandell LL, Sanderson BW, Moiseyev G, Johnson T, Mushegian A, Young K, et al. RDH10 is essential for synthesis of embryonic retinoic acid and is required for limb, craniofacial, and organ development. Genes Dev 2007;21:1113-24.

[76] Strate I, Min TH, Iliev D, Pera EM. Retinol dehydrogenase 10 is a feedback regulator of retinoic acid signalling during axis formation and patterning of the central nervous system. Development 2009;136:461-72.

[77] Theodosiou M, Laudet V, Schubert M. From carrot to clinic: an overview of the retinoic acid signaling pathway. Cell Mol Life Sci 2010;67:1423-45.

[78] Niederreither K, Abu-Abed S, Schuhbaur B, Petkovich M, Chambon P, Dollé P. Genetic evidence that oxidative derivatives of retinoic acid are not involved in retinoid signaling during mouse development. Nat Genet 2002;31:84-8.

[79] Reijntjes S, Blentic A, Gale E, Maden M. The control of morphogen signalling: regulation of the synthesis and catabolism of retinoic acid in the developing embryo. Dev Biol 2005;285:224-37.

[80] Feng L, Hernandez RE, Waxman JS, Yelon D, Moens CB. Dhrs3a regulates retinoic acid biosynthesis through a feedback inhibition mechanism. Dev Biol 2010;338:1-14.

[81] Napoli JL. Retinoic acid biosynthesis and metabolism. FASEB J 1996;10:993-1001.

[82] Levin MS. Cellular retinol-binding proteins are determinants of retinol uptake and metabolism in stably transfected caco-2 cells. J Biol Chem 1993;268:8267-76.

[83] Delva L, Bastie JN, Rochette-Egly C, Kraïba R, Balitrand N, Despouy G, et al. Physical and functional interactions between cellular retinoic acid binding protein II and the retinoic acid-dependent nuclear complex. Mol Cell Biol 1999;19:7158-67.

[84] Gronemeyer H, Gustafsson JA, Laudet V. Principles for modulation of the nuclear receptor superfamily. Nat Rev Drug Discov 2004;3:950-64.

[85] Mic FA, Molotkov A, Benbrook DM, Duester G. Retinoid activation of retinoic acid receptor but not retinoid X receptor is sufficient to rescue lethal defect in retinoic acid synthesis. Proc Natl Acad Sci USA 2003;100:7135-40.

[86] Calléja C, Messaddeq N, Chapellier B, Yang H, Krezel W, Li M, et al. Genetic and pharmacological evidence that a retinoic acid cannot be the RXR-activating ligand in mouse epidermis keratinocytes. Genes Dev 2006;20:1525-38.

[87] Balmer JE, Blomhoff R. A robust characterization of retinoic acid response elements based on a comparison of sites in three species. J Steroid Biochem Mol Biol 2005;96:347-54.

[88] Balmer JE, Blomhoff R. Gene expression regulation by retinoic acid. J Lipid Res 2002;43:1773-808.

[89] Wang Y, Zolfaghari R, Ross AC. Cloning of rat cytochrome P450RAI (CYP26) cDNA and regulation of its gene expression by all-*trans*-retinoic acid *in vivo*. Arch Biochem Biophys 2002;401:235-43.

[90] Niederreither K, McCaffery P, Dräger UC, Chambon P, Dollé P. Restricted expression and retinoic acid-induced downregulation of the retinaldehyde dehydrogenase type 2 (RALDH-2) gene during mouse development. Mech Dev 1997;62:67-78.

[91] Dollé P. Developmental expression of retinoic acid receptors (RARs). Nucl Recept Signal 2009;7:e006.

[92] McCaffery P, Wagner E, O'Neil J, Petkovich M, Dräger UC. Dorsal and ventral retinal territories defined by retinoic acid synthesis, break-down and nuclear receptor expression. Mech Dev 1999;82:119-30.

[93] White RJ, Nie Q, Lander AD, Schilling TF. Complex regulation of *cyp26a1* creates a robust retinoic acid gradient in the zebrafish embryo. PLoS Biol 2007;5:e304.

[94] Yasuda Y, Nishi N, Takahashi JA, Konishi H, Ohara I, Fujita H, et al. Induction of avascular yolk sac due to reduction of basic fibroblast growth factor by retinoic acid in mice. Dev Biol 1992;150:397-413.

[95] Ribes V, Le Roux I, Rhinn M, Schuhbaur B, Dollé P. Early mouse caudal development relies on crosstalk between retinoic acid, Shh and Fgf signalling pathways. Development 2009;136:665-76.

[96] Engberg N, Kahn M, Petersen DR, Hansson M, Serup P. Retinoic acid synthesis promotes development of neural progenitors from mouse embryonic stem cells by suppressing endogenous, Wnt-dependent nodal signaling. Stem Cells 2010;28:1498-509.

[97] Sheng N, Xie Z, Wang C, Bai G, Zhang K, Zhu Q, et al. Retinoic acid regulates bone morphogenic protein signal duration by promoting the degradation of phosphorylated Smad1. Proc Natl Acad Sci USA 2010;107:18886-91.

[98] Duester G. Retinoic acid regulation of the somitogenesis clock. Birth Defects Res C Embryo Today 2007;81:84-92.

[99] Gavalas A, Krumlauf R. Retinoid signalling and hindbrain patterning. Curr Opin Genet Dev 2000;10:380-6.

[100] Kumar S, Duester G. Retinoic acid signaling in perioptic mesenchyme represses Wnt signaling via induction of *Pitx2* and *Dkk2*. Dev Biol 2010;340:67-74.

[101] Lewandoski M, Mackem S. Limb development: the rise and fall of retinoic acid. Curr Biol 2009;19:R558-61.

[102] Birkholz DA, Olesnicky Killian EC, George KM, Artinger KB. *Prdm1a* is necessary for posterior pharyngeal arch development in zebrafish. Dev Dyn 2009;238:2575-87.

[103] Kennedy KA, Porter T, Mehta V, Ryan SD, Price F, Peshdary V, et al. Retinoic acid enhances skeletal muscle progenitor formation and bypasses inhibition by bone morphogenetic protein 4 but not dominant negative β-catenin. BMC Biol 2009;7:67.

[104] Alfano G, Conte I, Caramico T, Avellino R, Arnò B, Pizzo MT, et al. *Vax2* regulates retinoic acid distribution and cone opsin expression in the vertebrate eye. Development 2011;138:261-71.

[105] Gonzalez-Quevedo R, Lee Y, Poss KD, Wilkinson DG. Neuronal regulation of the spatial patterning of neurogenesis. Dev Cell 2010;18:136-47.

[106] Numayama-Tsuruta K, Arai Y, Takahashi M, Sasaki-Hoshino M, Funatsu N, Nakamura S, et al. Downstream genes of Pax6 revealed by comprehensive transcriptome profiling in the developing rat hindbrain. BMC Dev Biol 2010;10:6.

[107] Lohnes D, Mark M, Mendelsohn C, Dollé P, Dierich A, Gorry P, et al. Function of the retinoic acid receptors (RARs) during development (I). Craniofacial and skeletal abnormalities in RAR double mutants. Development 1994;120:2723-48.

[108] Partanen J, Schwartz L, Rossant J. Opposite phenotypes of hypomorphic and Y766 phosphorylation site mutations reveal a function for *Fgfr1* in anteroposterior patterning of mouse embryos. Genes Dev 1998;12:2332-44.

[109] Lohnes D. The Cdx1 homeodomain protein: an integrator of posterior signaling in the mouse. Bioessays 2003;25:971-80.

[110] Ikeya M, Takada S. Wnt-3a is required for somite specification along the anteroposterior axis of the mouse embryo and for regulation of *cdx-1* expression. Mech Dev 2001;103:27-33.

[111] Houle M, Prinos P, Iulianella A, Bouchard N, Lohnes D. Retinoic acid regulation of Cdx1: an indirect mechanism for retinoids and vertebral specification. Mol Cell Biol 2000;20:6579-86.

[112] Keenan ID, Sharrard RM, Isaacs HV. FGF signal transduction and the regulation of Cdx gene expression. Dev Biol 2006;299:478-88.

Retinoic Acid Signaling in Development: Tissue-Specific Functions and Evolutionary Origins

Florent Campo-Paysaa, Ferdinand Marlétaz, Vincent Laudet, Michael Schubert

Abstract

Retinoic acid (RA) is a vitamin A-derived morphogen important for axial patterning and organ formation in developing vertebrates and invertebrate chordates (tunicates and cephalochordates). Recent analyses of genomic data have revealed that the molecular components of the RA signaling cascade are also present in other invertebrate groups, such as hemichordates and sea urchins. In this review, we reassess the evolutionary origins of the RA signaling pathway by examining the presence of key factors of this signaling cascade in different metazoan genomes and by comparing tissue-specific roles for RA during development of different animals. This discussion of genomic and developmental data suggests that RA signaling might have originated earlier in metazoan evolution than previously thought. Based on this hypothesis, we conclude by proposing a scenario for the evolution of RA functions during development, which highlights functional gains and lineage-specific losses during metazoan diversification.

Introduction

Clinical and basic research carried out in the last fifty years has established that the vitamin A-derived morphogen retinoic acid (RA) is an important modulator of cell survival, cellular proliferation, differentiation, regionalization and organogenesis in the developing embryo (reviewed in Blomhoff and Blomhoff, 2006; Glover *et al.*, 2006; Mark *et al.*, 2006). Moreover, effects of RA on epithelial and tumor cells have also led to the use of vitamin A derivatives (generally referred to as retinoids) as anti-tumor agents and as therapeutic products to cure skin illnesses (Kligman, 1997). For example, in 1982, 13-*cis* RA (also called isotretinoin) was licensed under the name Accutane® for treatment of severe, cystic acne. Taken in early pregnancy, Accutane® also has strong teratogenic effects on the developing human fetus resulting in high percentages of spontaneous abortions and embryonic malformations (known as RA embryopathy). This RA-induced phenotype includes craniofacial malformations as well as abnormal central nervous system, heart and thymus development (Lammer *et al.*, 1985).

The roles of RA during development have been extensively investigated in vertebrates (chiefly in rodents, chicken and quail, frogs and fish) (reviewed in Maden, 2002; Glover *et al.*, 2006; Mark *et al.*, 2006). While vitamin A deficiency models and pharmacological approaches, such as treatment with vitamin A or RA, were initially used to study RA signaling, the development of genetic tools, such as targeted gene knockouts, has allowed an even more detailed assessment of the roles of RA during vertebrate embryogenesis. Targeted inactivation of RA signaling components has produced, for example, mouse embryos lacking RA synthesizing and degrading enzymes or the receptors mediating the RA signal (reviewed in Mark *et al.*, 2006). However, although it has been known for quite a while that RA signaling is not limited to vertebrates (Denucé, 1991; Holland and Holland, 1996), with only a few exceptions, the roles for RA signaling in embryonic development of invertebrates remain largely unknown.

Vertebrates belong to the phylum Chordata, which also include two invertebrate groups: the cephalochordates (e.g. amphioxus) and the tunicates or urochordates (e.g. ascidians, appendicularians and thaliaceans), with cephalochordates being the sister group of tunicates plus vertebrates (i.e. olfactores) (Bourlat *et al.*, 2006; Delsuc *et al*., 2006; reviewed in Schubert *et al*, 2006a). Together with the ambulacrarians (hemichordates and echinoderms) the chordates belong to the deuterostomes. Deuterostomes and protostomes, which include lophotrochozoans (e.g. annelids and mollusks) and ecdysozoans (e.g. insects and nematodes), represent the two major clades of bilaterian animals, while at the base of these two groups, there are three main clades of non-bilaterian metazoans: cnidarians, ctenophores and sponges (Dunn *et al.*, 2008).

After it became clear that RA signaling and function is not limited to vertebrates, RA signaling was considered to be chordate-specific, until some very recent analyses identified molecular components of the RA pathway in non-chordate deuterostomes, such as echinoderms and hemichordates (Cañestro *et al*., 2006; Howard-Ashby *et al*., 2006; Marlétaz *et al*., 2006; Ollikainen *et al*., 2006; Simões-Costa *et al*., 2008). In this review, we use these recent findings as a basis to discuss our current understanding of the evolution of the RA signaling cascade. We first analyze the evolutionary history of the molecular components constituting the RA signaling cascade, then review conserved and non-conserved functions of RA signaling in vertebrate and invertebrate development. Finally, by combining

molecular genomic and developmental data, we try to retrace the evolution of RA-dependent developmental mechanisms in metazoans.

An Overview over the RA Signaling Cascade

The retinoid family contains molecules derived from vitamin A (i.e. retinol), such as retinal and RA, which can exist in alternate stereoisomeric forms (e.g. all-*trans,* 9-*cis* or 13-*cis*) (Fig. 1a). During animal development, RA is synthesized from retinol in two steps of NAD-dependent oxidation, each catalyzed by different families of enzymes (Fig. 1b) (reviewed in Simões-Costa *et al*., 2008). After cellular uptake of ingested retinol by a system involving retinol binding protein (RBP) and the transmembrane protein STRA6 (Kawaguchi *et al*., 2007), the first step of RA synthesis is the reversible oxidation of retinol to retinal by enzymes of two different families: the alcohol dehydrogenases (ADHs/RolDHs) and the short-chain dehydrogenases/reductases (SDRs) (Kim *et al*., 1992; Boleda *et al*., 1993; Yang *et al*., 1994). ADHs can metabolize 9-*cis* and 13-*cis* retinol as well as all-*trans* retinol (Yang *et al.,* 1994), whereas SDRs can only catalyze the oxidation of all-*trans* retinol. The second step is the irreversible oxidation of retinal to retinoic acid (RA) predominantly by enzymes of the aldehyde dehydrogenase family (ALDH) (Duester, 1996). ALDHs catalyze the oxidation of both 9-*cis* and all-*trans* retinal (Labrecque *et al*., 1995). In vertebrates, studies on ALDHs have focused on ALDH1a1, 2, 3 and ALDH8 also known as, respectively, RALDH1, 2, 3 and 4 (for retinaldehyde dehydrogenases) (Lee *et al*., 1991; Niederreither *et al*., 1997; Lin *et al*., 2003). RALDH-dependent RA synthesis plays a critical role during the embryonic development of vertebrates. For example, loss of *raldh2* in mice results in severe developmental phenotypes that are almost completely rescued by maternal RA administration (Niederreither *et al*., 1999). In addition to ALDH enzymes, some members of the CYP450 family also catalyze the oxidations of retinol to retinal and of retinal to RA. *In vitro*, this has been shown for CYP1A (Roberts *et al*., 1992) and CYP2J4 (Raner *et al*., 1996), while *in vivo* synthesis of RA has recently been reported for CYP1B1 in mouse embryos (Chambers *et al*., 2007).

Other proteins, such as cellular retinoid binding proteins (CRBP and CRABP), also affect RA synthesis through their interaction with retinoids (retinol interacts with CRBP, while RA associates with CRABP) (Napoli, 1999), but the exact role of these factors still remains elusive. It has been suggested that CRBP and CRABP regulate the availability of RA *in vivo*: for example, the retinol dehydrogenase activity of SDRs is activated in the presence of CRBPI in mice (Yost *et al*., 1988), while CRABPII mediates the transport of RA to the nucleus (Donovan *et al*., 1995) and is involved in regulating the transcriptional activity of RA signaling (Delva *et al*., 1999).

The availability of endogenous RA is also controlled through its degradation by proteins of the cytochrome P450 family, chiefly in the CYP26 subfamily (White *et al*., 1997). These enzymes catalyze the oxidation of RA into a wide variety of metabolites, such as 4-oxo-RA, 4-OH-RA or 18-OH-RA. Whether these metabolites have a biological activity is still controversial (Niederreither *et al*., 2002; Reijntjes *et al*., 2005). Even if most studies have focused on the roles of CYP26 proteins during development, it is known that other CYPs, such as CYP2B6, CYP2C8, CYP3A4/5 or CYP2A6, can also catabolize RA (Marill *et al*., 2000).

RA functions during development are mediated by heterodimers of two members of the nuclear hormone receptors superfamily – RAR and RXR (Fig. 1b) (reviewed in Gronemeyer *et al.*, 2004). RAR/RXR heterodimers bind to specific DNA

elements in the regulatory regions of target genes, called RA response elements (RAREs). Most RAREs consist of two direct repeats (DRs) with the canonical nucleotide sequence (A/G)G(G/T)TCA separated by a variable number of nucleotidic spacers (usually either 1, 2 or 5 nucleotides) (reviewed in Balmer and Blomhoff, 2005; Chambon, 1996; Ross *et al*., 2000). Fixation of RA activates the ligand-dependent transcription factor function of the RAR/RXR heterodimer and leads to the binding of a coactivator complex to the heterodimer (reviewed in Gronemeyer *et al*., 2004). *In vitro,* RAR can bind both all-*trans* and 9-*cis* RA, whereas RXR only binds 9-*cis* RA. However, it is still unclear whether 9-*cis* RA is actually present in the embryo and thus whether this compound and its binding to RXR have a biological function during development (Mic *et al.*, 2003; Calleja *et al*., 2006).

In vertebrates, RA activates, directly or indirectly, the expression of a wide range of target genes (reviewed in Balmer and Blomhoff, 2002). Two types of these targets are of particular importance during embryonic development: (i) genes coding for factors involved in the RA pathway (synthesis, metabolism and signaling), and (ii) homeobox-containing transcription factors. Thus, as part of the RA pathway, RA directly controls expression of *cyp26a1, rarα, rarβ, rarγ* and *crabp2* (Balmer and Blomhoff, 2002; Blomhoff and Blomhoff, 2006), and probably indirectly regulates the expression of *raldh2* (Balmer and Blomhoff, 2002) suggesting that feedback loops are an important part of the RA signaling control machinery. Indeed, RA signaling inhibits *raldh* expression (Niederreither *et al*., 1997), while directly activating *cyp26* (Wang *et al*., 2002). This indicates that the availability of RA *in vivo* is under the tight control of a negative feedback loop restricting RA activity to specific zones of the developing embryo (White *et al.*, 2007). In addition, RA acts directly on the expression of transcription factor genes containing homeobox domains (e.g. *hoxa1, hoxb1, hoxa4, hoxd4, cdx1* and *pit1*), which have crucial roles in establishing anteroposterior (AP) positional information in the developing embryo (reviewed in Balmer and Blomhoff, 2002). Thus, another important function of the RA signaling cascade is the control of regional patterning during development by means of homeobox transcription factors. Finally, RA and other retinoids also have non-genomic functions, for example as pigments in the visual cycle (reviewed in Blomhoff and Blomhoff, 2006) or as substrate for retinoylation of proteins (reviewed in Breitman and Takahashi, 1996).

During development, RA has long been thought to act over a distance in the form of a gradient, hence controlling the establishment of positional identity along the anteroposterior (AP) body axis in the vertebrate embryo. This idea is based on studies on the distribution of endogenous RA in vertebrate embryos that have shown that RA concentrations are generally higher posteriorly than anteriorly (Chen *et al.*, 1994; Maden *et al.*, 1998). In this scenario, RA diffuses towards the anterior pole of the embryo creating a gradient along the AP axis. This RA gradient then activates expression of a variety of genes along this gradient and thus along the AP axis: high concentrations in the posteriormost part of the embryo activate (or repress) a specific group of genes, while low concentrations in the anterior part activate (or repress) other genes. Although some studies have supported this simple model, more recent work has questioned this production-diffusion gradient model for RA-dependent regionalization along the whole AP axis during development. For example, endogenous RA deficiency can be rescued *in vivo* by treatment with a uniform concentration of RA (Gale *et al*., 1999; Niederreither *et al*., 2000). Moreover, it has been shown that restricted RA degradation is crucial for the establishment of localized RA activity zones (Hernandez *et al*., 2007; White *et al*., 2007). Thus, local

expression of *cyp26* genes during hindbrain development is fundamental for establishing differential RA signaling activities along the AP axis of the hindbrain (reviewed in Glover *et al*., 2006). In addition, it is now clear that while RA synthesis takes place in a very broad posterior domain during development, the main zone of RA production being the somitic mesoderm (Berggren *et al*., 1999), RA synthesis also occurs, for example, in the retina (McCaffery *et al*., 1999), the testis (Vernet *et al*., 2006) and the heart (reviewed in Xavier-Neto *et al*., 2001). Thus, the simple model of a RA gradient coming from a posterior source is in fact more complex than anticipated.

In addition to the tightly controlled synthesis and degradation of RA during development in vertebrates, RA activity is further regulated by *crbp* and *crabp* as well as *rarβ and rarγ,* which have spatially and temporally restricted expression domains during development (Dolle *et al*., 1990; Ruberte *et al*., 1991). Taken together, this suggests that RA activity during development is far more complex than initially anticipated and probably does not correspond to a simple concentration gradient of endogenous RA. Instead, the embryo appears to be subdivided into several modules of different RA responsiveness, which depend on the presence of local sources and sinks of RA and on the sensitivity of a given tissue to RA, which varies both temporally and spatially during development. Thus, the action of RA signaling during the development appears to be both stage- and tissue-specific, as we will discuss below in more detail.

Evolutionary Origins of RA Signaling

RA signaling was long thought to be vertebrate-specific, but studies in invertebrate chordates (e.g. amphioxus and different tunicate species) provided evidence for roles of RA during development of all chordates (Denucé, 1991; De Bernardi *et al.*, 1994; Katsuyama *et al.*, 1995; Holland and Holland, 1996; Hinman and Degnan, 1998; Cañestro and Postlethwait, 2007). Moreover, recent studies have suggested a possible presence of this pathway in other deuterostomes: bioinformatic analyses have revealed *rar*, *cyp26* and *raldh* homologs in the genome of the sea urchin *Strongylocentrotus purpuratus* and *raldh* and *cyp26* homologs in EST sequences from the hemichordate *Saccoglossus kowalevskii* (Cañestro *et al*., 2006; Howard-Ashby *et al*., 2006; Marlétaz *et al*., 2006; Ollikainen *et al*., 2006; Simões-Costa *et al*., 2008). Thus, RA signaling probably already existed in the last common ancestor of all deuterostomes. To further investigate the evolution of the RA signaling cascade, we decided to expand our bioinformatic searches for molecular components of the RA signaling cascade to genomic data from various invertebrates, including two lophotrochozoans (the gastropod *Lottia gigantea* and the annelid worm *Capitella capitata*), three ecdysozoans (the fruit fly *Drosophila melanogaster,* the branchiopod *Daphnia pulex* and the nematode worm *Caenorhabditis elegans*) and one cnidarian (the sea anemone *Nematostella vectensis*). Using the Basic Local Alignment and Search Tool (BLAST), we searched the available genomic data for major components of the canonical vertebrate RA signaling cascade, namely *rar, rxr, aldh1* and *aldh8* (as putative *raldhs*) and *cyp26* (Fig. 2). For all these genes, the BLAST results were validated by phylogenetic analyses (data not shown).

For *aldh1* and *aldh8* as putative *raldhs*, we found candidates in all sampled species except the appendicularian *O. dioica*, where both genes seem to be absent, the insect *D. melanogaster* that lacks an *aldh8* and the sea urchin *S. purpuratus*, where the *aldh1* gene might be absent (Fig. 2). Moreover, while *cyp26* genes seem

absent from the tunicate *O. dioica*, the three ecdysozoans (*D. melanogaster, D. pulex* and *C. elegans*) and the cnidarian *N. vectensis*, we found hits for *cyp26* genes in both lophotrochozoan species (*L. gigantea* and *C. capitata*) (Fig. 2).

In this analysis, we were particularly interested in identifying homologs of *rar* and *rxr*, since these genes are the main mediators of RA signaling. While we found no *rar* homologs in the three ecdysozoan (*D. melanogaster, D. pulex* and *C. elegans*), the appendicularian (*Oikopleura dioica*) or the cnidarian (*N. vectensis*) genomes, to our great surprise we found homologs of *rar* in the genomes of both lophotrochozoan species (*L. gigantea* and *C. capitata*). Comparison of the amino acid residues responsible for contacting RA in the RAR ligand binding pocket (Renaud *et al.*, 1995; Escrivà *et al.*, 2006) revealed a significant degree of conservation between the two lophotrochozoan RARs and deuterostome RARs (Table 1). Of the 14 sites described for human RARγ (Renaud *et al.*, 1995), 13 are conserved in the two other human paralogs (RARα and RARβ), 12 in the RAR of *C. intestinalis*, 9 in amphioxus, at least 5 in sea urchin (Ollikainen *et al.*, 2006) and 8 in both *L. gigantea* and *C. capitata*. Moreover, in the binding pocket of human RARγ Lys236, Arg278 and Ser289 directly interact with the carboxylate moiety of the RA molecule (Renaud *et al.*, 1995). At these sites, identical amino acids are found in human RARα and RARβ as well as in the RAR of *C. intestinalis*. In contrast, only two of these residues (Arg278 and Ser289) are conserved in *L. gigantea* and only one (Arg278) in amphioxus, sea urchin and *C. capitata* (Table 1) (Ollikainen *et al*., 2006). However, in *C. capitata* Ser289 is replaced by a threonine, which represents a conservative amino acid substitution (Styczynski *et al*., 2008). Thus, the RARs from *L. gigantea* and *C. capitata* display a conservation of ligand binding residues with human RARγ that is comparable to that of amphioxus RAR, a receptor bound and activated by RA (Escrivà *et al.*, 2002, 2006). In sum, although there is so far no experimental evidence suggesting that RAR is a functional receptor for RA in lophotrochozoans, it is conceivable that the RARs of *L. gigantea* and *C. capitata* might bind RA.

For *rxr* genes, we found homologs in almost all sampled species except in the nematode *C. elegans* and in the cnidarian *N. vectensis*. However, the absence of *rxr* in these two animals might be due to species-specific losses, since likely *rxr* homologs have been identified before in both cnidarians (Kostrouch *et al.*, 1998) and nematodes (Shea *et al.*, 2004). In general, this wide distribution of *rxr* genes in metazoan animals is consistent with a role of RXR proteins as heterodimeric binding partners of a whole suite of different nuclear receptors including RAR (reviewed in Gronemeyer *et al*., 2004).

One interesting feature of this analysis is the confirmation of the absence of most components of the RA machinery in appendicularian tunicates (Cañestro *et al*., 2006). Thus, *O. dioica* is an exception within chordates, because it lacks most elements of the RA signaling cascade (*rar*, *aldh1* and *cyp26*) and because it does not seem to use the same pathway in AP patterning during embryonic development (Cañestro and Postlethwait, 2007). Instead, exogenous RA affects organogenesis in *O. dioica*, which has been interpreted as a non-classical (i.e. RAR-independent) morphogenetic role of RA in this animal (Cañestro and Postlethwait, 2007). Since RA signaling mediated by RAR plays important roles in the development of vertebrates, cephalochordates (e.g. *Branchiostoma floridae*) and other tunicates, such as ascidians (e.g. *Ciona intestinalis*) (Fujiwara, 2006), the absence of *rar* in appendicularian tunicates is evidently a derived character resulting from lineage-specific gene loss. This loss strongly suggests that at least some tunicate models

cannot be used for studying the evolution of RA signaling in chordates and its roles during development.

In contrast to this lineage-specific loss in appendicularian tunicates, we found candidates for all key players of the RA signaling cascade, in particular *rar*, in lophotrochozoans, which changes our perspective on the evolutionary origins of this signaling cascade. Thus, RA signaling is probably not chordate- or deuterostome-specific, but instead, this pathway might be present in both deuterostomes and lophotrochozoans. This suggests that RA signaling was already present in Urbilateria, the last common ancestor of protostomes and deuterostomes. Since both the mediator of RA signaling (*rar*) and the RA-degrading machinery (*cyp26*) are probably absent from the genomes of all three ecdysozoans studied, these main components of the RA pathway might have secondarily been lost in the ecdysozoan lineage. Although the prospects of a protostome RA signaling cascade are very tempting, the *in silico* results require further analysis and experimental verification to clearly establish whether RA signaling is functional and active during lophotrochozoan development. Moreover, to more precisely reconstruct the evolutionary history of the RA signaling cascade, it will be important to search for components of this pathway in organisms located at basal positions within both the lophotrochozoans and ecdysozoans.

Tissue-Specific Roles of RA Signaling during Development

It is now well established that, in addition to vertebrates, RA signaling is also important for the development of invertebrate chordates, such as amphioxus and ascidians. However, the role of RA signaling during development has been extensively studied only in vertebrates and relatively little is known in other groups. In the next part of the review, we focus on the roles of RA signaling in the developing embryo – especially on the tissue-specific functions of RA – and compare these functional roles between different animals.

Mesoderm, somitogenesis and left/right asymmetry

The most important tissues for RA production during embryonic development of vertebrates are the presomitic mesoderm (PSM) and the newly formed somites (reviewed in Dubrulle and Pourquie, 2004). In these tissues, RA is mainly synthesized by RALDH2 (Niederreither *et al*., 1997). Even if this mesodermal production also acts on other tissues, RA originating from PSM strongly influences the process of somitogenesis in the mesoderm itself (Fig. 3). In vertebrate somitogenesis, mesodermal units form along the AP embryonic axis. These repeated units give rise to various muscles and to the axial skeleton. A determination front, which travels along the AP extent of the PSM, controls the AP identity of the somites and activates a segmentation clock involving *notch* signaling that subsequently triggers somite segmentation and differentiation (reviewed in Dubrulle and Pourquie, 2004).

In vertebrates, RA signaling acts through two mechanisms in the establishment of the directional, periodic and synchronous segmentation of the PSM into somites: (i) the RA signaling pathway, together with FGF signaling, controls the position along the AP body axis of the traveling determination front (reviewed in Dubrulle and Pourquie, 2004) and (ii) RA signaling is required for symmetric formation of the somites along the left/right body axis (Kawakami *et al*., 2005; Vermot and Pourquie, 2005; Vermot *et al*., 2005). The first mechanism consists of reciprocal

inhibition of *fgf8* expression by RA and of *raldh2* expression by FGF8, which leads to the formation of opposite gradients of RA (anterior) and FGF8 (posterior) in the PSM with the determination front of segmentation corresponding to the limit between these two “gradients” (Diez del Corral *et al*., 2003). In general, RA signaling inhibits the signals from the caudal source of FGF8, hence triggering somite formation by activation of segmentation genes (Moreno and Kintner, 2004). Recent studies have also proposed Wnt signaling as a key player in reciprocal inhibition of RA and FGF: in chicken, *wnt8c* expression in the transition zone is activated by FGF8 emanating from the posterior stem zone with WNT8C in turn activating expression of *raldh2* at more anterior levels of the AP (Olivera-Martinez and Storey, 2007).

Moreover, an integral part of the segmentation and differentiation process during somitogenesis is the differential control in the PSM of *hox* gene expression by the RA/FGF system. While FGF8 seems to preferentially activate the expression of posterior *hox* genes at the 5’ end of the cluster, RA triggers expression of anterior *hox* genes in the 3’ part of the cluster. In the zone between the respective sources of RA and FGF8, where low concentrations of RA and FGF8 coexist, the combination of both signaling pathways triggers the induction of *hox* genes located at the center of the cluster. Thus, AP patterning mediated by collinear *hox* gene expression is linked to the segmentation machinery involved in somitogenesis (Dubrulle *et al*., 2001; Zákány *et al*., 2001; Diez del Corral and Storey, 2004). The differential induction of *hox* genes by RA/FGF along the AP axis of the paraxial mesoderm, resulting in a specific *hox* code that defines differential AP positions, is essential for the global AP patterning of the embryo (Fig. 3). Combinations of the expression of several *hox* genes at any given point in the presomitic mesoderm will give rise to different positional identities along the AP axis. This patterning of the paraxial mesoderm leads, for example, to the AP specification of vertebral identities. This specification involving RA signaling and *hox* genes also requires other factors, such as *cdx1* (Houle *et al*., 2000; Allan *et al*., 2001).

The role of RA in somitogenesis is well conserved within vertebrates, since similar mechanisms have been identified in different vertebrate models, such as mice, chicken and frogs (Dubrulle *et al*., 2001; Dubrulle and Pourquie, 2004; Moreno and Kintner, 2004). However, it is not certain whether outside vertebrates, somitogenesis involves RA signaling, because, while tunicates do not have somites at all (Passamaneck and Di Gregorio, 2005), somitogenesis in cephalochordates seems to differ somewhat from that in vertebrates. In the cephalochordate amphioxus, there is no PSM and somite formation can be divided into two different phases: the anteriormost somites arise from the dorsolateral wall of the archenteron by enterocoelous outpocketing, while the remaining somites form asymmetrically one at a time by budding directly from the tail bud (Schubert *et al*., 2001). Treatment with exogenous RA does not appear to affect the number of somites, but seems to be required for the proper positioning of somites along the AP axis, as RA treatment shifts the somites anteriorly (Schubert *et al*., 2006b). Taken together, it is conceivable that a function for RA signaling in segmentation of the somites evolved specifically in the vertebrate lineage, probably in parallel with the elaboration of PSM. In this context, it will be useful to study putative roles for RA signaling in mesodermal segmentation of other invertebrates, such as annelids.

The second role for RA signaling in vertebrate somite formation is to enable the symmetric formation of somites along the left/right body axis. Disruption of RA synthesis in the PSM results in desynchronization of somite formation in mouse, chicken and zebrafish embryos (Kawakami *et al*., 2005; Vermot and Pourquie, 2005;

Vermot *et al*., 2005) resulting in asymmetric and asynchronous production of somites on the left and on the right side of the body. Surprisingly, supernumerary somites resulting from RA deficiency do not appear on the same side in chicken and mice (Vermot and Pourquie, 2005). Thus, symmetric and synchronous somitogenesis in vertebrates requires RA signaling as protection against asymmetric signals in the segmenting PSM (Vermot and Pourquie, 2005). In amphioxus, the left and right somite series are normally offset by half a segment (Schubert *et al.*, 2001). Thus, a role for RA signaling in establishing a symmetric arrangement of somites on the right and left sides might have evolved specifically in the vertebrate lineage.

In addition to playing a role in maintaining symmetry during somite formation, the RA signaling pathway might also be involved in early determination of global left/right asymmetry during vertebrate development (Fig. 3). FGF-dependent transport of vesicles loaded with RA and SHH to one side of the early embryo has been suggested to be essential for the establishment of the left/right asymmetry in vertebrates (Tanaka *et al*., 2005), although this hypothesis has subsequently been challenged (Sirbu and Duester, 2006). In contrast, RA signaling probably controls expression of genes involved in setting up left/right asymmetry in the early vertebrate embryo, such as *lefty*, *pitx* and *nodal* (Chazaud *et al*., 1999; Tsukui *et al*., 1999). In invertebrate chordates, as in other invertebrate deuterostomes, the establishment of left/right asymmetry probably involves the same gene hierarchy as in vertebrates (Chea *et al.*, 2005; Duboc *et al.*, 2004, 2005). For example, as in vertebrates, homologs of *shh, pitx* and *nodal* are all expressed asymmetrically during amphioxus development (Schubert *et al*., 2005; Yu *et al.*, 2007). However, both in amphioxus (Schubert *et al*., 2005) and ascidians (Hinman and Degnan, 1998) treatments with exogenous RA do not obviously alter the establishment of the left/right body axis suggesting that a role for RA signaling in establishing global left/right asymmetry in the embryo is a vertebrate innovation (Schubert *et al*., 2005).

In addition to its roles in AP patterning and establishment of left/right asymmetry, in vertebrate mesoderm, RA signaling is also involved in the development of various organs and mesodermal derivatives (Fig. 3), such as kidneys (Burrow, 2000; Wingert *et al*., 2007), limb buds (Tabin, 1995), skin appendages (Chuong, 1993), muscles (Hamade *et al*., 2006) and the heart (Xavier-Neto *et al*., 2001). For example, in the developing heart, RA acts directly on cardiac field specification (Niederreither *et al.*, 1999; Xavier-Neto *et al*., 2001), AP patterning (Rosenthal and Xavier-Neto, 2000) and heart looping (Chazaud *et al*., 1999; Wasiak and Lohnes, 1999). At least the role for RA signaling in AP patterning of the heart is well conserved within vertebrates, as treatments with exogenous RA during development result in similar phenotypes in mice (Xavier-Neto *et al.*, 1999), chicken (Osmond *et al*., 1991), zebrafish (Stainier and Fishman, 1992) and lampreys (Kuratani *et al*., 1998). In the developing vertebrate heart, RA diffusing from posterior paraxial and lateral mesoderm specifies posterior cardiac precursors to a cardiac inflow chamber fate (i.e. atrium and sinus venosus). Moreover, a caudorostral wave of *raldh2* expression in the posterior cardiac field further increases RA concentrations at the time these heart fates are established (Hochgreb *et al*., 2003). Since very few studies have been focusing on the role of RA signaling during organogenesis in invertebrates, it remains to be established whether this intricate RA-dependent mechanism is vertebrate-specific or whether invertebrates pattern their hearts and pumping organs with similar molecular mechanisms (reviewed in Simões-Costa *et al*., 2005).

Endoderm patterning and pharyngeal development

Retinoid deficiency during embryonic development triggers severe defects of endodermal structures. For instance, the pharyngeal endoderm is expanded in vitamin A deficient quail embryos, while the postpharyngeal endoderm (i.e. the posterior foregut) is reduced (Quinlan *et al*., 2002). Comparable phenotypes were observed in zebrafish mutants lacking *raldh2* activity (Stafford and Prince, 2002). Conversely, addition of exogenous RA posteriorizes the foregut transforming anterior regions into pancreas and liver, which usually form in the trunk endoderm (i.e. in the midgut) (Stafford and Prince, 2002). In mice, high concentrations of RA are required for the patterning of branchial arches 3 to 6, while the differentiation of the second arch requires lower concentrations and the first branchial arch has undetectable levels of RA (Wendling *et al*., 2000; reviewed in Mark *et al*., 2004). These data suggest that in the pharynx very localized activity of RA along the AP axis is necessary for proper patterning. Thus, in vertebrates, RA controls AP patterning of the endoderm and the action of RA signaling is crucial anteriorly, in the developing pharynx and pharyngeal endoderm (Fig. 3). It is now well established that the role of RA signaling in patterning the pharyngeal endoderm is independent of the action of neural crest cells (NCCs) (Veitch *et al*., 1999; Wendling *et al*., 2000; Matt *et al*., 2003) that migrate from the central nervous system (CNS) into the pharyngeal region and contribute to branchial arch morphogenesis (Noden, 1983): patterning of the branchial arches occurs in the absence of NCCs (Veitch *et al*., 1999) and is dependent on signals provided directly by the mesendoderm (Piotrowski and Nusslein-Volhard, 2000). One of these signals is most likely RA (Wendling *et al*., 2000), whose action in vertebrate pharyngeal patterning might, at least partially, be mediated by activation of *hox* genes (Wendling *et al*., 2000; Matt *et al*., 2003).

In the cephalochordate amphioxus, which lacks migrating NCCs, patterning of the pharyngeal endoderm is also RA-dependent (Holland and Holland, 1996; Escrivà *et al*., 2002; Schubert *et al*., 2005). Treatments with exogenous RA lead to a loss of pharyngeal structures (e.g. mouth and gill slits) and push the posterior expression limits of pharyngeal marker genes (e.g. *pax1/9* and *otx*) anteriorly, while treatments with an RA antagonist as well as functional knockdown of *hox1* result in posterior expansion of both pharyngeal structures and marker gene expression in the pharyngeal endoderm (Holland and Holland, 1996; Escrivà *et al*., 2002; Schubert *et al*., 2005). Thus, in amphioxus foregut endoderm, RA signaling acts via *hox1* to limit the expression of genes required for pharyngeal specification (such as *pax1/9* and *otx*) to anterior, pharyngeal endoderm (Schubert *et al*., 2005). Interestingly, effects of RA on the development of pharyngeal structures in ascidians are similar to those in amphioxus and vertebrates. As in amphioxus, *pax1/9* and *otx* are expressed in the ascidian pharynx (Ogasawara *et al*., 1999; Hinman and Degnan, 2000) and in the ascidian *Herdmania curvata* exogenous RA decreases *otx* expression and induces a posteriorization of the endoderm leading to a complete loss of the branchial basket (Hinman and Degnan, 1998, 2000). Moreover, in some tunicates, such as the ascidian *Ecteinascidia turbinata*, a migratory cell population resembling vertebrate NCCs has recently been described (Jeffery *et al*., 2004), but a role for these cells in pharynx formation of tunicates remains elusive. These observations suggest that RA signaling was already implicated in AP patterning of the endoderm in the last common ancestor of all chordates. The involvement of NCCs in the formation of pharyngeal structures probably appeared later in evolution (reviewed in Graham and Smith, 2001; Trainor and Krumlauf, 2001) in parallel with the elaboration of new developmental roles for NCCs (Jeffery *et al*., 2004; reviewed in Marlétaz *et al*., 2006;

Schubert *et al.*, 2006a), the fourth embryonic tissue layer (Hall, 2000). To better understand the evolutionary origins of RA signaling in the endoderm it will be very important to investigate the roles of this signaling cascade in endodermal patterning of other invertebrates, such as hemichordates, mollusks or annelids.

In vertebrates, it has been shown that RA signals from the adjacent mesoderm are also involved in organogenesis of endodermal derivatives (Fig. 3). For example, pancreas and liver differentiation in the anterior trunk endoderm (i.e. in the anterior midgut) requires RA signals from the lateral plate mesoderm (reviewed in Kumar and Melton, 2003; Kumar *et al*., 2003). In contrast, more posteriorly located organs (e.g. the intestine) do not seem to need RA signaling to form. Thus, at least in vertebrates, patterning of the endoderm requires differential RA signaling activity along the AP axis with RA responsive zones (posterior pharynx, anterior gut) and zones not responsive to RA (anterior pharynx, posterior gut) (Wendling *et al*., 2000; Stafford and Prince, 2002; Kumar *et al*., 2003; reviewed in Kumar and Melton, 2003; Mark *et al*., 2004). These RA reactivity zones are created by localized sources of RA in presumptive fore- and midgut and adjacent mesoderm (produced by RALDH1, RALDH2 and RALDH3) combined with specific sinks of RA in the pharynx and organ primordia (such as the lung) (established by CYP26A1, CYP26B1 and CYP26C1) (MacLean *et al*., 2001; Abu-Abed *et al*., 2002; Blentic *et al*., 2003; Reijntjes *et al*., 2005). In addition, localized endodermal expression of different *rar* paralogs further fine-tunes this RA signaling activity along the AP axis of the developing endoderm (Mollard *et al*., 2000; Bertrand *et al*., 2007). Finally, localized expression of *rars* in dorsal and ventral areas of the vertebrate endoderm may even create differential RA signaling activities along the dorsoventral (DV) body axis (Pan *et al*., 2007). To understand how and when these complex patterns of RA signaling activity evolved in the endoderm, it will be crucial to study in much greater detail the functions of RA in the endoderm of different invertebrate systems.

Patterning of the central nervous system and neuronal differentiation

Functions of RA signaling in the central nervous system (CNS) have mainly been investigated during the establishment of the AP axis. In the CNS, RA does not globally specify the complete length of the AP axis, but instead has a more restricted and specific action on different zones of the CNS (Fig. 3), most importantly on the hindbrain (HB) (Godsave *et al*., 1998). Within the HB, which in vertebrates is both morphologically and molecularly subdivided into rhombomeres (Lumsden and Krumlauf, 1996), AP patterning is generally controlled by a *hox* code, involving nested collinear expression of *hox* genes along the AP axis, which defines the identity of each rhombomere (Trainor and Krumlauf, 2001). It is well established that RA regulates this collinear expression of *hox* genes in the vertebrate HB (Murakami *et al*., 2004; Glover *et al*., 2006). Treatments with RA induce an anterior expansion of the HB and the spinal cord (SC) and inhibit formation of anterior brain structures such as the forebrain (FB) and the midbrain (MB) (Durston *et al*., 1989). In addition, RA affects expression of marker genes of specific brain regions. For example, *otx2* and *en2* expression is repressed in the presence of exogenous RA, while HB marker genes such as *hoxb3* and *hoxb4* are activated (Godsave *et al*., 1998). RA probably acts in two steps during the development of the HB: (i) before somitogenesis, RA induces the physical segmentation of rhombomeres and (ii) after the onset of somitogenesis, RA defines rhombomere identities by controlling the establishment of the *hox* code (Marshall *et al*., 1992; Wood *et al*., 1994).

HB patterning also requires different levels of RA signaling with the differentiation of posterior rhombomeres requiring higher concentrations of RA than that of more anterior rhombomeres (Dupé and Lumsden, 2001). Thus, RA functions within the vertebrate HB probably rely on the formation of different zones of RA activity. Recent publications suggest that differential RA responsiveness within the HB relies on dynamic expression of *cyp26* genes (Abu-Abed *et al*., 2001; Emoto *et al*., 2005; Hernandez *et al*., 2007). Thus, a simple gradient of RA diffusing along the AP axis is probably not responsible for creating differential RA activities in the developing vertebrate HB (reviewed in Glover *et al*., 2006). In addition to the control of RA availability by CYP26 enzymes, RA signaling in the HB is also influenced by FGF signaling. For example, FGF8 originating from the MB antagonizes the action of RA and thus inhibits RA-dependent activation of *hox* genes, which is crucial for defining the anterior limits of *hox* gene expression in the HB (Irving and Mason, 2000).

In the amphioxus HB, collinear expression of *hox* genes is also regulated by RA. Exogenous RA shifts *hox* expression anteriorly and RA antagonist pushes *hox* expression posteriorly (Schubert *et al*., 2006b). Moreover, the expression domain of *otx*, which marks the amphioxus FB, MB and anterior HB, is reduced by RA treatment, while RA antagonist treatment expands the *otx* domain further into the amphioxus HB. Thus, RA signaling in the amphioxus HB most likely limits expression of *otx* to the amphioxus FB, MB and anterior HB. This particular function of RA signaling in amphioxus is probably directly mediated by *hox* genes, such as *hox1*, since functional RAREs have been described in the *cis*-regulatory regions of both amphioxus *hox1* and *hox3* (Manzanares *et al*., 2000; Wada *et al*., 2006) and knockdown of *hox1* has the same effect as treatment with an RA antagonist (Schubert *et al*., 2006b). In contrast, tunicates (e.g. ascidians and appendicularians) have lost several *hox* genes and both cluster organization and collinear expression have at least been partially lost (Ikuta *et al*., 2004; Seo *et al*., 2004; reviewed in Cañestro *et al*., 2007; Deschamps, 2007). Similarly, regulation of *hox* genes by RA signaling in tunicates is derived. For example, the conserved RARE in the *cis*-regulatory regions of *hox1* genes has been lost in *C. intestinalis* (Wada *et al*., 2006). Even so, treatments with RA affect the development of the anterior CNS in *Phallusia mammilata* (De Bernardi *et al*., 1994) and the closure of the neural tube in *C. intestinalis* (Nagatomo *et al*., 2003). Exogenous RA also activates and shifts anteriorly the expression of *hox1* in *Halocynthia roretzi* and *C. intestinalis* (Katsuyama *et al*., 1995; Katsuyama and Saiga, 1998; Nagatomo and Fujiwara, 2003). Taken together, these data suggest that a role for RA-dependent *hox* codes in specifying AP identities along the anterior CNS was probably present at the base of the chordates. It is very likely that, in the tunicate lineage, this patterning mechanism has at least partially been lost (reviewed in Hinman and Degnan, 2001; Fujiwara, 2006). To further address these hypotheses, future studies should focus on the roles of RA in the development of the nervous systems of other invertebrates, such as hemichordates, mollusks and annelids, with particular focus on the connection between RA signaling and collinear expression of *hox* genes in establishing positional identities along the AP body axis.

In addition to AP patterning, RA signaling in vertebrates is also involved in dorsoventral (DV) patterning of the posterior CNS (i.e. the SC) as well as in neurogenesis and neuronal specification within the developing CNS (reviewed in Glover *et al.*, 2006) (Fig. 3). In the SC, RA from somitic mesoderm specifies ventral regions by activating expression of DV patterning genes (e.g. *pax6, nkx6.1, pax3*)

(Wilson *et al*., 2004). This RA-dependent activation is either direct through induction of target genes or indirect by disrupting transcriptional repression mediated by FGF signaling (Diez del Corral *et al*., 2003; Novitch *et al*., 2003). Moreover, RA signaling induces neurogenesis in the developing SC by inhibiting the antineurogenic FGF8 signal and by activating expression of proneural genes like *NeuroM* (Diez del Corral *et al.*, 2003). RA also promotes primary neurogenesis in frogs, such as *Xenopus laevis* (Papalopulu and Kintner, 1996; Sharpe and Goldstone, 1997). Thus, RA treatment of frog embryos increases the number of primary neurons in posterior neuroectoderm and spreads them anteriorly, while reducing RA signaling activity has the opposite effect (Sharpe and Goldstone, 2000; reviewed in Maden, 2002). Finally, RA also acts on differentiation and specification of several neuronal populations, such as motoneurons, in the developing vertebrate HB and SC. For example, RA from paraxial mesoderm is required in the SC for the development of motoneuron progenitors (Novitch *et al*., 2003), while RA signals emanating directly from differentiating SC neurons are required for specification of a motoneuron fate (Sockanathan *et al.*, 2003). In the developing HB, RA signaling seems to be required for the specification of particular motoneuron subtypes. RA treatment reduces the number of HB branchial and visceral motoneurons by respecifying them as somatic motoneurons. Inhibition of CYP26 and grafting of RA-soaked beads under the rostral HB similarly leads to the formation of somatic motoneurons in the anterior part of HB explants (Guidato *et al.*, 2003a, b).

Involvement of RA signaling in motoneuron differentiation has also been studied in the cephalochordate amphioxus. RA treatment reduces the number of dorsal compartment motoneurons in the amphioxus MB and HB homolog. Conversely, inhibition of RA signaling, as well as disruption of *hox1* function, increases the number of motoneurons (Schubert *et al*., 2006b). These effects are comparable to those on somatic motoneurons in the vertebrate HB. Unfortunately, no other studies have addressed the roles of RA signaling in neuronal specification in amphioxus, tunicates or any other invertebrates. It is thus difficult to propose a scenario for the evolutionary origins of this role of RA signaling in development.

In addition to influencing neuronal differentiation and specification, RA also promotes neurite outgrowth in developing vertebrates (Fig. 3). For example, vitamin A deficient rat embryos exhibit defective axonal projections (White *et al*., 2000) and vitamin A deficient quail embryos are characterized by both defective axonal projections and impaired motoneuron outgrowth (Maden *et al.*, 1996; Wilson *et al*., 2003). Moreover, RA has been shown to activate neurite outgrowth in cultures of various chicken and mouse cell types originating either from the developing CNS (Wuarin *et al*., 1990; Corcoran *et al*., 2000) or from the developing peripheral nervous system (Haskell *et al*., 1987; Quinn and De Boni, 1991; Plum and Clagett-Dame, 1996; Corcoran *et al*., 2000). Neuronal cell populations of invertebrates seem to exhibit a similar response to RA: for example, treatments with RA induce neurite outgrowth and growth cone turning in neurons isolated of the mollusk *Lymnaea stagnalis*. Moreover, exogenous RA also enhances the viability of the mollusk neurons in culture (Dmetrichuk *et al*., 2006), which might be mediated by RAR, since our *in silico* analysis identified a putative *rar* homolog in the genome of *L. gigantea*, which belongs to a different mollusk subclass (Fig. 2). If the action of RA on neurite outgrowth and survival in mollusks is indeed mediated by a RAR/RXR heterodimer, this function of RA signaling could well be conserved between vertebrates and lophotrochozoans. This role of RA in neuronal outgrowth and cell survival might thus

represent an ancestral function of RA signaling during development that was already present in the last common ancestor of lophotrochozoans and deuterostomes.

Neural crest, placodes and their derivatives

Many evolutionary novelties of vertebrates arise from two cell populations: NCCs and placodes. Placodes are neurogenic thickenings of the ectoderm, while NCCs originate at the dorsal neural tube, delaminate and migrate through the embryo, carrying the RA-dependent *hox* code into other tissue layers. Cranial cartilage, bone and connective tissue are among the vertebrate-specific structures derived from NCCs and placodes. RA signaling has different functions in NCCs and placodes. For example, in the zebrafish mutant *neckless* (deficient for *raldh2*), NCCs undergo abnormal apoptosis leading to cartilage defects in larval fish (Begemann *et al*., 2001). Thus, RA signaling probably controls cell survival of NCCs (Fig. 3). Moreover, exogenous RA triggers abnormal migration of NCCs from rhombomere 4 into the first branchial arch, which results in repatterning of the cartilage there (Plant *et al*., 2000), probably due to alterations in the *hox* code of the NCCs. In vertebrates, RA signaling also acts on NCCs that contribute to the formation of placode-derived structures. For example, eye morphogenesis and retinal differentiation are controlled by RA (McCaffery *et al*., 1999), which seems to act primarily on the NCC-derived periocular mesenchyme. In this periocular mesenchyme, RA activates two specific RAR/RXR heterodimers (RARβ/RXRα and RARγ/RXRα) that control expression of genes important for optic development (e.g. *foxc1* and *pitx2*) (Matt *et al*., 2005). RA signaling is also implicated in the development of other placodes, such as the olfactory (Anchan *et al*., 1997; Mic *et al*., 2000), the otic (Dupé *et al*., 1999; White *et al*., 2000; Romand *et al*., 2006) and the lateral line placodes (Gibbs and Northcutt, 2004) (Fig. 3).

Although definitive NCCs and placodes did not appear before the emergence of vertebrates, possible evolutionary precursors of these tissues may be present in some invertebrates (Holland and Holland, 1996; Jeffery *et al*., 2004; reviewed in Schlosser, 2005; Marlétaz *et al*., 2006; Schubert *et al*., 2006a; Bassham *et al*., 2008). It is thus interesting to assess whether RA signaling affects the development of these invertebrate tissues. Such tissues include, for example, the adhesive papillae (or palps) in the anteriormost general ectoderm of ascidians, which is thought to be homologous either to the vertebrate olfactory placode (Mazet *et al*., 2005) or to vertebrate cement and hatching glands (Manni *et al*., 2004). However, the vertebrate cement and hatching glands might not be placodes in a strict sense, because they do not share a common developmental origin with placodes (reviewed in Schlosser, 2005).

In several ascidian species, RA treatment leads to a complete loss of anterior adhesive papillae (Denucé, 1991; De Bernardi *et al*., 1994; Katsuyama *et al*., 1995; Hinman and Degnan, 1998; Katsuyama and Saiga, 1998; Yagi and Makabe, 2002; Nagatomo *et al*., 2003), which is accompanied by upregulation of *hox1* in the general ectoderm including the region destined to give rise to the adhesive organ (Katsuyama *et al*., 1995; Katsuyama and Saiga, 1998). Intriguingly, exogenous RA leads to a loss of both the cement and the hatching gland in frogs (Sive *et al*., 1990; Drysdale and Elinson, 1991), which, as in ascidians, is accompanied by upregulation of *hox1* expression in the ectoderm. In frogs, RA treatment at the mid-blastula results in ubiquitous ectodermal expression of *hox1* genes during gastrulation (Kolm and Sive, 1995).

In amphioxus, the other group of invertebrate chordates, RA signaling in the general ectoderm controls collinear expression of *hox* genes as well as specification of neuronal subtypes (Schubert *et al.*, 2004) (Fig. 3). Thus, while exogenous RA shifts collinear expression of *hox* genes anteriorly, RA antagonist treatment downregulates transcription of *hox* genes in the general ectoderm. Moreover, Schubert *et al.* (2004) demonstrated that in the amphioxus general ectoderm the differentiation of sensory neurons is sensitive to RA signals. During normal development, these cells are scattered along the entire AP axis of the developing general ectoderm and express a combinatorial code of different proneural genes including *coe*, *islet* and *err*. Treatments with RA and RA antagonist shift the AP distribution of sensory neurons in the general ectoderm and alter the combinatorial code of proneural genes expressed in these cells (Schubert *et al*., 2004). The effect of RA on the distribution of sensory neurons in the amphioxus general ectoderm is reminiscent of the posteriorizing action of exogenous RA on lateral line neurons of axolotls (Gibbs and Northcutt, 2004).

Concluding Remarks

In this review, we have presented our current understanding of the structure and functions of the RA signaling during animal development. Moreover, we have shown that RA signaling is neither vertebrate- nor chordate-specific and we have presented data suggesting that RA signaling may not even be deuterostome-specific, because the main players of RA signaling, in particular *rar*, are probably present in the genomes of both annelids and mollusks. Since some effects of exogenous RA have already been described in lophotrochozoans, especially on neurite outgrowth and neuronal survival in mollusks (Dmetrichuk *et al*., 2006), it is conceivable that RA signaling mediated by RAR/RXR heterodimers is functional in lophotrochozoans. Although an experimental verification of this hypothesis is still lacking, these data suggest that the origin of RA signaling should probably be pushed back to the last common ancestor of all bilaterian animals (Urbilateria), where RA signaling might have played a role in neurite outgrowth and neuronal survival during development (Fig. 4). After the divergence of ecdysozoans, lophotrochozoans and deuterostomes, RA signaling has evolved very differently in each of the three groups. Hence, while RA signaling was apparently lost in ecdysozoans, it persisted in the two other clades. In deuterostomes, very little is known about RA signaling in ambulacrarians (echinoderms and hemichordates). In the chordate lineage, novel functions for RA signaling probably arose in neuronal specification in the general ectoderm and the CNS. Moreover, RA has evolved a fundamental role in AP patterning of the chordate endoderm, general ectoderm and CNS. In all three tissue layers, these functions of RA are mediated by *hox* genes suggesting that roles for RA signaling in AP patterning and in regulation of the *hox* code might have evolved simultaneously. In the lineage leading to vertebrates and tunicates (e.g. ascidians and appendicularians), RA signaling has acquired roles in controlling NCC survival and placode development. However, the exact timing of these events remains elusive. Within tunicates, the sister group of vertebrates, RA signaling is present and functional in ascidian embryos, while the RA machinery has been lost in appendicularians. Further functions for RA signaling during development arose in the vertebrate lineage, particularly in mesodermal (AP patterning, somitogenesis, left/right asymmetry) and neural tissues (neurogenesis, DV patterning of the SC).

This putative evolutionary history of the RA signaling cascade suggests at least a few important directions for future investigations. Due to its presence and roles in development of a wide range of animals, RA signaling could probably be regarded as one of the core regulatory networks controlling animal development. However, the independent loss of the RA machinery in two different lineages (i.e. appendicularians and ecdysozoans) shows that embryos can also be patterned in the absence of RA signaling (Cañestro and Postlethwait, 2007). Moreover, the evolution of vertebrates is characterized by the acquisition of a number of different RA-dependent functions (Fig. 4). This elaboration of novel roles for RA during development is probably correlated with the massive gene duplications that characterize the evolution of vertebrates (for example, while there is only one *rar* and one *rxr* in the cephalochordate amphioxus, tetrapods have three copies of each gene) (Ohno, 1970). Although some studies have already addressed the relationship between duplication of genes involved in RA signaling and acquisition of new functions during development (Escrivà *et al*., 2006), further work is needed to decipher this interplay between gene number and acquisition of novel RA-dependent functions in vertebrates.

Finally, most studies on roles of RA signaling during development have focused on vertebrate model organisms. Although there are now some data on developmental roles of RA signaling emerging from invertebrate chordates, there is still an obvious lack of information on the role(s) of RA in non-chordates. Even if low concentrations of 9-*cis* and all-*trans* RA have been observed in regenerating limb blastemas of the crab *Uca pugilator* (Hopkins, 2001) and effects of treatments with RA agonists or antagonists have been described in sea urchins (Sciarrino and Matranga, 1995; Kuno *et al*., 1999), mollusks (Créton *et al*., 1993) crustaceans (Chung *et al*., 1998; Hopkins, 2001; Söderhäll *et al*., 2006), insects (Sun *et al*., 1993; Picking *et al*., 1996; Shim *et al*., 1997), planarians (Romero and Bueno, 2001), cnidarians (Müller, 1984) and sponges (Imsiecke *et al*., 1994; Nikko *et al*., 2001; Wiens *et al*., 2003), the actual presence and putative roles for RA signaling during development in these taxa remain elusive. Thus, to fully understand the origin and evolution of RA signaling during embryonic development, we need to broaden the sampling of animal taxa and apply more sophisticated experimental tools to non-vertebrate model systems.

Acknowledgements

The authors would like to thank Linda Z. Holland, Nicholas D. Holland, Gérard Benoit, Maria Theodosiou and Mathilde Paris for fruitful discussions, helpful comments and critical reading of the manuscript.

Table 1. Conservation of residues contacting retinoic acid (RA) in the ligand binding pocket of retinoic acid receptors (RARs) from different metazoan animals. The numbering of amino acid residues is according to Renaud *et al.* (1995). Amino acids identical to those of human RARγ are shown in red and residues directly interacting with the carboxylate moiety of the RA molecule (Lys236, Arg278 and Ser289) are highlighted in blue.

Human RARγ	Phe 230	Leu 233	Lys 236	Cys 237	Leu 271	Met 272	Arg 274	Ile 275	Arg 278	Phe 288	Ser 289	Phe 304	Gly 393	Leu 400
Human RARα	Phe	Leu	Lys	Cys	Leu	Ile	Arg	Ile	Arg	Phe	Ser	Phe	Gly	Leu
Human RARß	Phe	Leu	Lys	Cys	Leu	Ile	Arg	Ile	Arg	Phe	Ser	Phe	Gly	Leu
C. intestinalis RAR	Phe	Leu	Lys	Cys	Leu	Phe	Arg	Ile	Arg	Phe	Ser	Phe	Gly	Val
Amphioxus RAR	Phe	Leu	Glu	Thr	Leu	Ile	Arg	Leu	Arg	Phe	Ile	Phe	Gly	Leu
Sea urchin RAR	Val	Met	Arg	Ala	Met	Ile	Arg	Ile	Arg	Phe	Thr	Phe	-	-
L. gigantea RAR	Ile	Leu	Gly	Gly	Met	Ile	Arg	Leu	Arg	Phe	Ser	Phe	Gly	Leu
C. capitata RAR	Val	Leu	Ser	Gly	Leu	Val	Arg	Leu	Arg	Phe	Thr	Phe	Gly	Leu

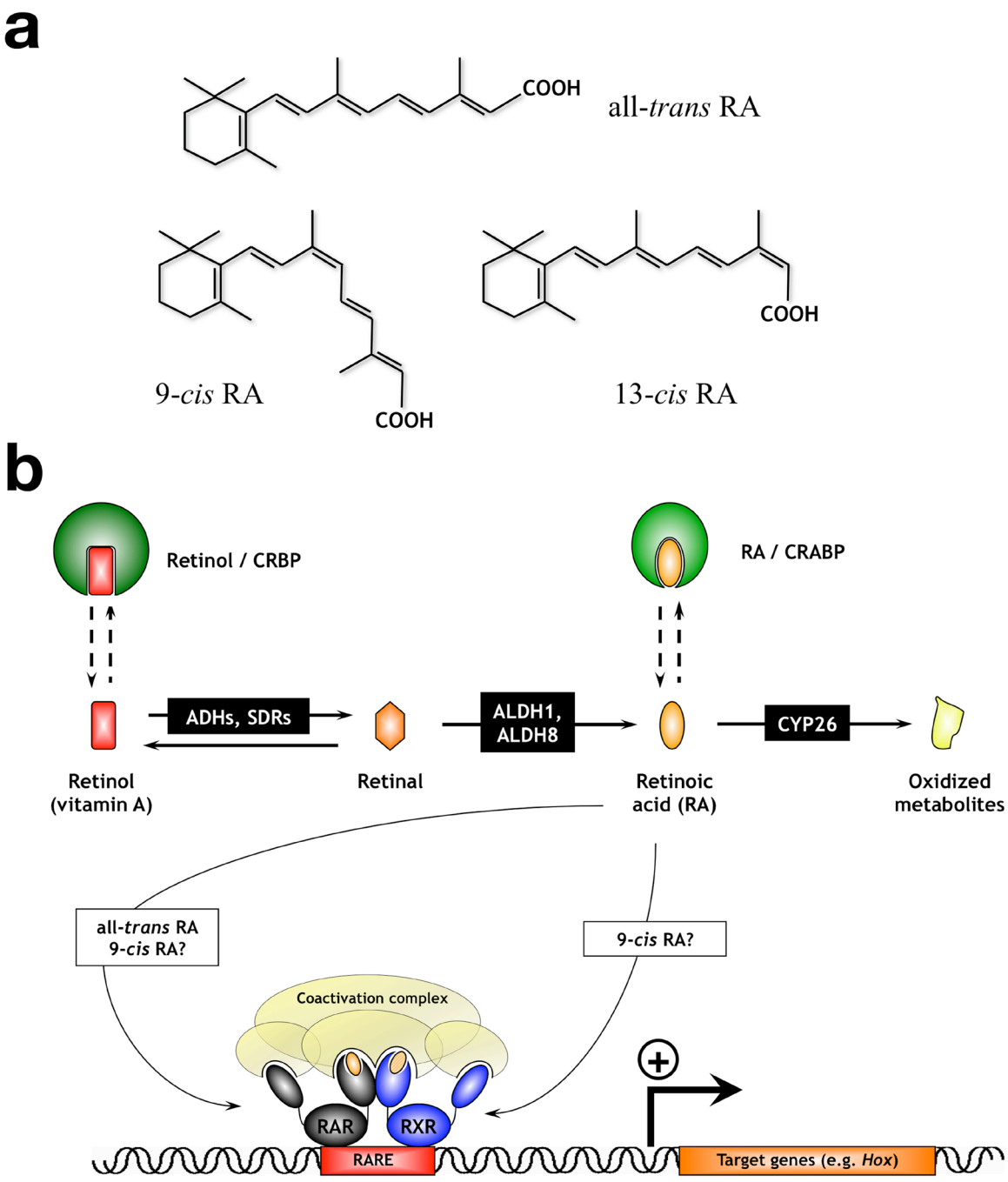


Fig. 1. Retinoic acid (RA) and RA signaling. (**a**) Molecular structure of three different stereoisomeric forms of RA. (**b**) Overview of the RA signaling cascade. Only the major actors of the pathway (enzymes involved in retinoid synthesis, degradation, binding and signaling) are shown. Enzymes implicated in RA synthesis and degradation are shown as black boxes. ADH: alcohol dehydrogenase; ALDH: aldehyde dehydrogenase; CYP26: Cytochrome P450, subfamily CYP26; CRBP: cellular retinol binding protein; CRABP: cellular retinoic acid binding protein; RA: retinoic acid; RAR: retinoic acid receptor; RARE: retinoic acid response element; RXR: retinoid X receptor; SDR: short-chain dehydrogenase/reductase.

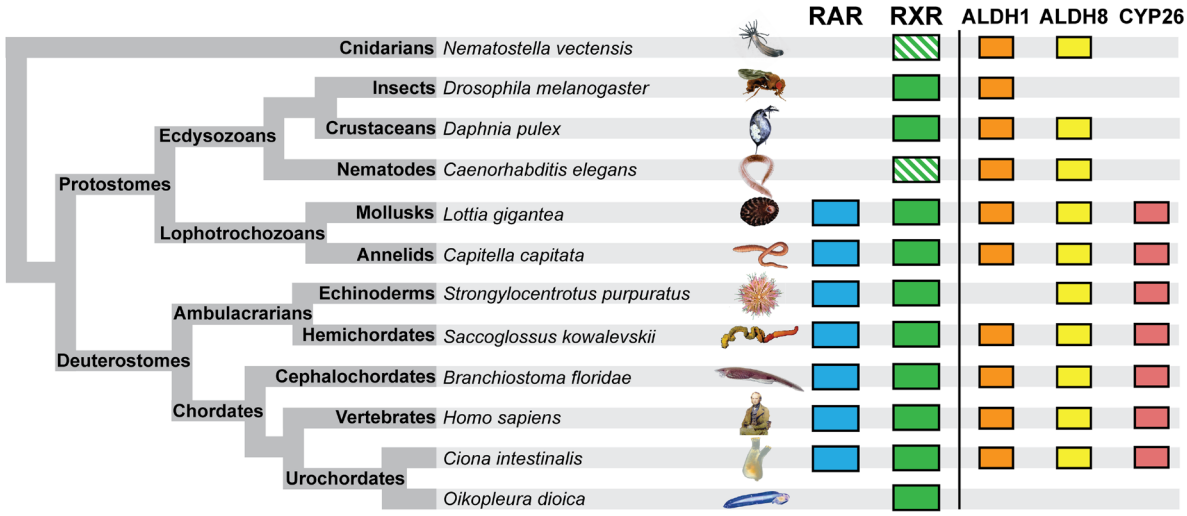
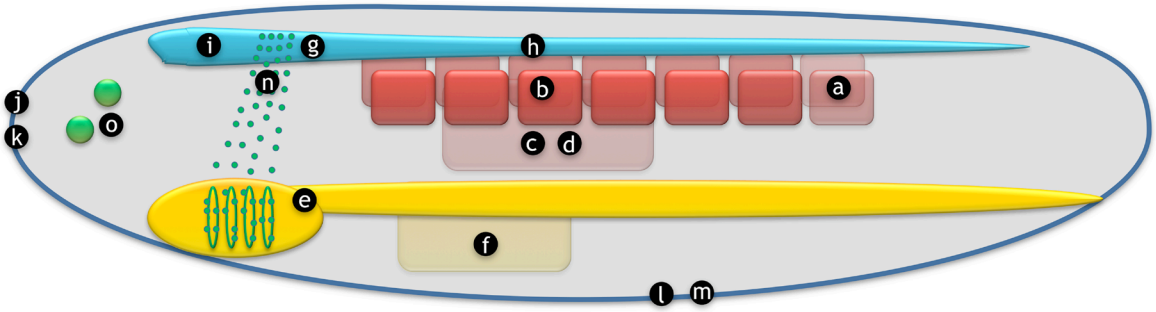


Fig. 2. Members of the retinoic acid (RA) signaling cascade in different metazoans. The presence of assayed genes (*rar*, *rxr*, *aldh1*, *aldh8*, *cyp26*) is shown with colored boxes. The information given in this figure is based on BLAST search results, which for *rar* and *rxr* have been verified by phylogenetic analyses. Hatched boxes indicate situations where a given gene is absent from the genome of an assayed species, but present in other species of the same phylum. ALDH: aldehyde dehydrogenase; CYP26: Cytochrome P450, subfamily CYP26; RAR: retinoic acid receptor; RXR: retinoid X receptor.



	Vertebrates	Urochordates	Amphioxus
Mesoderm	a Somitogenesis b AP patterning c Left/right asymmetry d Organogenesis	d Organogenesis	
Endoderm	e AP patterning and pharyngeal development f Organogenesis	e AP patterning and pharyngeal development	e AP patterning and pharyngeal development
Central nervous system	g AP patterning h Neuronal specification and DV patterning i Neurite outgrowth	g AP patterning?	g AP patterning h Neuronal specification
General ectoderm	j Cement and hatching gland morphogenesis	k Palp morphogenesis	l AP patterning m Neuronal specification
Neural crest and placodes	n Neural crest survival o Placode development		

Fig. 3. Summary of the roles of retinoic acid (RA) signaling during chordate development. The figure shows a diagrammatic representation of a chordate embryo with letters referring to tissue-specific functions of RA in different chordate phyla. These different roles for RA during chordate development are summarized in the table. In both diagram and table, the different tissues are color coded: red for mesoderm, yellow for endoderm, light blue for central nervous system, dark blue for general ectoderm, green for neural crest and placodes. AP: anteroposterior; DV: dorsoventral.

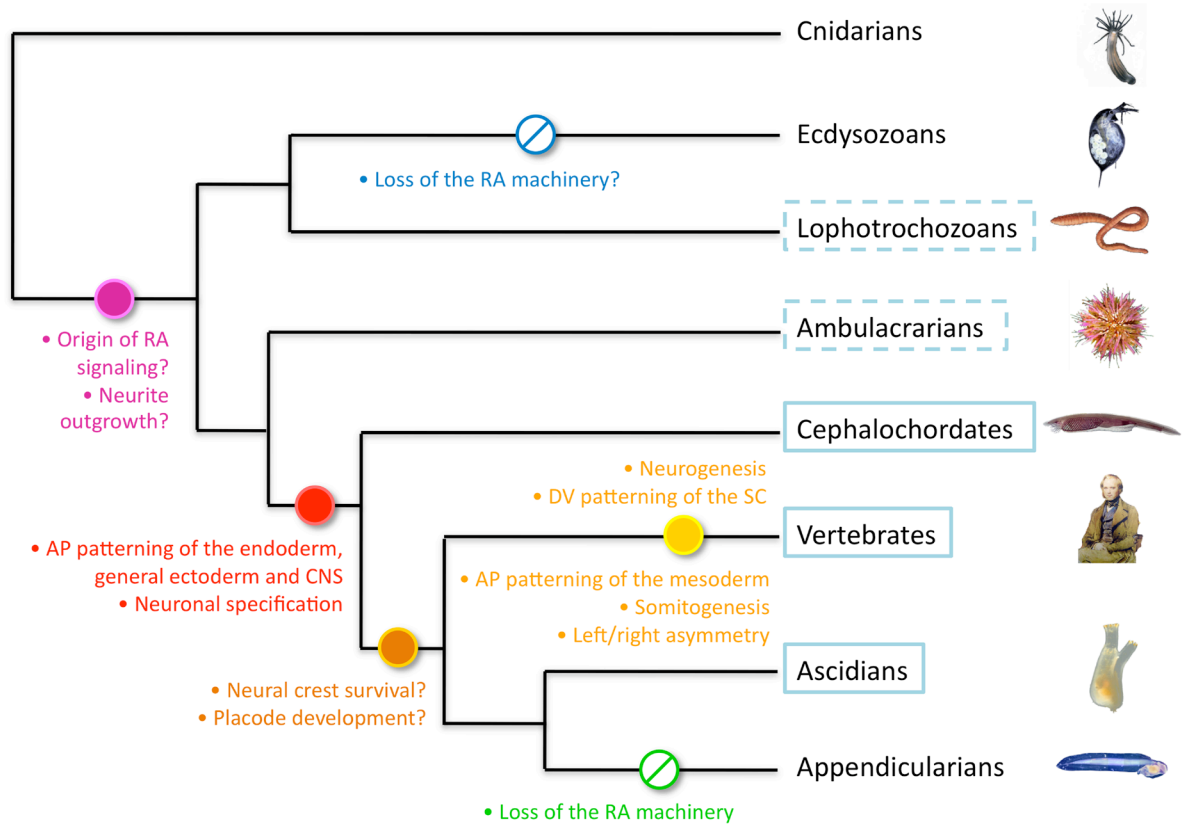


Fig. 4. Origin and evolution of the roles of retinoic acid (RA) signaling in metazoan development. Acquisition of new roles of RA signaling in a lineage is marked with filled circles. Boxes indicate phyla where presence of RA signaling has already been demonstrated (continuous line) or still awaits experimental verification (dashed line). "/" indicates absence or loss of RA signaling in a given lineage. AP: anteroposterior; CNS: central nervous system; DV: dorsoventral; RA: retinoic acid; SC: spinal cord.

References

Abu-Abed S, Dollé P, Metzger D, Beckett B, Chambon P, Petkovich M. 2001. The retinoic acid-metabolizing enzyme, CYP26A1, is essential for normal hindbrain patterning, vertebral identity, and development of posterior structures. Genes Dev 15:226-240.

Abu-Abed S, MacLean G, Fraulob V, Chambon P, Petkovich M, Dollé P. 2002. Differential expression of the retinoic acid-metabolizing enzymes CYP26A1 and CYP26B1 during murine organogenesis. Mech Dev 110:173-177.

Allan D, Houle M, Bouchard N, Meyer BI, Gruss P, Lohnes D. 2001. RARγ and Cdx1 interactions in vertebral patterning. Dev Biol 240:46-60.

Anchan RM, Drake DP, Haines CF, Gerwe EA, LaMantia AS. 1997. Disruption of local retinoid-mediated gene expression accompanies abnormal development in the mammalian olfactory pathway. J Comp Neurol 379:171-184.

Balmer JE, Blomhoff R. Gene expression regulation by retinoic acid. 2002. J Lipid Res 43:1773-1808.

Balmer JE, Blomhoff R. A robust characterization of retinoic acid response elements based on a comparison of sites in three species. 2005. J Steroid Biochem Mol Biol 96:347-354.

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.

Begemann G, Schilling TF, Rauch GJ, Geisler R, Ingham PW. 2001. The zebrafish *neckless* mutation reveals a requirement for raldh2 in mesodermal signals that pattern the hindbrain. Development 128:3081-3094.

Berggren K, McCaffery P, Drager U, Forehand CJ. 1999. Differential distribution of retinoic acid synthesis in the chicken embryo as determined by immunolocalization of the retinoic acid synthetic enzyme, RALDH-2. Dev Biol 210:288-304.

Bertrand S, Thisse B, Tavares R, Sachs L, Chaumot A, Bardet PL, Escrivà H, Duffraisse M, Marchand O, Safi R, Thisse C, Laudet V. 2007. Unexpected Novel Relational Links Uncovered by Extensive Developmental Profiling of Nuclear Receptor Expression. PloS Genet 3:e188.

Blentic A, Gale E, Maden M. 2003. Retinoic acid signalling centres in the avian embryo identified by sites of expression of synthesising and catabolising enzymes. Dev Dyn 227:114-127.

Blomhoff R, Blomhoff HK. 2006. Overview of retinoid metabolism and function. J Neurobiol 66:606-630.

Boleda MD, Saubi N, Farres J, Pares X. 1993. Physiological substrates for rat alcohol dehydrogenase classes: aldehydes of lipid peroxidation, ω-hydroxyfatty acids, and retinoids. Arch Biochem Biophys 307:85-90.

Bourlat SJ, Juliusdottir T, Lowe CJ, Freeman R, Aronowicz J, Kirschner M, Lander ES, Thorndyke M, Nakano H, Kohn AB, Heyland A, Moroz LL, Copley RR, Telford MJ. 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature 444:85-88.

Breitman TR, Takahashi N. 1996. Retinoylation of proteins in mammalian cells. Biochem Soc Trans 24:723-727.

Burrow CR. 2000. Regulatory molecules in kidney development. Pediatr Nephrol 14:240-253.

Calleja C, Messaddeq N, Chapellier B, Yang H, Krezel W, Li M, Metzger D, Mascrez B, Ohta K, Kagechika H, Endo Y, Mark M, Ghyselinck NB, Chambon P. 2006. Genetic and pharmacological evidence that a retinoic acid cannot be the RXR-activating ligand in mouse epidermis keratinocytes. Genes Dev 20:1525-1538.

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, 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.

Chambers D, Wilson L, Maden M, Lumsden A. 2007. RALDH-independent generation of retinoic acid during vertebrate embryogenesis by CYP1B1. Development 134:1369-1383.

Chambon P. 1996. A decade of molecular biology of retinoic acid receptors. FASEB J 10:940-954.

Chazaud C, Chambon P, Dollé P. 1999. Retinoic acid is required in the mouse embryo for left-right asymmetry determination and heart morphogenesis. Development 126:2589-2596.

Chea HK, Wright CV, Swalla BJ. 2005. Nodal signaling and the evolution of deuterostome gastrulation. Dev Dyn 234: 269-278.

Chen Y, Huang L, Solursh M. 1994. A concentration gradient of retinoids in the early *Xenopus laevis* embryo. Dev Biol 161:70-76.

Chung AC, Durica DS, Clifton SW, Roe BA, Hopkins PM. 1998. Cloning of crustacean ecdysteroid receptor and retinoid-X receptor gene homologs and elevation of retinoid-X receptor mRNA by retinoic acid. Mol Cell Endocrinol 139:209-227.

Chuong CM. 1993. The making of a feather: homeoproteins, retinoids and adhesion molecules. Bioessays 15:513-521.

Corcoran J, Shroot B, Pizzey J, Maden M. 2000. The role of retinoic acid receptors in neurite outgrowth from different populations of embryonic mouse dorsal root ganglia. J Cell Sci 113:2567-2574.

Créton R, Zivkovic D, Zwaan G, Dohmen R. 1993. Polar ionic currents around embryos of *Lymnaea stagnalis* during gastrulation and organogenesis. Int J Dev Biol 37:425-431.

De Bernardi F, Sotgia C, Ortolani G. 1994. Retinoic acid treatment of ascidian embryos: effects on larvae and metamorphosis. Anim Biol 3:75-81.

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

Delva L, Bastie JN, Rochette-Egly C, Kraïba R, Balitrand N, Despouy G, Chambon P, Chomienne C. 1999. Physical and functional interactions between cellular retinoic acid binding protein II and the retinoic acid-dependent nuclear complex. Mol Cell Biol 19:7158-7167.

Denucé JM. 1991. Teratogenic and metamorphosis inhibiting activity of retinoic acid in Ciona intestinalis. Z Naturforsch 46:1094-1100.

Deschamps J. 2007. Ancestral and recently recruited global control of the Hox genes in development. Curr Opin Genet Dev 17:422-427.

Diez del Corral R, Storey KG. 2004. Opposing FGF and retinoid pathways: a signalling switch that controls differentiation and patterning onset in the extending vertebrate body axis. Bioessays 26:857-869.

Diez del Corral R, Olivera-Martinez I, Goriely A, Gale E, Maden M, 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 JM, Carlone RL, Spencer GE. 2006. Retinoic acid induces neurite outgrowth and growth cone turning in invertebrate neurons. Dev Biol 294:39-49.

Dollé P, Ruberte E, Leroy P, Morriss-Kay G, Chambon P. 1990. Retinoic acid receptors and cellular retinoid binding proteins. I. A systematic study of their differential pattern of transcription during mouse organogenesis. Development 110:1133-1151.

Donovan M, Olofsson B, Gustafson AL, Dencker L, Eriksson U. 1995. The cellular retinoic acid binding proteins. J Steroid Biochem Mol Biol 53:459-465.

Drysdale TA, Elinson RP. 1991. Development of the *Xenopus laevis* hatching gland and its relationship to surface ectoderm patterning. Development 111:469-478.

Duboc V, Röttinger E, Besnardeau L, Lepage T. 2004. Nodal and BMP2/4 signaling organizes the oral-aboral axis of the sea urchin embryo. Dev Cell 6:397-410.

Duboc V, Röttinger E, Lapraz F, Besnardeau L, Lepage T. 2005. Left-right asymmetry in the sea urchin embryo is regulated by nodal signaling on the right side. Dev Cell 9:147-158.

Dubrulle J, Pourquié O. 2004. Coupling segmentation to axis formation. Development 131:5783-5793.

Dubrulle J, McGrew MJ, Pourquié O. 2001. FGF signaling controls somite boundary position and regulates segmentation clock control of spatiotemporal Hox gene activation. Cell 106:219-232.

Duester G. 1996. Involvement of alcohol dehydrogenase, short-chain dehydrogenase/reductase, aldehyde dehydrogenase, and cytochrome P450 in the control of retinoid signaling by activation of retinoic acid synthesis. Biochemistry 35:12221-12227.

Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, Sørensen MV, Haddock SH, Schmidt-Rhaesa A, Okusu A, Kristensen RM, Wheeler WC, Martindale MQ, Giribet G. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452:745-749.

Dupé V, Lumsden A. 2001. Hindbrain patterning involves graded responses to retinoic acid signalling. Development 128:2199-2208.

Dupé V, Ghyselinck NB, Wendling O, Chambon P, Mark M. 1999. Key roles of retinoic acid receptors alpha and beta in the patterning of the caudal hindbrain, pharyngeal arches and otocyst in the mouse. Development 126:5051-5059.

Durston AJ, Timmermans JP, Hage WJ, Hendriks HF, de Vries NJ, Heideveld M, Nieuwkoop PD. 1989. Retinoic acid causes an anteroposterior transformation in the developing central nervous system. Nature 340:140-144.

Emoto Y, Wada H, Okamoto H, Kudo A, Imai Y. 2005. Retinoic acid-metabolizing enzyme Cyp26a1 is essential for determining territories of hindbrain and spinal cord in zebrafish. Dev Biol 278:415-427.

Escrivà H, Holland ND, Gronemeyer H, Laudet V, Holland LZ. 2002. The retinoic acid signaling pathway regulates anterior/posterior patterning in the nerve cord and

pharynx of amphioxus, a chordate lacking neural crest. Development 129:2905-2916.
Escrivà H, Bertrand S, Germain P, Robinson-Rechavi M, Umbhauer M, Cartry J, Duffraisse M, Holland L, Gronemeyer H, Laudet V. 2006. Neofunctionalization in vertebrates: the example of retinoic acid receptors. PloS Genet 2:e102.
Fujiwara S. 2006. Retinoids and nonvertebrate chordate development. J Neurobiol 66:645-652.
Gale E, Zile M, Maden M. 1999. Hindbrain respecification in the retinoid-deficient quail. Mech Dev 89:43-54.
Gibbs MA, Northcutt RG. 2004. Retinoic acid repatterns axolotl lateral line receptors. Int J Dev Biol 48:63-66.
Glover JC, Renaud JS, Rijli FM. 2006. Retinoic acid and hindbrain patterning. J Neurobiol 66:705-725.
Godsave SF, Koster CH, Getahun A, Mathu M, Hooiveld M, van der Wees J, Hendriks J, Durston AJ. 1998. Graded retinoid responses in the developing hindbrain. Dev Dyn 213:39-49.
Graham A, Smith A. 2001. Patterning the pharyngeal arches. Bioessays 23:54-61.
Gronemeyer H, Gustafsson JA, Laudet V. 2004. Principles for modulation of the nuclear receptor superfamily. Nat Rev Drug Discov 3:950-964.
Guidato S, Barrett C, Guthrie S. 2003a. Patterning of motor neurons by retinoic acid in the chick embryo hindbrain in vitro. Mol Cell Neurosci 23:81-95.
Guidato S, Prin F, Guthrie S. 2003b. Somatic motoneurone specification in the hindbrain: the influence of somite-derived signals, retinoic acid and Hoxa3. Development 130:2981-2996.
Hall BK. 2000. The neural crest as a fourth germ layer and vertebrates as quadroblastic not triploblastic. Evol Dev 1:3-5.
Hamade A, Deries M, Begemann G, Bally-Cuif L, Genet C, Sabatier F, Bonnieu A, Cousin X. 2006. Retinoic acid activates myogenesis in vivo through Fgf8 signalling. Dev Biol 289:127-140.
Haskell BE, Stach RW, Werrbach-Perez K, Perez-Polo JR. 1987. Effect of retinoic acid on nerve growth factor receptors. Cell Tissue Res 247:67-73.
Hernandez RE, Putzke AP, Myers JP, Margaretha L, Moens CB. 2007. Cyp26 enzymes generate the retinoic acid response pattern necessary for hindbrain development. Development 134:177-187.
Hinman VF, Degnan BM. 1998. Retinoic acid disrupts anterior ectodermal and endodermal development in ascidian larvae and postlarvae. Dev Genes Evol 208:336-345.
Hinman VF, Degnan BM. 2000. Retinoic acid perturbs *Otx* gene expression in the ascidian pharynx. Dev Genes Evol 210:129-139.
Hochgreb T, Linhares VL, Menezes DC, Sampaio AC, Yan CY, Cardoso WV, Rosenthal N, Xavier-Neto J. 2003. A caudorostral wave of RALDH2 conveys anteroposterior information to the cardiac field. Development 130:5363-5374.
Holland LZ, Holland ND. 1996. Expression of *AmphiHox-1* and *AmphiPax-1* in amphioxus embryos treated with retinoic acid: insights into evolution and patterning of the chordate nerve cord and pharynx. Development 122:1829-1838.
Hopkins PM. 2001. Limb regeneration in the fiddler crab, *Uca pugilator*: hormonal and growth factor control. Amer Zool 41:389-398.

Houle M, Prinos P, Iulianella A, Bouchard N, Lohnes D. 2000. Retinoic acid regulation of Cdx1: an indirect mechanism for retinoids and vertebral specification. Mol Cell Biol 20:6579-6586.

Howard-Ashby M, Materna SC, Brown CT, Chen L, Cameron RA, Davidson EH. 2006. Gene families encoding transcription factors expressed in early development of *Strongylocentrotus purpuratus*. Dev Biol 300:90-107.

Ikuta T, Yoshida N, Satoh N, Saiga H. 2004. *Ciona intestinalis Hox* gene cluster: its dispersed structure and residual colinear expression in development. Proc Natl Acad Sci U S A 101:15118-15123.

Imsiecke G, Borojevic R, Custudio M, Müller WEG. 1994. Retinoic acid acts as a morphogen in freshwater sponges. Invert Reprod Dev 26:89-98.

Irving C, 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.

Jeffery WR, Strickler AG, Yamamoto Y. 2004. Migratory neural crest-like cells form body pigmentation in a urochordate embryo. Nature 431:696-699.

Katsuyama Y, Saiga H. 1998. Retinoic acid affects patterning along the anterior-posterior axis of the ascidian embryo. Dev Growth Differ 40:413-422.

Katsuyama Y, Wada S, Yasugi S, Saiga H. 1995. Expression of the *labial* group *Hox* gene *HrHox-1* and its alteration induced by retinoic acid in development of the ascidian *Halocynthia roretzi*. Development 121:3197-3205.

Kawaguchi R, Yu J, Honda J, Hu J, Whitelegge J, Ping P, Wiita P, Bok D, Sun H. 2007. A membrane receptor for retinol binding protein mediates cellular uptake of vitamin A. Science 315:820-825.

Kawakami Y, Raya A, Raya RM, Rodríguez-Esteban C, Belmonte JC. 2005. Retinoic acid signalling links left-right asymmetric patterning and bilaterally symmetric somitogenesis in the zebrafish embryo. Nature 435:165-171.

Kim CI, Leo MA, Lieber CS. 1992. Retinol forms retinoic acid via retinal. Arch Biochem Biophys 294:388-393.

Kligman AM. 1997. The treatment of acne with topical retinoids: one man's opinions. J Am Acad Dermatol 36:S92-S95.

Kolm PJ, Sive HL. 1995. Regulation of the *Xenopus* labial homeodomain genes, *HoxA1* and *HoxD1*: activation by retinoids and peptide growth factors. Dev Biol 167:34-49.

Kostrouch Z, Kostrouchova M, Love W, Jannini E, Piatigorsky J, Rall JE. 1998. Retinoic acid X receptor in the diploblast, Tripedalia cystophora. Proc Natl Acad Sci U S A 95:13442-13447.

Kumar M, Melton D. 2003. Pancreas specification: a budding question. Curr Opin Genet Dev 13:401-407.

Kumar M, Jordan N, Melton D, Grapin-Botton A. 2003. Signals from lateral plate mesoderm instruct endoderm toward a pancreatic fate. Dev Biol 259:109-122.

Kuno S, Kawamoto M, Okuyama M, Yasumasu I. 1999. Outgrowth of pseudopodial cables induced by all-*trans* retinoic acid in micromere-derived cells isolated from sea urchin embryos. Dev Growth Differ 41:193-199.

Kuratani S, Ueki T, Hirano S, Aizawa S. 1998. Rostral truncation of a cyclostome, *Lampetra japonica*, induced by all-*trans* retinoic acid defines the head/trunk interface of the vertebrate body. Dev Dyn 211:35-51.

Labrecque J, Dumas F, Lacroix A, Bhat PV. 1995. A novel isoenzyme of aldehyde dehydrogenase specifically involved in the biosynthesis of 9-*cis* and all-*trans* retinoic acid. Biochem J 305:681-684.

Lammer EJ, Chen DT, Hoar RM, Agnish ND, Benke PJ, Braun JT, Curry CJ, Fernhoff PM, Grix AW Jr, Lott IT, Richard JM, Sun SC. 1985. Retinoic acid embryopathy. N Engl J Med 313:837-841.

Lee MO, Manthey CL, Sladek NE. 1991. Identification of mouse liver aldehyde dehydrogenases that catalyze the oxidation of retinaldehyde to retinoic acid. Biochem Pharmacol 42:1279-1285.

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.

Lumsden A, Krumlauf R. 1996. Patterning the vertebrate neuraxis. Science 274:1109-1115.

MacLean G, Abu-Abed S, Dollé P, Tahayato A, Chambon P, Petkovich M. 2001. Cloning of a novel retinoic-acid metabolizing cytochrome P450, *Cyp26B1*, and comparative expression analysis with *Cyp26A1* during early murine development. Mech Dev 107:195-201.

Maden M. 2002. Retinoid signalling in the development of the central nervous system. Nat Rev Neurosci 3:843-853.

Maden M, Gale E, Kostetskii I, Zile M. 1996. Vitamin A-deficient quail embryos have half a hindbrain and other neural defects. Curr Biol 6:417-426.

Maden M, Sonneveld E, van der Saag PT, Gale E. 1998. The distribution of endogenous retinoic acid in the chick embryo: implications for developmental mechanisms. Development 125:4133-4144.

Manni L, Lane NJ, Joly JS, Gasparini F, Tiozzo S, Caicci F, Zaniolo G, Burighel P. 2004. Neurogenic and non-neurogenic placodes in ascidians. J Exp Zoolog B Mol Dev Evol 302:483-504.

Manzanares M, Wada H, Itasaki N, Trainor PA, Krumlauf R, Holland PW. 2000. Conservation and elaboration of Hox gene regulation during evolution of the vertebrate head. Nature 408:854-857.

Marill J, Cresteil T, Lanotte M, Chabot GG. 2000. Identification of human cytochrome P450s involved in the formation of all-*trans*-retinoic acid principal metabolites. Mol Pharmacol 58:1341-1348.

Mark M, Ghyselinck NB, Chambon P. 2004. Retinoic acid signalling in the development of branchial arches. Curr Opin Genet Dev 14:591-598.

Mark M, Ghyselinck NB, Chambon P. 2006. Function of retinoid nuclear receptors: lessons from genetic and pharmacological dissections of the retinoic acid signaling pathway during mouse embryogenesis. Annu Rev Pharmacol Toxicol 46:451-480.

Marlétaz F, Holland LZ, Laudet V, Schubert M. 2006. Retinoic acid signaling and the evolution of chordates. Int J Biol Sci 2:38-47.

Marshall H, Nonchev S, Sham MH, Muchamore I, Lumsden A, Krumlauf R. 1992. Retinoic acid alters hindbrain *Hox* code and induces transformation of rhombomeres 2/3 into a 4/5 identity. Nature 360:737-741.

Matt N, Ghyselinck NB, Wendling O, Chambon P, Mark M. 2003. Retinoic acid-induced developmental defects are mediated by RARβ/RXR heterodimers in the pharyngeal endoderm. Development 130:2083-2093.

Matt N, Dupé V, Garnier JM, Dennefeld C, Chambon P, Mark M, Ghyselinck NB. 2005. Retinoic acid-dependent eye morphogenesis is orchestrated by neural crest cells. Development 132:4789-4800.

Mazet F, Hutt JA, Milloz J, Millard J, Graham A, Shimeld SM. 2005. Molecular evidence from *Ciona intestinalis* for the evolutionary origin of vertebrate sensory placodes. Dev Biol 282:494-508.

McCaffery P, Wagner E, O'Neil J, Petkovich M, Drager UC. 1999. Dorsal and ventral retinal territories defined by retinoic acid synthesis, break-down and nuclear receptor expression. Mech Dev 82:119-130.

Mic FA, Molotkov A, Fan X, Cuenca AE, Duester G. 2000. RALDH3, a retinaldehyde dehydrogenase that generates retinoic acid, is expressed in the ventral retina, otic vesicle and olfactory pit during mouse development. Mech Dev 97:227-230.

Mic FA, Molotkov A, Benbrook DM, Duester G. 2003. Retinoid activation of retinoic acid receptor but not retinoid X receptor is sufficient to rescue lethal defect in retinoic acid synthesis. Proc Natl Acad Sci U S A 100:7135-7140.

Mollard R, Viville S, Ward SJ, Décimo D, Chambon P, Dollé P. 2000. Tissue-specific expression of retinoic acid receptor isoform transcripts in the mouse embryo. Mech Dev 94:223-232.

Moreno TA, Kintner C. 2004. Regulation of segmental patterning by retinoic acid signaling during *Xenopus* somitogenesis. Dev Cell 6:205-218.

Müller WA. 1984. Retinoids and pattern formation in a hydroid. J Embryol Exp Morphol 81:253-271.

Murakami Y, Pasqualetti M, Takio Y, Hirano S, Rijli FM, Kuratani S. 2004. Segmental development of reticulospinal and branchiomotor neurons in lamprey: insights into the evolution of the vertebrate hindbrain. Development 131:983-995.

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.

Nagatomo K, Ishibashi T, Satou Y, Satoh N, Fujiwara S. 2003. Retinoic acid affects gene expression and morphogenesis without upregulating the retinoic acid receptor in the ascidian *Ciona intestinalis*. Mech Dev 120:363-372.

Napoli JL. 1999. Retinoic acid: its biosynthesis and metabolism. Prog Nucleic Acid Res Mol Biol 63:139-188.

Niederreither K, McCaffery P, Drager UC, Chambon P, Dollé P. 1997. Restricted expression and retinoic acid-induced downregulation of the retinaldehyde dehydrogenase type 2 (RALDH-2) gene during mouse development. Mech Dev 62:67-78.

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

Niederreither K, Vermot J, Schuhbaur B, Chambon P, Dollé P. 2000. Retinoic acid synthesis and hindbrain patterning in the mouse embryo. Development 127:75-85.

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

Nikko E, Van de Vyver G, Richelle-Maurer E. 2001. Retinoic acid down-regulates the expression of *EmH-3* homeobox-containing gene in the freshwater sponge *Ephydatia muelleri*. Mech Ageing Dev 122:779-794.

Noden DM. 1983. The role of the neural crest in patterning of avian cranial skeletal, connective, and muscle tissues. Dev Biol 96:144-165.

Novitch BG, Wichterle H, Jessell TM, Sockanathan S. 2003. A requirement for retinoic acid-mediated transcriptional activation in ventral neural patterning and motor neuron specification. Neuron 40:81-95.

Ogasawara M, Wada H, Peters H, Satoh N. 1999. Developmental expression of *Pax1/9* genes in urochordate and hemichordate gills: insight into function and evolution of the pharyngeal epithelium. Development 126:2539-2550.

Ohno S. 1970. Evolution by gene duplication. New York: Springer-Verlag. 160p.

Ollikainen N, Chansawangbhuwana C, Baker ME. 2006. Evolution of the thyroid hormone, retinoic acid, ecdysone and liver X receptors. Integr Comp Biol 46:815-826.

Olivera-Martinez I, Storey KG. 2007. Wnt signals provide a timing mechanism for the FGF-retinoid differentiation switch during vertebrate body axis extension. Development. 134:2125-2135.

Osmond MK, Butler AJ, Voon FC, Bellairs R. 1991. The effects of retinoic acid on heart formation in the early chick embryo. Development 113:1405-1417.

Pan FC, Chen Y, Bayha E, Pieler T. 2007. Retinoic acid-mediated patterning of the pre-pancreatic endoderm in *Xenopus* operates via direct and indirect mechanisms. Mech Dev 124:518-531.

Papalopulu N, Kintner C. 1996. A posteriorising factor, retinoic acid, reveals that anteroposterior patterning controls the timing of neuronal differentiation in *Xenopus* neuroectoderm. Development 122:3409-3418.

Passamaneck YJ, Di Gregorio A. 2005. *Ciona intestinalis*: chordate development made simple. Dev Dyn 233:1-19.

Picking WL, Chen DM, Lee RD, Vogt ME, Polizzi JL, Marietta RG, Stark WS. 1996. Control of Drosophila opsin gene expression by carotenoids and retinoic acid: northern and western analyses. Exp Eye Res 63:493-500.

Piotrowski T, Nüsslein-Volhard C. 2000. The endoderm plays an important role in patterning the segmented pharyngeal region in zebrafish (*Danio rerio*). Dev Biol 225:339-356.

Plant MR, MacDonald ME, Grad LI, Ritchie SJ, Richman JM. 2000. Locally released retinoic acid repatterns the first branchial arch cartilages in vivo. Dev Biol 222:12-26.

Plum LA, Clagett-Dame M. 1996. All-trans retinoic acid stimulates and maintains neurite outgrowth in nerve growth factor-supported developing chick embryonic sympathetic neurons. Dev Dyn 205 :52-63.

Quinlan R, Gale E, Maden M, Graham A. 2002. Deficits in the posterior pharyngeal endoderm in the absence of retinoids. Dev Dyn 225:54-60.

Quinn SD, De Boni U. 1991. Enhanced neuronal regeneration by retinoic acid of murine dorsal root ganglia and of fetal murine and human spinal cord in vitro. In Vitro Cell Dev Biol 27:55-62.

Raner GM, Vaz AD, Coon MJ. 1996. Metabolism of all-trans, 9-*cis*, and 13-*cis* isomers of retinal by purified isozymes of microsomal cytochrome P450 and mechanism-based inhibition of retinoid oxidation by citral. Mol Pharmacol 49:515-522.

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.

Renaud JP, Rochel N, Ruff M, Vivat V, Chambon P, Gronemeyer H, Moras D. 1995. Crystal structure of the RAR-gamma ligand-binding domain bound to all-trans retinoic acid. Nature 378:681-689.

Roberts ES, Vaz AD, Coon MJ. 1992. Role of isozymes of rabbit microsomal cytochrome P-450 in the metabolism of retinoic acid, retinol, and retinal. Mol Pharmacol 41:427-433.
Romand R, Dollé P, Hashino E. 2006. Retinoid signaling in inner ear development. J Neurobiol 66:687-704.
Romero R, Bueno D. 2001. Disto-proximal regional determination and intercalary regeneration in planarians, revealed by retinoic acid induced disruption of regeneration. Int J Dev Biol 45:669-673.
Rosenthal N, Xavier-Neto J. 2000. From the bottom of the heart: anteroposterior decisions in cardiac muscle differentiation. Curr Opin Cell Biol 12:742-746.
Ross SA, McCaffery PJ, Drager UC, De Luca LM. 2000. Retinoids in embryonal development. Physiol Rev 80:1021-1054.
Ruberte E, Dollé P, Chambon P, Morriss-Kay G. 1991. Retinoic acid receptors and cellular retinoid binding proteins. II. Their differential pattern of transcription during early morphogenesis in mouse embryos. Development 111:45-60.
Schlosser G. 2005. Evolutionary origins of vertebrate placodes: insights from developmental studies and from comparisons with other deuterostomes. J Exp Zoolog B Mol Dev Evol 304:347-399.
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 240:262-273.
Schubert M, Holland ND, Escrivà H, Holland LZ, Laudet V. 2004. Retinoic acid influences anteroposterior positioning of epidermal sensory neurons and their gene expression in a developing chordate (amphioxus). Proc Natl Acad Sci U S A 101:10320-10325.
Schubert M, Yu JK, Holland ND, Escrivà H, Laudet V, Holland LZ. 2005. Retinoic acid signaling acts via *Hox1* to establish the posterior limit of the pharynx in the chordate amphioxus. Development 132:61-73.
Schubert M, Escrivà H, Xavier-Neto J, Laudet V. 2006a. Amphioxus and tunicates as evolutionary model systems. Trends Ecol Evol 21:269-277.
Schubert M, Holland ND, Laudet V, Holland LZ. 2006b. A retinoic acid-*Hox* hierarchy controls both anterior/posterior patterning and neuronal specification in the developing central nervous system of the cephalochordate amphioxus. Dev Biol 296:190-202.
Sciarrino S, Matranga V. 1995. Effects of retinoic acid and dimethylsulfoxide on the morphogenesis of the sea urchin embryo. Cell Biol Int 19:675-680.
Seo HC, Edvardsen RB, Maeland AD, Bjordal M, Jensen MF, Hansen A, Flaat M, Weissenbach J, Lehrach H, Wincker P, Reinhardt R, Chourrout D. 2004. *Hox* cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*. Nature 431:67-71.
Sharpe CR, Goldstone K. 1997. Retinoid receptors promote primary neurogenesis in *Xenopus*. Development 124:515-523.
Sharpe C, Goldstone K. 2000. The control of *Xenopus* embryonic primary neurogenesis is mediated by retinoid signalling in the neurectoderm. Mech Dev 91:69-80.
Shea C, Hough D, Xiao J, Tzertzinis G, Maina CV. 2004. An *rxr/usp* homolog from the parasitic nematode, Dirofilaria immitis. Gene 324:171-182.
Shim K, Picking WL, Kutty RK, Thomas CF, Wiggert BN, Stark WS. 1997. Control of Drosophila retinoid and fatty acid binding glycoprotein expression by retinoids

and retinoic acid: northern, western and immunocytochemical analyses. Exp Eye Res 65:717-727.
Simões-Costa MS, Vasconcelos M, Sampaio AC, Cravo RM, Linhares VL, Hochgreb T, Yan CY, Davidson B, Xavier-Neto J. 2005. The evolutionary origin of cardiac chambers. Dev Biol 277:1-15.
Simões-Costa MS, Azambuja AP, Xavier-Neto J. 2008. The search for non-chordate retinoic acid signaling: lessons from chordates. J Exp Zoolog B Mol Dev Evol 310:54-72.
Sirbu IO, Duester G. 2006. Retinoic-acid signalling in node ectoderm and posterior neural plate directs left-right patterning of somitic mesoderm. Nat Cell Biol 8:271-277.
Sive HL, Draper BW, Harland RM, Weintraub H. 1990. Identification of a retinoic acid-sensitive period during primary axis formation in *Xenopus laevis*. Genes Dev 4:932-942.
Sockanathan S, Perlmann T, Jessell TM. 2003. Retinoid receptor signaling in postmitotic motor neurons regulates rostrocaudal positional identity and axonal projection pattern. Neuron 40:97-111.
Söderhäll I, Tangprasittipap A, Liu H, Sritunyalucksana K, Prasertsan P, Jiravanichpaisal P, Söderhäll K. 2006. Characterization of a hemocyte intracellular fatty acid-binding protein from crayfish (*Pacifastacus leniusculus*) and shrimp (*Penaeus monodon*). FEBS J 273:2902-2912.
Stafford D, Prince VE. 2002. Retinoic acid signaling is required for a critical early step in zebrafish pancreatic development. Curr Biol 12:1215-1220.
Stainier DY, Fishman MC. 1992. Patterning the zebrafish heart tube: acquisition of anteroposterior polarity. Dev Biol 153:91-101.
Styczynski MP, Jensen KL, Rigoutsos I, Stephanopoulos G. 2008. BLOSUM62 miscalculations improve search performance. Nat Biotechnol 26:274-275.
Sun D, Chen DM, Harrelson A, Stark WS. 1993. Increased expression of chloramphenicol acetyltransferase by carotenoid and retinoid replacement in *Drosophila* opsin promoter fusion stocks. Exp Eye Res 57:177-187.
Tabin C. 1995. The initiation of the limb bud: growth factors, *Hox* genes, and retinoids. Cell 80:671-674.
Tanaka Y, Okada Y, Hirokawa N. 2005. FGF-induced vesicular release of Sonic hedgehog and retinoic acid in leftward nodal flow is critical for left-right determination. Nature 435:172-177.
Trainor PA, Krumlauf R. 2001. *Hox* genes, neural crest cells and branchial arch patterning. Curr Opin Cell Biol 13:698-705.
Tsukui T, Capdevila J, Tamura K, Ruiz-Lozano P, Rodriguez-Esteban C, Yonei-Tamura S, Magallon J, Chandraratna RA, Chien K, Blumberg B, Evans RM, Belmonte JC. 1999. Multiple left-right asymmetry defects in *Shh*$^{-/-}$ mutant mice unveil a convergence of the Shh and retinoic acid pathways in the control of *Lefty-1*. Proc Natl Acad Sci U S A 96:11376-11381.
Veitch E, Begbie J, Schilling TF, Smith MM, Graham A. 1999. Pharyngeal arch patterning in the absence of neural crest. Curr Biol 9:1481-1484.
Vermot J, Pourquié O. 2005. Retinoic acid coordinates somitogenesis and left-right patterning in vertebrate embryos. Nature 435:215-220.
Vermot J, Gallego Llamas J, Fraulob V, Niederreither K, Chambon P, Dollé P. 2005. Retinoic acid controls the bilateral symmetry of somite formation in the mouse embryo. Science 308:563-566.

Vernet N, Dennefeld C, Rochette-Egly C, Oulad-Abdelghani M, Chambon P, Ghyselinck NB, Mark M. 2006. Retinoic acid metabolism and signaling pathways in the adult and developing mouse testis. Endocrinology 147:96-110.

Wada H, Escrivà H, Zhang S, Laudet V. 2006. Conserved RARE localization in amphioxus *Hox* clusters and implications for *Hox* code evolution in the vertebrate neural crest. Dev Dyn 235:1522-1531.

Wang Y, Zolfaghari R, Ross AC. 2002. Cloning of rat cytochrome P450RAI (CYP26) cDNA and regulation of its gene expression by all-trans-retinoic acid in vivo. Arch Biochem Biophys 401:235-243.

Wasiak S, Lohnes D. 1999. Retinoic acid affects left-right patterning. Dev Biol 215:332-342.

Wendling O, Dennefeld C, Chambon P, Mark M. 2000. Retinoid signaling is essential for patterning the endoderm of the third and fourth pharyngeal arches. Development 127:1553-1562.

White JA, Beckett-Jones B, Guo YD, Dilworth FJ, Bonasoro J, Jones G, Petkovich M. 1997. cDNA cloning of human retinoic acid-metabolizing enzyme (hP450RAI) identifies a novel family of cytochromes P450. J Biol Chem 272:18538-18541.

White JC, Highland M, Kaiser M, Clagett-Dame M. 2000. Vitamin A deficiency results in the dose-dependent acquisition of anterior character and shortening of the caudal hindbrain of the rat embryo. Dev Biol 220:263-284.

White RJ, Nie Q, Lander AD, Schilling TF. 2007. Complex regulation of *cyp26a1* creates a robust retinoic acid gradient in the zebrafish embryo. PloS Biol 5:e304.

Wiens M, Batel R, Korzhev M, Müller WE. 2003. Retinoid X receptor and retinoic acid response in the marine sponge *Suberites domuncula*. J Exp Biol 206:3261-3271.

Wilson L, Gale E, Maden M. 2003. The role of retinoic acid in the morphogenesis of the neural tube. J Anat 203:357-368.

Wilson L, Gale E, Chambers D, Maden M. 2004. Retinoic acid and the control of dorsoventral patterning in the avian spinal cord. Dev Biol 269:433-446.

Wingert RA, Selleck R, Yu J, Song HD, Chen Z, Song A, Zhou Y, Thisse B, Thisse C, McMahon AP, Davidson AJ. 2007. The *cdx* genes and retinoic acid control the positioning and segmentation of the zebrafish pronephros. PLoS Genet 10:1922-1938.

Wood H, Pall G, Morriss-Kay G. 1994. Exposure to retinoic acid before or after the onset of somitogenesis reveals separate effects on rhombomeric segmentation and 3' *HoxB* gene expression domains. Development 120:2279-2285.

Wuarin L, Sidell N, de Vellis J. 1990. Retinoids increase perinatal spinal cord neuronal survival and astroglial differentiation. Int J Dev Neurosci 8:317-326.

Xavier-Neto J, Neville CM, Shapiro MD, Houghton L, Wang GF, Nikovits W Jr, Stockdale FE, Rosenthal N. 1999. A retinoic acid-inducible transgenic marker of sino-atrial development in the mouse heart. Development 126:2677-2687.

Xavier-Neto J, Rosenthal N, Silva FA, Matos TG, Hochgreb T, Linhares VL. 2001. Retinoid signaling and cardiac anteroposterior segmentation. Genesis 31:97-104.

Yagi K, Makabe KW. 2002. Retinoic acid differently affects the formation of palps and surrounding neurons in the ascidian tadpole. Dev Genes Evol 212:288-292.

Yang ZN, Davis GJ, Hurley TD, Stone CL, Li TK, Bosron WF. 1994. Catalytic efficiency of human alcohol dehydrogenases for retinol oxidation and retinal reduction. Alcohol Clin Exp Res 18:587-591.

Yost RW, Harrison EH, Ross AC. 1988. Esterification by rat liver microsomes of retinol bound to cellular retinol-binding protein. J Biol Chem 263:18693-18701.

Yu JK, Satou Y, Holland ND, Shin-I T, Kohara Y, Satoh N, Bronner-Fraser M, Holland LZ. 2007. Axial patterning in cephalochordates and the evolution of the organizer. Nature 445 :613-617.

Zákány J, Kmita M, Alarcon P, de la Pompa JL, Duboule D. 2001. Localized and transient transcription of *Hox* genes suggests a link between patterning and the segmentation clock. Cell 106:207-217.

Research Chapters

Differential Regulation of ParaHox Genes by Retinoic Acid in the Invertebrate Chordate Amphioxus (*Branchiostoma floridae*)

Peter W. Osborne, Gérard Benoît, Vincent Laudet, Michael Schubert†, David E. K. Ferrier†
†These authors contributed equally to this work.

Abstract

The ParaHox cluster is the evolutionary sister to the Hox cluster. Like the Hox cluster, the ParaHox cluster displays spatial and temporal regulation of the component genes along the anterior/posterior axis in a manner that correlates with the gene positions within the cluster (a feature called colinearity). The ParaHox cluster is however a simpler system to study because it is composed of only three genes. We provide a detailed analysis of the amphioxus ParaHox cluster and, for the first time in a single species, examine the regulation of the cluster in response to a single developmental signalling molecule, retinoic acid (RA). Embryos treated with either RA or RA antagonist display altered ParaHox gene expression: *AmphiGsx* expression shifts in the neural tube, and the endodermal boundary between *AmphiXlox* and *AmphiCdx* shifts its anterior/posterior position. We identified several putative retinoic acid response elements and in vitro assays suggest some may participate in RA regulation of the ParaHox genes. By comparison to vertebrate ParaHox gene regulation we explore the evolutionary implications. This work highlights how insights into the regulation and evolution of more complex vertebrate arrangements can be obtained through studies of a simpler, unduplicated amphioxus gene cluster.

Introduction

Within chordates, a single intact ParaHox cluster of three genes has been conserved in tetrapods and amphioxus, whose lineages separated over 500 million years ago (Brooke et al., 1998; Coulier et al., 2000; Ferrier et al., 2005). The three ParaHox gene families, Gsx (or Gsh or ind), Xlox (or Xlhbox8, IPF1, PDX1, IDX1, STF1 or Lox3) and Cdx (or caudal/cad) are widespread within Bilateria (Ferrier and Holland, 2001a) and at the origin of chordates the ParaHox genes were linked in a tight cluster with Xlox located between Gsx and Cdx (Brooke et al., 1998; Finnerty and Martindale, 1999; Ferrier and Holland, 2001b). Unlike the single cluster of amphioxus however, tetrapods also have several extra ParaHox genes (an extra Gsx and two extra Cdx genes) as a result of two rounds of whole genome duplications followed by gene loss. Teleost fish in contrast, have disrupted the structure of their ParaHox cluster because of an extra round of whole genome duplications and subsequent gene loss, although the overall number and complement of vertebrate ParaHox genes has been retained in the teleost lineage (Mulley et al., 2006; Prohaska and Stadler, 2006; Siegel et al., 2007). The single hagfish ParaHox cluster is also in the process of degeneration, with Xlox mutated to a pseudo-gene (Furlong et al., 2007). Urochordates and echinoderms have not undergone genome duplications but have also disrupted their ParaHox cluster, due to rearrangement of their respective genomes. These invertebrate deuterostome lineages have thus retained all of their ParaHox genes, but not the ParaHox cluster, which contrasts with the mechanism of cluster disruption via gene loss in some vertebrates (Ferrier and Holland, 2002; Arnone et al., 2006). Although widely conserved, the ParaHox cluster, like its paralogue the Hox cluster, is therefore not completely immune to disruption (Ferrier and Minguillon, 2003).

Outside of the chordates, lophotrochozoans have representatives of all three genes (Ferrier and Holland, 2001a) while only representatives of the Cdx and Gsx families have been identified so far in ecdysozoans. In both groups no intact cluster has yet been identified. Overall, the ancestral bilaterian clearly had three ParaHox genes that were most likely arranged in a cluster similar to that maintained by amphioxus. However, it remains a mystery as to why some groups have retained a ParaHox cluster over long periods of time, while other groups have not. Either a selective constraint has kept the ParaHox cluster together, for example through shared cis-regulation (Mulley et al., 2006; Duboule, 2007), or alternatively they have simply been retained as an evolutionary relic via slow genomic evolution in some lineages. The single ParaHox cluster of amphioxus constitutes a good system to investigate these possible regulatory mechanisms because it will avoid the problem of functional redundancy between duplicates inherent with studying vertebrate ParaHox genes (Schubert et al., 2006a). By identifying amphioxus regulatory mechanisms conserved with other chordates we can also reconstruct the ancestral regulation of the cluster and begin to answer questions on how and why the ParaHox cluster has been maintained in certain lineages.

Retinoic acid (RA) is a derivative of vitamin A (retinol) involved in regulating chordate anterior-posterior (A/P) patterning (Mark et al., 2006; Marletaz et al., 2006; Maden, 2007). RA function is mediated by retinoic acid receptor (RAR) and retinoid X receptor (RXR) heterodimers binding to specific regulatory DNA elements (called RA response elements or RAREs) and activating transcription in a ligand-dependent manner (Mark et al., 2006). RAREs consist of two direct repeats (DR), with the canonical nucleotide sequence (A/G)G(G/T)TCA, that are usually separated by either

2 or 5 nucleotide spacers (Bastien and Rochette-Egly, 2004). Compared with Hox genes, there are relatively few examples of RA regulating ParaHox genes in chordates and the ParaHox cluster as a whole has not been examined.

Currently there are no reports of vertebrate Gsx genes responding to RA but there are several studies examining the interaction between RA and the other two ParaHox genes. In mice, *Cdx1* is directly regulated by RA and RA treatment causes an early induction and a posterior expansion of *Cdx1* (Houle et al., 2000). The direct regulation of *Cdx1* by RA signalling is partially mediated by an atypical DR5-type RARE located upstream of the *Cdx1* gene (Houle et al., 2000; Lickert and Kemler, 2002; Houle et al., 2003; Pilon et al., 2007). Interestingly, from *Cdx1* RARE null mutants a second RA signalling pathway was identified and found to regulate *Cdx1* (Houle et al., 2003), probably through a DR2-type RARE conserved in the mouse and chicken *Cdx1/CdxA* introns (Gaunt et al., 2003). In contrast to these results in mouse and chicken, one of the two zebrafish *Cdx1* genes is not regulated by RA (at least not in the tissues or stages studied) (Stafford and Prince, 2002). In addition, the other two mouse Cdx paralogues are not induced by RA (Roelen et al., 2002). Instead, excess RA represses *Cdx4* anteriorly (Iulianella et al., 1999) and may subtly reduce *Cdx2* expression (Roelen et al., 2002). Finally, recent work indicates zebrafish Cdx4 may actually prevent RA signal transduction in posterior endoderm (Kinkel et al., 2008). This variation of Cdx responses to RA signalling indicates the regulation of these genes has evolved since the ParaHox cluster duplicated, making an analysis of the unduplicated amphioxus Cdx gene critical for understanding the ancestral regulation of the gene.

In RA-depleted embryos, Xlox expression is abolished in the dorsal pancreatic anlage in zebrafish (Stafford and Prince, 2002) and mice (Martin et al., 2005; Molotkov et al., 2005). Conversely, maternal replacement of RA in deficient mice allows recovery of Xlox expression (Martin et al., 2005; Molotkov et al., 2005). Furthermore, excess RA also causes up-regulation of Xlox expression in mouse cell lines (Tulachan et al., 2003; Micallef et al., 2005). RA has also been demonstrated to induce ectopic Xlox expression in chick/quail chimera experiments (Kumar et al., 2003).

In this report, we aimed to examine the evolution of RA regulation of ParaHox genes in chordates. We undertook a comprehensive examination of all three ParaHox genes from the Florida amphioxus, *Branchiostoma floridae,* and assessed the response to exogenous RA and RA antagonist. The treatment of amphioxus with RA has been well characterised in the past. Embryos treated with excess RA at gastrulation display a stereotypical change in development (Holland and Holland, 1996). The most distinctive changes are the loss of mouth and gill slits and a massive reduction of the pharyngeal endoderm. This is despite a lack of neural crest in amphioxus which was long believed to be the main mediator for the effects of RA in the vertebrate branchial arches (Escriva et al., 2002). In a similar manner to vertebrates, gene expression is also dramatically altered in amphioxus treated with RA: for instance Hox genes are shifted anteriorly by RA in the neural tube (Holland and Holland, 1996, Schubert et al. 2006b). In contrast, the development of the notochord and musculature of amphioxus is not clearly affected by the addition of RA. Moreover, while *AmphiRXR* is weakly expressed throughout the amphioxus embryo (Escriva et al., 2002), *AmphiRAR* is expressed throughout the mesendoderm at gastrula stages and by neurula stages is widely expressed in the posterior neural plate and most mesendodermal derivatives (Escriva et al., 2002). During subsequent development, RAR is downregulated anteriorly and posteriorly and is restricted to

more central regions of the neural tube, somites and endoderm (Escriva et al., 2002). Thus, RA is likely to be acting in locations relevant to the expression of the amphioxus ParaHox genes.

We provide a full description of the expression of amphioxus ParaHox genes in embryogenesis and early larval development, revealing a novel expression domain of *AmphiGsx* that has not previously been documented. For the first time in a single species we show the response of all three ParaHox genes to a single signalling molecule. All three genes are regulated by RA, with RA causing an A/P shift in some expression domain boundaries. In addition, *AmphiCdx* and *AmphiXlox* share a RA-sensitive boundary of expression in the posterior endoderm. We have also begun to establish whether this RA regulation is via a direct mechanism by identifying putative RAREs in the amphioxus ParaHox cluster.

Materials and Methods

Embryology, in situ hybridisation, and microscopy

Embryos of the Florida amphioxus (*Branchiostoma floridae*) were treated at the very late blastula stage with $1x10^{-6}$ M RA or RA antagonist (BMS009) diluted in DMSO or with DMSO alone (as described previously in Escriva et al., 2002; Schubert et al. 2006b). In situ hybridisations were performed essentially as described elsewhere (Holland et al., 1996). Staining was conducted at 4°C. The *AmphiCdx* clone (pC15.1) was a kind gift from Jordi Garcia-Fernández, initially cloned from an amphioxus cDNA library (Langeland et al., 1998). First and second exons of *AmphiGsx* and *AmphiXlox* were amplified from the two amphioxus ParaHox PACs (Genbank accession numbers AC129948 and AC129947) and subsequently fused by PCR. Total sizes of the clones were: *AmphiXlox* 1174 bp and *AmphiGsx* 1018 bp. Primers used to clone *AmphiGsx* and *AmphiXlox* were: BfGsx5'-F GTCGAACGCCTTTGTGAAGT; BfGsx5'-R TGATGCCACTCCAGAGGGGAAAGGTAGG; BfGsx3'-F TCCCCTCTGGAGTGGCATCAGATGGTC; BfGsx3'-R TACGACAACGCAAAGTAACG; BfXlox5'-F TTCAAACGATACCGGACAAAC; BfXlox5'-R AAGGACGCACCACCTGGCCATTGAGAC; BfXlox3'-F GGCCAGGTGGTGCGTCCTTTGCTGTTG; BfXlox3'-R ATGAAAAACCACCTGCGTTG. Stained amphioxus embryos were visualised on a Zeiss Axioskop2 microscope. The percentage of body length expressing each gene was calculated from measurements of gene expression domains and embryo length using Axiovision 4. One-way ANOVA tests were undertaken on raw (cells numbers expressing *AmphiGsx*) or log transformed data (relative lengths of expression of *AmphiXlox* or *AmphiCdx*) to determine if there were significant differences between treatments at each stage. See Supplementary Tables S2-S4 for numbers of embryos used in the analysis.

Identification of RAREs

NHR scan (Sandelin and Wasserman, 2005) identified four DR5 and two DR2 element sequences in the amphioxus ParaHox cluster (Supplementary Table S1). These six putative RAREs were further analysed, as was a putative DR5 element located upstream of *AmphiGsx* that was identified by a manual search (using Gene Palette (Rebeiz and Posakony, 2004)) for the consensus sequence (A/G)G(G/T)TCA[2/5](A/G)G(G/T)TCA (Perlmann et al., 1993).

Electrophoretic mobility shift assays (EMSAs)

EMSAs were carried out as previously described (Vanacker et al., 1999). Complementary primers (Supplementary Table S1) corresponding to the putative RAREs (plus flanking sequences) with 5' overhanging restriction sites were annealed and labelled with γ^{32}P-ATP. Radiolabelled RAREs were bound to the amphioxus RAR and RXR proteins at 4°C (in Hepes, 2 mM; NaCl, 50 mM; KCl, 50mM; $MgSO_4$, 3 mM; glycerol, 10%; DTT, 2 mM; poly dIdC, 0.04 µg/µL) and run on a 6% acrylamide gel, followed by gel drying and signal detection.

Transactivation assays

Reporter construct cloning and cell culture experiments were performed as described (Wang et al., 2003). Complementary primers (Supplementary Table S1) corresponding to the putative RAREs (plus flanking sequence) with 5' overhanging restriction sites were annealed and cloned into the pGL2 promoter vector (Promega). RARE island 1 (7702 bp) was PCR amplified using the primers RI1F-TTCTTTCGGCGTCGTTATTC and RI1R-CTCGAGAACCGCAAGAGCAACAC which are 1163 bp upstream and 2033 bp downstream of DR5b and DR5c respectively. RARE island 2 (7061 bp) was PCR amplified using the primers RI2F-GGATCCCGCAATAGATACGTCAACA and RI2R-GGATCCCACCAGCTCACCCTAAA which are 363 bp upstream and 1482 bp downstream of DR5e and DR2b respectively. Both PCR fragments were subcloned into the pGL2 promoter vector upstream of a minimal promoter element and the luciferase reporter gene to assay possible activities as enhancer elements. The human embryonic kidney cell line 293 was grown in DMEM + 10% foetal calf serum and antibiotics in 2 cm^2 wells to 80-90% confluence. All cells were co-transfected with a DNA solution of vector containing the RARE construct to be tested plus a plasmid encoding β-galactosidase as a baseline control for transfection efficiency between treatments. Transfection was conducted in the presence of lipofectamine and Plus Reagent (Invitrogen) but in the absence of foetal calf serum. In addition, controls were transfected with either *AmphiRAR* (in pSG5), *AmphiRXR* (in pSG5) or empty vector alone, whereas experimental treatments contained both *AmphiRAR* and *AmphiRXR* constructs. Cells were transfected for five hours and then foetal calf serum was added to all cells, while all-*trans* RA at a final concentration of $5x10^{-7}$ M was added to half of the wells. After 24 hours, levels of b-galactosidase and luciferase activity were measured.

Results

Wild-type expression of amphioxus ParaHox genes

To examine the regulation of amphioxus ParaHox genes, we first undertook a detailed description of the endogenous gene expression. *AmphiCdx* is the first ParaHox gene to be expressed. It initiates during mid-gastrulation in a ring around the blastopore and this expression continues through blastopore closure (Fig. 1L). As neurulation begins, expression is strongest dorsally in the neuroectoderm, but there is also ventral and lateral expression throughout the posterior of the embryo in ectoderm and to a lesser extent in ventral endoderm (Fig. 1M). During early stages of somitogenesis, expression is present in the neuroectoderm, the walls of the posterior archenteron and in the tail bud region in a continuous domain (Fig. 1N). Neuroectoderm expression is strong posteriorly and weaker at more anterior levels, such that *AmphiCdx* expression in the neuroectoderm is graded from posterior to

anterior. Moreover, the anterior limit of expression is generally more rostral in the neuroectoderm than in the archenteron. This difference in A/P levels is maintained in late embryos (Fig. 1O-P), although after one week of development neural expression is down-regulated anteriorly, hence reversing these A/P level differences (Fig. 1Q). Throughout development there is expression in the tail bud, and the posterior elongation of the embryo results in expansion of expression so that *AmphiCdx* is also expressed throughout the hindgut and posterior neural tube until at least 7 days post fertilisation, the oldest stage examined (Fig. 1M-Q).

The earliest *AmphiXlox* expression was observed after gastrulation in the ventral posterior archenteron wall at low levels (data not shown). This domain subsequently strengthens and at approximately the same A/P position a weaker domain develops dorsally (Fig. 1E). This dorsal expression subsequently increases in intensity and spreads to include the neuroectoderm and mesendoderm (Fig. 1F). During neurulation, expression is initiated posteriorly in the embryo, linking the dorsal and ventral domains. There is also strong *AmphiXlox* expression in two cells of the neural tube approximately level with the anterior boundary of somite five (where the first pigment spot will form) (Fig. 1G,H). In late neurulae, *AmphiXlox* expression is maintained in the neural tube and the posterior expression in the embryo separates again into two distinct domains. The first is a ventral domain in the posterior gut, while the second is located more posteriorly in the dorsal mesendoderm close to the tail bud (Fig. 1I). In early larvae, neural expression is down-regulated and the mesendodermal domain lost. Larval growth causes an A/P expansion of *AmphiXlox* endodermal expression so that it now marks a specific domain in the posterior midgut/anterior hindgut (Fig. 1J). This distinctive domain is maintained until at least seven days post fertilisation with strong expression at its anterior limit, which weakens posteriorly (Fig. 1K).

AmphiGsx is the last gene to be activated, and it is expressed in a more restricted pattern. There are two temporally distinct domains of *AmphiGsx* expression, an ‘early’ and a ‘late’ domain. The early domain arises during neurulation and consists of four cells in the neural tube, level with somite five (Fig. 1B,C). This early *AmphiGsx* domain has not previously been identified and overlaps with *AmphiXlox* neural expression, although *AmphiGsx* is expressed in slightly more cells (compare Fig. 1C and 1H). In late embryos, a transient expression domain is detectable in the most anterior part of the neural tube, the cerebral vesicle (CV) (Fig. 1D). Expression is in the centre of the CV (along the A/P axis) on both left and right sides and remains until early larval stages (Fig. 1D inset). This detailed description of amphioxus ParaHox expression patterns extends and improves the initial description provided by Brooke et al. (1998).

RA changes amphioxus ParaHox expression in the endoderm

To assess whether RA regulates amphioxus ParaHox gene expression we examined the expression of all three ParaHox genes in amphioxus embryos treated with either RA, RA antagonist (BMS009) or DMSO (control). For both *AmphiCdx* and *AmphiXlox*, the most conspicuous effects of RA and BMS009 treatments are detectable in the late embryo endoderm (Figs 2,3). At this stage, a significant response to varying RA signalling levels for both *AmphiCdx* (one-way ANOVA, $p=0.001$) and *AmphiXlox* (one-way ANOVA, $p<0.0001$) is observed in the posterior endoderm. These alterations remain significant during subsequent development (Figs 2,3). Both genes respond in a similar but complementary manner to changing RA signalling levels (Figs 2,3). The anterior limit of endodermal *AmphiCdx*

expression in RA-treated embryos is shifted posteriorly, and consistent with this response, the anterior limit of expression is shifted rostrally in BMS009-treated embryos (Fig. 2). Interestingly, it is the posterior limit of *AmphiXlox* that is changed in response to RA and BMS009 rather than the anterior boundary (Fig. 3). In RA-treated amphioxus embryos, *AmphiXlox* is expanded posteriorly, and BMS009 treatment moves the posterior expression limit rostrally.

From these complementary shifts in expression, it appears RA signalling influences the boundary of expression between the two ParaHox genes. In an attempt to elucidate the details of this shift, we examined the expression of both genes simultaneously. In single colour double in situ hybridisation experiments on control embryos (Fig. 4A,B) and on embryos treated with either RA or BMS009 (data not shown), only a single continuous domain of expression can be observed, indicating that there is no gap in expression between the two genes. From two colour double in situ hybridisations the expression patterns clearly overlap in late neurulae in the ventral archenteron walls (data not shown). However as the pharynx begins to form, the overlap shrinks so by the late embryo stage little if any overlap in the expression domains in the archenteron can be seen (Figs 4E,F). This is also obvious in RA- and BMS009-treated embryos (Figs 4C,D,G,H) implying that subsequent to the initial activation of *AmphiCdx* and *AmphiXlox*, there is a secondary refinement of the limits of their expression domains, so that the two domains become adjacent to each other. It is the A/P position of this boundary during later stages of amphioxus development that is sensitive to RA.

RA alters AmphiGsx *expression in the neuroectoderm*

Unlike the endodermal modification of *AmphiCdx* and *AmphiXlox*, RA does not regulate expression of these two ParaHox genes in the neuroectoderm. However, RA does have a profound effect on *AmphiGsx* in the neuroectoderm. Expression of *AmphiGsx* is not modified in the cerebral vesicle in response to either RA or BMS009 but RA signalling does modify the A/P position of the early *AmphiGsx* domain (Fig. 5). Treatment with exogenous RA shifts and expands *AmphiGsx* expression anteriorly by increasing the number of *AmphiGsx*-expressing cells. An anterior shift was observed in all RA-treated embryos (Fig. 5A); however some embryos displayed a pattern of individual cells with high levels of expression interspersed by cells with low or no observable expression. In control embryos, the anterior and posterior limits of *AmphiGsx* expression were level with the anterior and posterior limits of somite 5 respectively (Fig. 5B). In RA-treated embryos the posterior limit of expression was shifted rostrally by approximately one somite length. The anterior limit of *AmphiGsx* was affected even more profoundly by RA, shifting rostrally by several somite lengths. An extreme effect was also observed in BMS009-treated embryos where the signal becomes completely undetectable by in situ hybridisation (even when stained for 30 days) (Fig. 5C). To quantify early *AmphiGsx* expression, the total number of *AmphiGsx*-expressing cells induced by RA and repressed by BMS009 were counted in treated embryos and compared to control embryos (Fig. 5G). The total number of cells expressing *AmphiGsx* are significantly different in response to treatment with RA or BMS009 (one-way ANOVA, $p<0.0001$).

Identification and analysis of putative ParaHox RAREs

As expression of all three ParaHox genes is clearly modified by RA treatment, we hypothesised a direct regulation and thus attempted to identify putative RAREs in the ParaHox cluster. Seven putative RAR/RXR binding sites were identified within

the amphioxus ParaHox cluster. Six RAREs were localised within two distinct genomic segments, a region (Island 1) containing three putative RAREs between *AmphiGsx* and *AmphiXlox* (DR5a, DR5b and DR5c) and a region (Island 2) of three more putative RAREs between *AmphiXlox* and *AmphiCdx* (DR5d, DR2a and DR2b). In addition, a single weak match to a RARE consensus sequence was identified upstream of *AmphiGsx* (DR5e) (Fig. 6A). As *AmphiGsx* expression was so dramatically altered by RA we tested this putative RARE as well, despite its lack of significance in the NHR scan. These seven putative ParaHox RAREs (Supplementary Table S1) were tested in vitro using a combination of EMSAs and heterologous cell culture transactivation experiments. In both assays, DR5c was the only single element that was bound weakly by the amphioxus RAR/RXR heterodimer (Fig. 6B, Supplementary Fig. S1). To further test DR5c, we analysed the ability of the amphioxus RAR/RXR heterodimer to activate transcription of a reporter construct containing single or tandem copies of DR5c. Moreover, we also tested the activity of DR5c in its native genomic context, ParaHox RARE Island 1 (Fig. 6B). The ParaHox Island 1 used in the analysis includes DR5b-DR5c (and several additional sequences with only a single nucleotide difference from the consensus DR5 which were not significant under NHR scan criteria). We also assayed the activity of ParaHox RARE Island 2 (Fig. 6A). A single copy of DR5c consistently causes more than 2-fold induction after stimulation with RA (Fig. 6B). However, multiple copies of the DR5c element did not significantly increase the level of induction of the reporter construct. The Island 1 construct (containing DR5c) consistently shows approximately 3-fold induction after treatment with RA. The RARE Island 2 construct is not significantly activated by RA, indicating it is not efficiently regulated by the amphioxus RAR/RXR heterodimer. Taken together, and considering that we are probably working at the limits of detection for such a heterologous assay system, this implies the amphioxus ParaHox DR5c element can be bound by amphioxus RAR/RXR and potentially mediate a direct RA-dependent response.

Discussion

Evolutionary constraints on the chordate ParaHox gene cluster

Intact ParaHox clusters have been identified in amphioxus, mice, humans (Brooke et al., 1998; Ferrier et al., 2005) and *Xenopus* (H.V. Isaacs personal communication) and it is generally assumed chordates have maintained this gene cluster as a result of selective constraints. It has been proposed these constraints are due to shared enhancer elements (Brooke et al., 1998). Alternatively it may be that the amphioxus and tetrapod lineages have simply retained the ParaHox cluster as an evolutionary relic. Indeed, recent demonstrations that both the hagfish and teleost lineages have lost their ParaHox clusters (Mulley et al., 2006; Furlong et al., 2007) along with cluster break-up in the urochordates (Ferrier and Holland, 2002) could argue against selective constraints maintaining chordate ParaHox clusters . A better understanding of the regulation of the ParaHox genes in intact clusters will allow us to understand whether the cluster is maintained by chance or through selective constraints, released in animals with unusual features (such as derived development, degenerate morphology or duplicated genomes).

Detailed descriptions of the expression of each ParaHox gene are an absolute prerequisite for studying their regulation. To this end we present a more extensive analysis of amphioxus ParaHox gene expression than previously described (Brooke et al., 1998). We show the importance of such careful examinations by identifying an

extra *AmphiGsx* expression domain in the neural tube that overlaps with the *AmphiXlox* neural domain. This newly discovered *AmphiGsx* domain does not break the spatial or temporal collinearity of the amphioxus ParaHox gene expression (Brooke et al., 1998). The genes are still activated temporally (first to last) and spatially (posterior to anterior) in the order *AmphiCdx*, *AmphiXlox* and then *AmphiGsx*. The *AmphiXlox* and *AmphiCdx* genes also overlap during early embryonic development and later share a boundary in the endoderm. These overlapping/abutting domains raise the possibility that pairs of ParaHox genes share upstream transcriptional regulators and perhaps even share enhancers, which if present could provide a selective constraint for cluster retention.

RA regulates endodermal expression of AmphiXlox *and* AmphiCdx

The *AmphiXlox* domain in the gut endoderm is expanded posteriorly in response to exogenous RA, whereas the *AmphiCdx* domain in the gut is reduced anteriorly. Double in situ hybridisations suggest RA is actually shifting the A/P position of the boundary between the two genes and these results imply a tight mechanistic link between the regulation of *AmphiXlox* and *AmphiCdx*. Previous studies have demonstrated RA signalling in amphioxus is involved in development of the anterior (pharyngeal) and fore/midgut endoderm (Holland and Holland, 1996; Cañestro et al., 2001; Schubert et al., 2005). However, this is the first demonstration that posterior endodermal marker genes also respond to RA in amphioxus, indicating the entire A/P axis of the amphioxus endoderm is patterned through RA-dependent mechanisms.

There are three major mechanisms that could explain the shift of the *AmphiCdx/AmphiXlox* boundary in the endoderm: (1) both ParaHox genes could be regulated directly and independently through one or more RAREs, (2) one of the ParaHox genes could be regulated directly through a RARE and the second ParaHox gene is regulated by the first, (3) neither ParaHox gene is directly regulated through RAREs and instead both are regulated by an upstream gene which is RA responsive. These mechanisms are not mutually exclusive and a combination could be acting on the *AmphiXlox/AmphiCdx* boundary.

Under scenario one, both *AmphiCdx* and *AmphiXlox* would be directly regulated by RA. This could be through a single RARE such as the identified DR5c upstream of *AmphiXlox,* or there may be additional atypical or distant RAREs regulating these genes not identified in this study.

Regulation of a boundary between two genes is possibly more likely to occur through cross-regulation between the genes (for example Toresson et al., 2000; Tour et al., 2002). This makes scenario two more likely and only one gene needs to be regulated by RA, with this gene subsequently regulating the A/P limit of the other gene. Although not definitive we postulate *AmphiXlox* is more likely to be directly regulated than *AmphiCdx* as *AmphiCdx* is repressed by ectopic RA, and RAR/RXR heterodimers normally mediate transcriptional activation rather than repression upon RA binding (Bastien and Rochette-Egly, 2004).

It is also possible that neither ParaHox gene is directly regulated by RA but instead an upstream, RA-responsive gene might regulate the *AmphiXlox/AmphiCdx* boundary. A good candidate for this would be *AmphiTR2/4* as it is RA responsive, competitively binds to RAREs and is expressed posteriorly in amphioxus embryos. In the *AmphiXlox/AmphiCdx* boundary region there are low-mid levels of both *AmphiRAR* and *AmphiTR2/4*. Under excess RA conditions, *AmphiRAR* is up-regulated throughout the posterior gut, while *AmphiTR2/4* decreases in the posterior

endoderm (Escriva et al., 2002). In contrast, with BMS009 treatment, *AmphiRAR* levels are dramatically decreased in the posterior endoderm and *AmphiTR2/4* is up-regulated throughout the gut. Alteration of RA levels therefore clearly changes the balance between AmphiTR2/4 and AmphiRAR/AmphiRXR binding to RAREs in the posterior endoderm implicating AmphiTR2/4 as a candidate for regulating the *AmphiCdx*/*AmphiXlox* boundary.

In *Xenopus, Xlox*/*Xlhbox8* and *Cdx2* are expressed in abutting locations, respectively in the stomach/pancreas and small intestine (Horb and Slack, 2001). Zebrafish *cad1* (which is unaffected by exogenous RA) is expressed immediately posterior to the pancreas (where *Xlox*/*Pdx1* is expressed) (Stafford and Prince, 2002) and zebrafish *cdx4* initially does not overlap with *pdx1*, though by 16 hours of development there are a few cells expressing both genes (Kinkel et al., 2008). In the mouse endoderm there appears to be a very small (if any) overlap of *Xlox/Pdx1* and *Cdx2* expression (Fang et al., 2006). The same is true for sea urchins, with expression of *Sp-Xlox* and *Sp-Cdx* meeting at the junction of the midgut sphincter (Arnone et al., 2006). One of the difficulties with examining boundary levels between Xlox and Cdx is the presence of three functionally redundant Cdx genes in vertebrates. This was partly addressed in zebrafish, where *Cdx4* null/MO-Cdx1a fish displayed an expansion of Pdx1, indicating a regulatory connection between these ParaHox genes (Kinkel et al., 2008). Altogether taking into account all the ParaHox genes of a single organism, an abutting boundary between Xlox and Cdx was probably the ancestral condition for at least all deuterostomes.

RA regulates early neural expression of AmphiGsx

Unlike the relatively late RA endodermal sensitivity of *AmphiXlox/AmphiCdx*, the early domain of *AmphiGsx* is RA responsive. The lack of change in late *AmphiGsx* expression in the CV is consistent with the hypothesis that gene expression in the amphioxus CV is protected against the influence of RA signalling, possibly through the activity of competitive RAR/RXR inhibitors, such as AmphiTR2/4 (Schubert et al., 2006b). Exogenous RA shifts the early *AmphiGsx* domain anteriorly and induces additional *AmphiGsx*-expressing cells, whereas BMS009 abolishes *AmphiGsx* expression. This loss of detectable expression in response to BMS009 treatment also occurs in epidermal sensory cells for *AmphiERR* and *AmphiHox1, 3, 4* and *6* (Schubert et al., 2004). AmphiHox genes are also expressed in the developing neural tube, with *AmphiHox1-3* expression overlapping the earlier A*mphiGsx* expression (Wada et al., 1999; Schubert et al., 2006b). Indeed, *AmphiHox3* has an anterior boundary that approximately coincides with the anterior boundary of *AmphiGsx*. All of the Hox genes examined so far in the amphioxus neural tube are up-regulated and shifted anteriorly by RA, in a similar manner to *AmphiGsx*, in addition to being down-regulated and shifted posteriorly by BMS009 (Schubert et al., 2006b). Various studies have also suggested that *AmphiHox1* and *AmphiHox3* are likely to be directly regulated by RA (Manzanares et al., 2000; Schubert et al., 2005; Schubert et al., 2006b; Wada et al., 2006) and perhaps the RA-induced anterior expansion of *AmphiGsx* expression is mediated by one of these Hox genes. The loss of the early *AmphiGsx* expression domain in BMS009-treated embryos could also be explained by the reduction and posterior shift of Hox expression in these embryos, as AmphiHox protein levels may have dropped below the threshold required to activate *AmphiGsx*. Intriguingly, a putative Hoxb1/Pbx binding site (AGATGGATGG) (Popperl et al., 1995) is located 1916 bp upstream of *AmphiGsx* (data not shown). In contrast to the *AmphiGsx* anterior limit, the posterior boundary shifts rostrally by only a single

somite length, indicating the Hox gene pathway is unlikely to be specifying this posterior limit, as we would expect to see a more dramatic rostral shift after treatment with excess RA. Thus, there appear to be separate regulatory mechanisms for the anterior and posterior limits of *AmphiGsx* expression as these boundaries shift differentially in response to RA. We cannot however exclude the possibility that *AmphiGsx* is directly regulated by RAR/RXR and under this scenario the decrease of *AmphiRAR* expression levels in BMS009-treated embryos could explain the loss of *AmphiGsx* expression.

It is intriguing that only *AmphiGsx* is regulated by RA in the neural tube despite all three ParaHox genes being expressed in this tissue. This implies the regulatory apparatus of *AmphiXlox* and *AmphiCdx* is modular, with an endodermal regulatory module and a neural tube regulatory module. Under this mechanism only the endodermal modules of both genes will include a RA responsive pathway. The lack of neural *AmphiXlox* RA responsiveness is especially important as *AmphiXlox* and *AmphiGsx* expression overlap in wild-type embryos. After treatment with RA however, *AmphiGsx* shifts anteriorly whilst neural *AmphiXlox* expression remains unaffected, thereby reducing the likelihood of shared *AmphiGsx/AmphiXlox* enhancers constraining the chordate ParaHox cluster.

Direct regulation of ParaHox genes by RA

We postulate that at least some of the RA mediated effects on ParaHox expression are mediated directly by the RAR/RXR heterodimer. Consistent with this suggestion, we show by gel shift experiments that the amphioxus RAR/RXR heterodimer weakly binds to a putative RARE located between *AmphiGsx* and *AmphiXlox*. In addition, this binding site alone or an 8 kb genomic region surrounding this site both mediate transcriptional activation of a reporter gene upon RA stimulation. Although the induction was modest, these in vitro and cell transfection results imply the possibility of amphioxus RAR/RXR binding within the ParaHox cluster. The identification of this potential direct RA regulation on the ParaHox cluster is important for understanding both the regulation of the ParaHox genes and some of the constraints on the cluster. It should be noted that the amphioxus RAR/RXR heterodimer may not be directly activating ParaHox gene expression, but instead could be required for the remodelling of the chromatin, thus allowing other transcription factors access to the ParaHox cluster and hence allowing RA responsive regulation to occur. There may also be additional RAREs contributing to the regulation of the amphioxus ParaHox cluster that were not detected by our bioinformatic searches. However, if there are additional RAREs within the ParaHox cluster they must be highly divergent from the consensus RARE sequence.

One final alternative scenario that should not be ignored is the possibility of 9-*cis*-RA mediating this effect. 9-*cis*-RA is easily converted from all-*trans* RA and is likely to be present in small amounts during exogenous RA treatments with all-*trans* RA. Also, amphioxus contains relatively high levels of endogenous 9-*cis*-RA, though still at half the level of all-*trans* RA (Dalfo et al., 2002). The physiological role of 9-*cis*-RA in vertebrates is still unclear, however, and also needs to be established in amphioxus (reviewed in Simões-Costa et al., 2008). It has recently been shown that amphioxus RXR can both bind and be activated by 9-*cis* RA, albeit with lower efficiency than vertebrate RXRs (Tocchini-Valentini et al., 2008), and it is thus conceivable that the RA signal in chordates might be transduced by as yet undefined RAREs that would have been missed in the present work.

Evolution of ParaHox regulation by RA

Studies of RA regulation of ParaHox genes from vertebrates are limited. No studies have examined Gsx in response to RA and direct regulation of Xlox by RA has never been identified. RA does however induce *Pdx1* in mouse ES cells (Micallef et al., 2005; Shiraki et al., 2008), and RA from mesoderm explants is sufficient to induce Pdx1 in anterior endoderm (Kumar et al., 2003). In addition, reduction of RA in both mice and zebrafish results in a loss of *Pdx1* expression in the dorsal pancreatic anlage, and replacement of RA restores mouse *Pdx1* expression (Stafford and Prince, 2002; Martin et al., 2005; Molotkov et al., 2005). In contrast to Xlox and Gsx, RA has been demonstrated to directly regulate both mouse and chicken *Cdx1/CdxA* genes through an atypical RARE upstream of mouse *Cdx1* and also through a conserved RARE within the intron of tetrapod *Cdx1/CdxA* genes (Houle et al., 2000; Gaunt et al., 2003; Houle et al., 2003). RA also causes a reduction of anterior *Cdx4* expression in mouse neural tissue and mesoderm (endodermal *Cdx4* expression was not examined) (Iulianella et al., 1999) and mouse *Cdx2* may also be slightly reduced (Roelen et al., 2002). Clear similarities exist for RA regulation between *AmphiCdx* and tetrapod *Cdx2* and *Cdx4* genes, with all of these genes being down-regulated in response to RA. However, changes in the regulation of Cdx by RA have obviously occurred, as *AmphiCdx* is not induced like the mouse *Cdx1* genes; and the zebrafish *Cdx4* gene appears to confer immunity against RA signalling in posterior endoderm (Kinkel et al., 2008). Moreover, the RAREs controlling the RA response of mouse and chicken Cdx1/CdxA genes are not conserved in the amphioxus ParaHox cluster (data not shown).

Here we have demonstrated all three amphioxus ParaHox genes are regulated by RA. These results are the first demonstration of all three ParaHox genes being affected by RA in a single species. In fact, no signalling pathway has previously been shown to affect the complete set of ParaHox genes from any animal. We also provide evidence that some of this regulation may be directly mediated through a RARE in the ParaHox cluster. Although there are differences between the regulation of vertebrate and amphioxus ParaHox genes by RA, it is likely that RA is upstream of all three ParaHox genes in both taxa (although vertebrate Gsx genes clearly need to be examined in more detail). Quite probably then, at the origin of the chordates the ParaHox cluster was already regulated by RA. It is clear however, that a substantial amount of work on chordate ParaHox regulation is still required, with consideration of the ParaHox cluster as a whole being paramount. It is also intriguing that the evolutionary sister of the ParaHox cluster (the Hox cluster) is also regulated by RA in both amphioxus and vertebrates (Wada et al., 2006). Recent discoveries of RAR genes from protostomes (Campo-Paysaa et al., 2008) have also pushed back the likely origin of RA signalling in animal evolution, allowing the intriguing hypothesis that perhaps RA regulated the original ProtoHox cluster. Further work on the control of ParaHox and Hox genes by RA within the animal kingdom, should establish whether regulation by RA signalling is a conserved trait of these sister gene clusters.

Acknowledgements

The authors are indebted to John M. Lawrence from the University of South Florida, USA, for providing laboratory space during the amphioxus spawning season and to Linda Z. Holland for help during amphioxus field work. The authors would also like to acknowledge Gérard Triqueneaux and Mathilde Paris for technical assistance with gel shift experiments, Ben Raymond for advice on statistical methods, Jordi

Garcia-Fernández for providing the *AmphiCdx* clone and Hinrich Gronemeyer for providing the BMS009 compound. This work was supported by grants from the BBSRC to DEKF, the ANR and the CNRS to MS, the MENRT to MS and VL as well as by CRESCENDO, a European Union Integrated Project of FP6. PWO is funded by the Clarendon, ORS and EPA Cephalosporin scholarships. The authors would like to thank the anonymous reviewers for their comments.

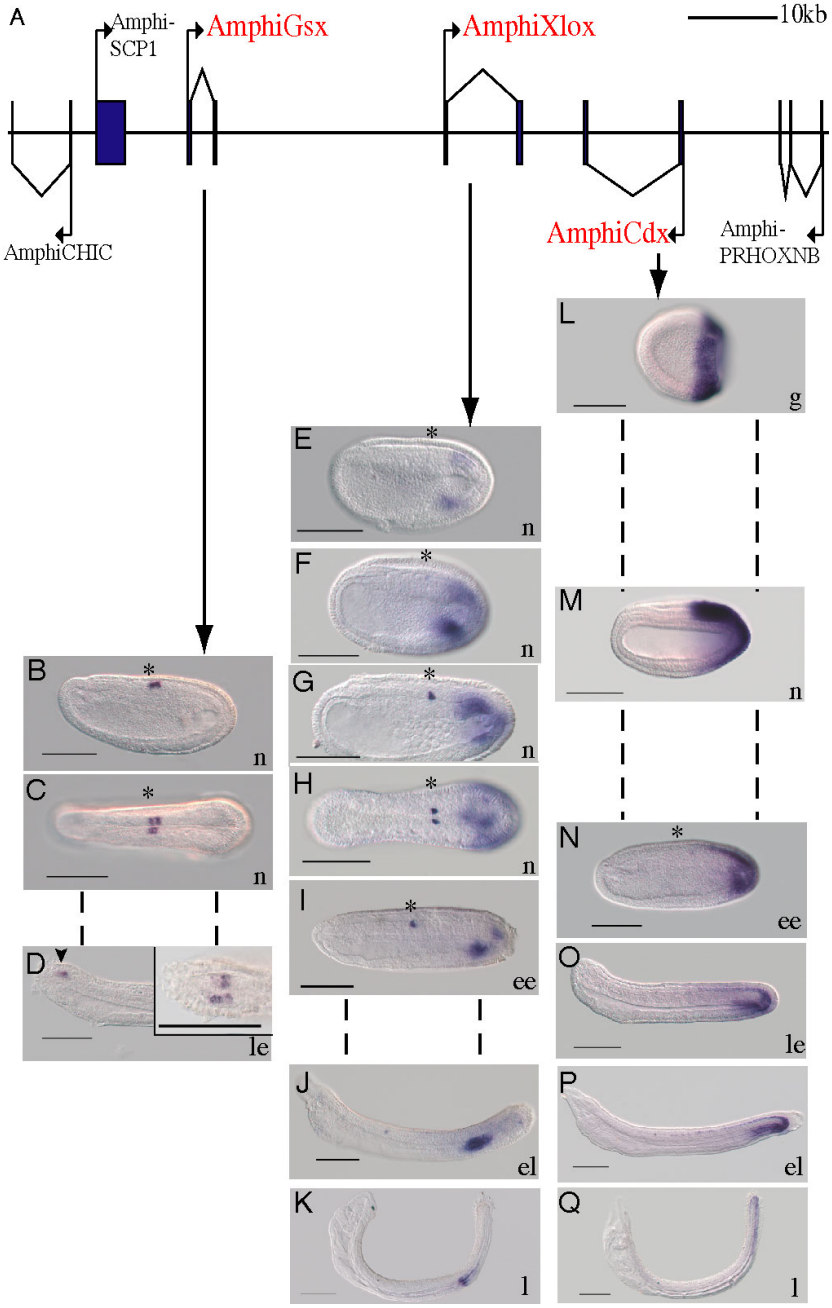


Fig. 1. Wild-type expression of the amphioxus ParaHox genes is related to the genomic organisation (A) of the ParaHox gene cluster. *AmphiGsx* (B-D) is expressed most anteriorly, *AmphiXlox* (E-K) is expressed more centrally and *AmphiCdx* (L-Q) is expressed most posteriorly. *AmphiCdx* is the first gene to be detectable during gastrulation around the closing blastopore (L) and then remains in a continuous domain in the posterior of the animal in the neural tube, hindgut and tail bud (M-Q). *AmphiXlox* expression commences slightly later than *AmphiCdx* in the posterior endoderm in two distinct domains, (dorsally and ventrally) (E-I) but later becomes restricted to a more central region of the gut in the developing larva (J, K). There is also transient expression in two neural tube cells in a region coinciding with the position of the future first pigment spot (G-I). The last ParaHox gene to be activated is *AmphiGsx*, which is initially expressed in the neural tube at the same level as the neural domain of *AmphiXlox* (B-C). Subsequently, *AmphiGsx* expression is down-regulated in this early domain and is activated later in the cerebral vesicle (D). The inset in D is a magnified dorsal view of the embryo at the level of the arrowhead. Embryos are presented as side views with anterior to the left except for dorsal views of the embryos in C and H. The anterior/posterior position of the boundary between somites 4 and 5 is indicated by an asterisk. Lowercase lettering denotes the developmental stage: g, gastrula; n, neurula, ee, early embryo; le, late embryo; el, early larvae; l, larvae, with similarly aged embryos/larvae aligned horizontally. Vertical dashed lines denote a continuation of the same expression pattern shown in the panels. Scale bars represent 100 µm.

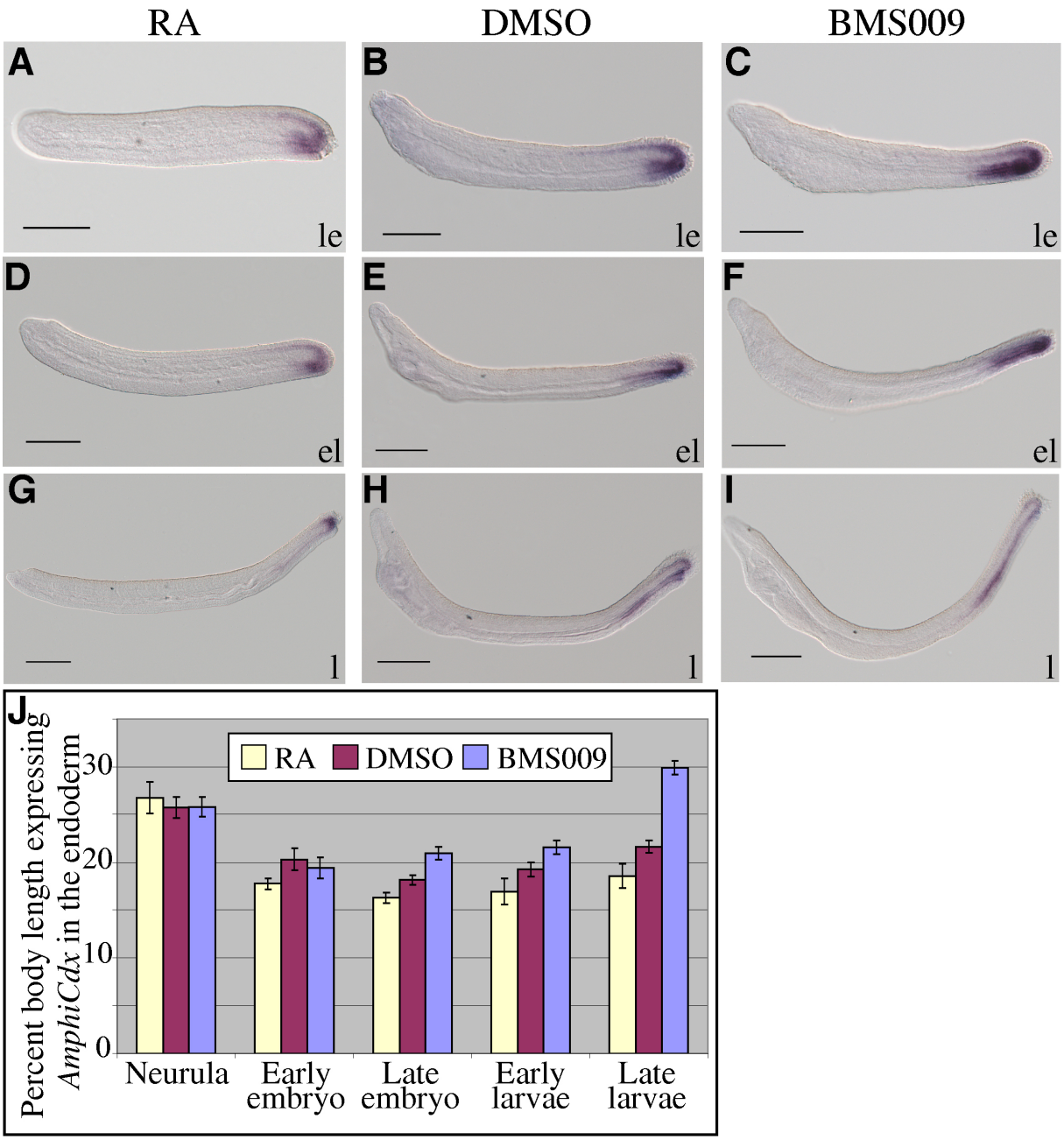


Fig. 2. Retinoic acid (RA) regulates the expression of *AmphiCdx* in the posterior endoderm. RA causes a posterior compression of the endodermal expression (A, D, G) relative to control embryos (B, E, H), whereas treatment with the RA antagonist BMS009 expands the *AmphiCdx* domain anteriorly (C, F, I). Embryos are presented as side views with anterior to the left. Lowercase lettering denotes the developmental stage: le, late embryo; el, early larvae; l, larvae. Scale bars represent 100 µm. Expression was not affected by treatment at early stages of development (J) with one-way ANOVA p-values of 0.86 and 0.24 for neurulae and early embryo, respectively. A significant difference exists between treatments at all other stages with p-values of less than 0.005. Error bars are standard error of the mean.

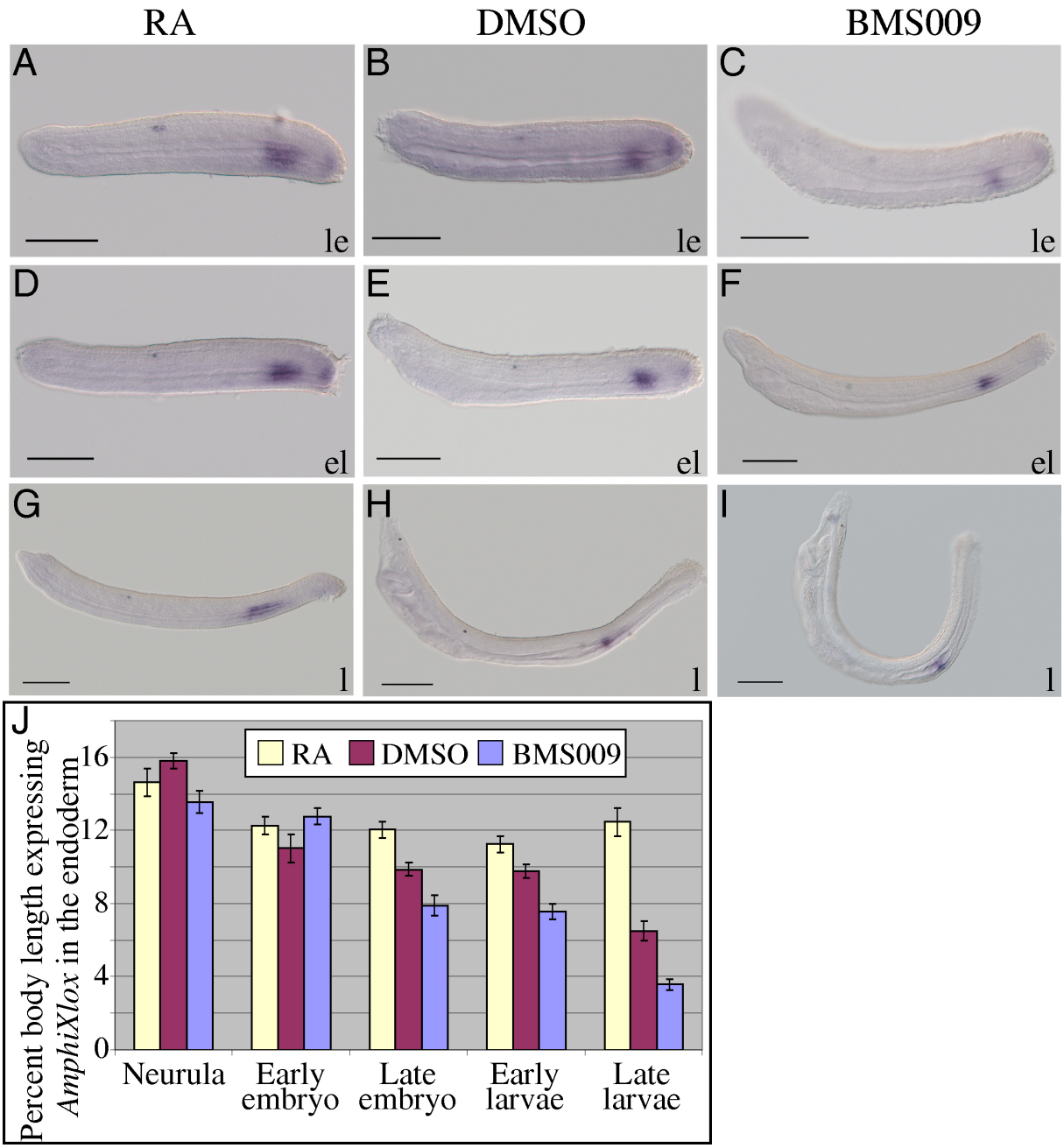


Fig. 3. Retinoic acid (RA) regulates the expression of *AmphiXlox* in the posterior endoderm. RA causes an expansion caudally of the posterior limit of endodermal *AmphiXlox* expression (A, D, G) relative to control embryos (B, E, H), whereas BMS009 treatment reduces the endodermal *AmphiXlox* domain posteriorly (C, F, I). Embryos are presented as side views with anterior to the left. Lowercase lettering denotes the developmental stage: le, late embryo; el, early larvae; l, larvae. Scale bars represent 100 μm. Expression was not affected by treatment at early stages of development (J) with one-way ANOVA p-values of 0.04 and 0.27 for mid-neurulae and late neurulae, respectively (the weak significance at the mid-neurula stage is due to the high variability amongst BMS009-treated embryos). A highly significant difference exists between treatments at all other stages with p-values of less than 0.0001. Error bars are standard error of the mean.

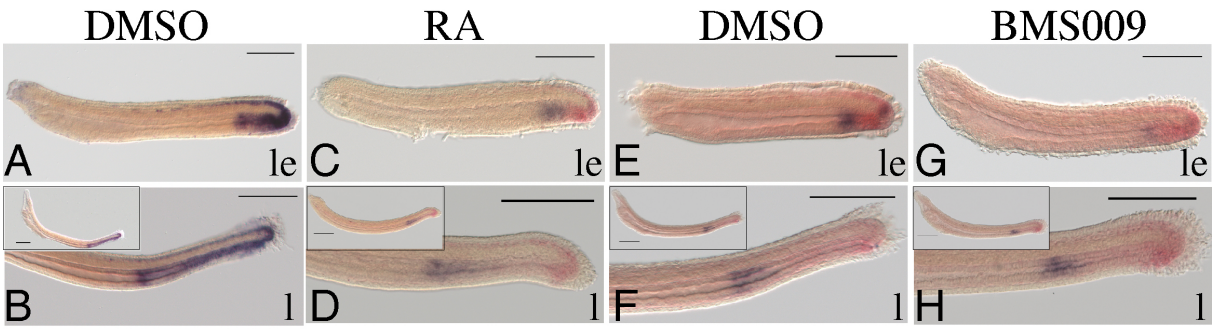


Fig. 4. Retinoic acid (RA) regulates the expression boundary between *AmphiXlox* and *AmphiCdx* in the posterior endoderm. Amphioxus larvae subjected to single colour double in situ hybridisations against both *AmphiCdx* and *AmphiXlox* reveal the expression domains are either overlapping or adjacent to each other,with no gap between the expression domains (A, B). Two colour double in situ hybridisations on early larvae show that the two genes have adjacent domains from the stages shown (E, F). Once the boundary between *AmphiXlox* (blue staining) and *AmphiCdx* (red staining) has formed in the late embryo, this boundary becomes responsive to treatment with RA or BMS009 (C, D, G, H). The images in B, D, F, H are views of the posterior half of the larvae displayed in the insets. Embryos are presented as side views with anterior to the left. Lowercase lettering denotes the developmental stage: le, late embryo; l, larvae. Scale bars represent 100 μm.

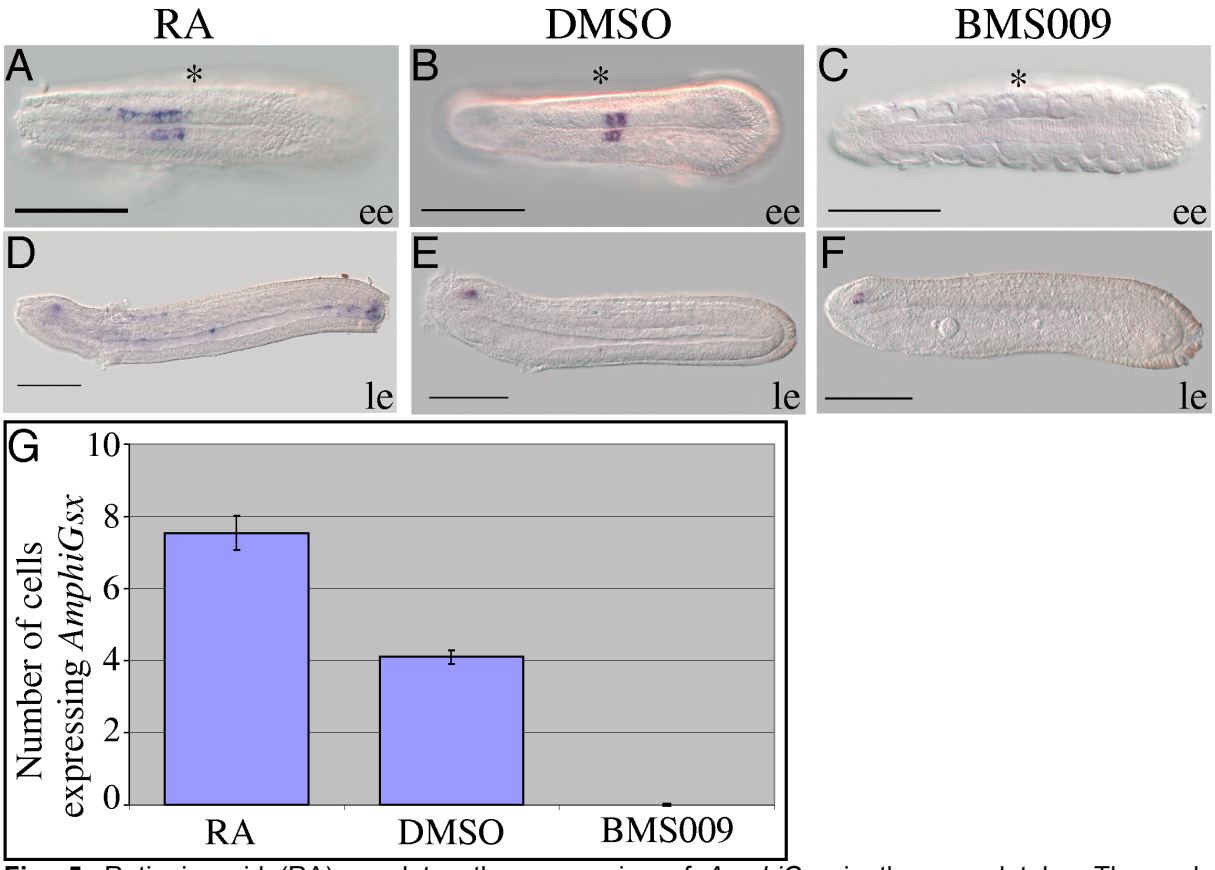


Fig. 5. Retinoic acid (RA) regulates the expression of *AmphiGsx* in the neural tube. The early *AmphiGsx* expression domain corresponds to four cells in the neural tube at the level of somite five (B). This domain is dramatically shifted and expanded anteriorly by treatment with RA (A) and reduced below the level of detection by in situ hybridisation by treatment with BMS009 (C). The embryo in C is at a different focal plane to those in A and B to display the somites and hence the stage of the embryo. The anterior/posterior position of the boundary between somites 4 and 5 is indicated by an asterisk. Treatment does not affect the later cerebral vesicle expression domain (D-F). Embryos are presented with anterior to the left, embryos in A-C are dorsal and in D-F are side views. Lowercase lettering denotes the developmental stage: ee, early embryo; le, late embryo. Scale bars represent 100 µm. Treatment affects the number of cells expressing *AmphiGsx* in the early domain causing a significant increase after RA treatment and a reduction after BMS009 treatment (one-way ANOVA p-value of less than 0.001) (G). Error bars are standard error of the mean.

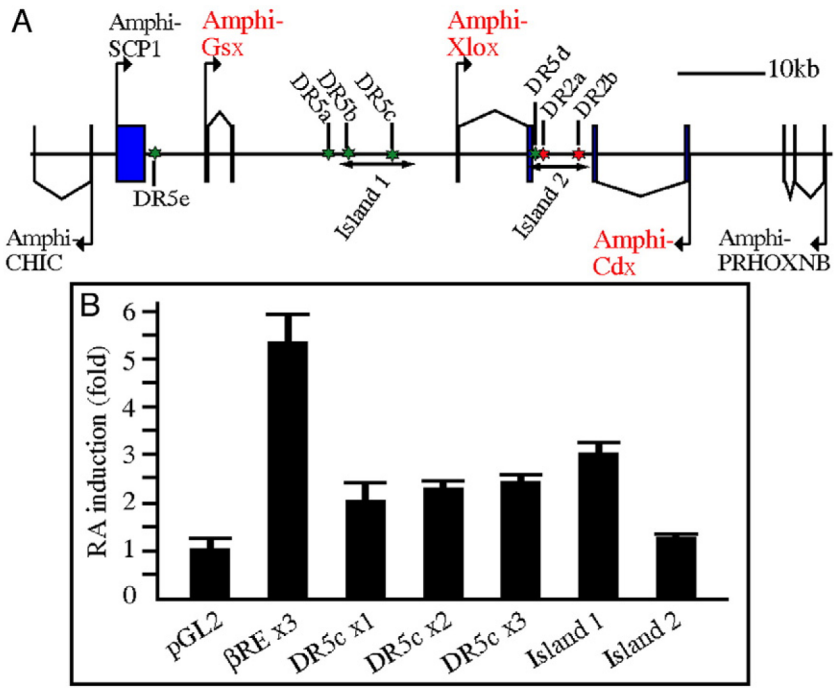


Fig. 6. Identification (A) and analysis (B) of putative retinoic acid response elements (RAREs), DR5a-e and DR2a-b, in the amphioxus ParaHox cluster. These putative RAREs are clustered in two islands, Island 1 upstream of *AmphiXlox* and Island 2 between *AmphiXlox* and *AmphiCdx*. In a heterologous cell culture transactivation assay, DR5c weakly activates transcription of a reporter gene after RA stimulation both as a single copy construct (DR5cx1) and when tested as a double (DR5cx2) or a triple (DR5cx3) copy construct. A construct containing the majority of Island 1 (including DR5b and DR5c) also weakly induces the reporter gene, while a construct of Island 2 (including DR5d, DR2a and DR2b) is less efficient in activating reporter gene expression. Empty pGL2-promoter vector was used as a negative control. A vector containing three copies of the RARE found in the human *RARβ2* regulatory region (βREx3) was used as a positive control for RA-stimulated, amphioxus RAR/RXR-dependent transcriptional activation in the cell culture transactivation assays. Error bars are standard error of the mean.

References

Arnone, M. I., Rizzo, F., Annunciata, R., Cameron, R. A., Peterson, K. J. and Martinez, P. (2006). Genetic organization and embryonic expression of the ParaHox genes in the sea urchin *S. purpuratus*: insights into the relationship between clustering and colinearity. *Developmental Biology* **300**, 63-73.

Bastien, J. and Rochette-Egly, C. (2004). Nuclear retinoid receptors and the transcription of retinoid-target genes. *Gene* **328**, 1-16.

Brooke, N. M., Garcia-Fernàndez, J. and Holland, P. W. H. (1998). The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. *Nature* **392**, 920-922.

Campo-Paysaa, F., Marlétaz, F., Laudet, V. and Schubert, M. (2008). Retinoic acid signaling in development: tissue-specific functions and evolutionary origins. *Genesis* **46**, 640-656.

Cañestro, C., Albalat, R., Escriva, H. and Gonzalez-Duarte, R. (2001). Endogenous β-galactosidase activity in amphioxus: a useful histochemical marker for the digestive system. *Development Genes and Evolution* **211**, 154-156.

Coulier, F., Burtey, S., Chaffanet, M., Birg, F. and Birnbaum, D. (2000). Ancestrally-duplicated paraHOX gene clusters in humans. *International Journal of Oncology* **17**, 439-444.

Dalfo, D., Albalat, R., Molotkov, A., Duester, G. and Gonzalez-Duarte, R. (2002). Retinoic acid synthesis in the prevertebrate amphioxus involves retinol oxidation. *Development Genes and Evolution* **212**, 388-393.

Duboule, D. (2007). The rise and fall of Hox gene clusters. *Development* **134**, 2549-2560.

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**, 2905-2916.

Fang, R. X., Olds, L. C. and Sibley, E. (2006). Spatio-temporal patterns of intestine-specific transcription factor expression during postnatal mouse gut development. *Gene Expression Patterns* **6**, 426-432.

Ferrier, D. E., Dewar, K., Cook, A., Chang, J. L., Hill-Force, A. and Amemiya, C. (2005). The chordate ParaHox, cluster. *Current Biology* **15**, R820-R822.

Ferrier, D. E. K. and Holland, P. W. H. (2001a). Sipunculan ParaHox genes. *Evolution & Development* **3**, 263-270.

Ferrier, D. E. K. and Holland, P. W. H. (2001b). Ancient origin of the Hox gene cluster. *Nature Reviews Genetics* **2**, 33-38.

Ferrier, D. E. K. and Holland, P. W. H. (2002). *Ciona intestinalis* ParaHox genes: evolution of Hox/ParaHox cluster integrity, developmental mode, and temporal colinearity. *Molecular Phylogenetics and Evolution* **24**, 412-417.

Ferrier, D. E. K. and Minguillon, C. (2003). Evolution of the Hox/ParaHox gene clusters. *International Journal of Developmental Biology* **47**, 605-611.

Finnerty, J. R. and Martindale, M. Q. (1999). Ancient origins of axial patterning genes: Hox genes and ParaHox genes in the Cnidaria. *Evolution & Development* **1**, 16-23.

Furlong, R. F., Younger, R., Kasahara, M., Reinhardt, R., Thorndyke, M. and Holland, P. W. (2007). A Degenerate ParaHox Gene Cluster in a Degenerate Vertebrate. *Molecular Biology and Evolution*. **24**, 2681-2686.

Gaunt, S. J., Drage, D. and Cockley, A. (2003). Vertebrate caudal gene expression gradients investigated by use of chick *cdx-A/lacZ* and mouse *cdx-1/lacZ* reporters in transgenic mouse embryos: evidence for an intron enhancer. *Mechanisms of Development* **120**, 573-586.

Holland, L. Z. and Holland, N. D. (1996). Expression of *AmphiHox-1* and *AmphiPax-1* in amphioxus embryos treated with retinoic acid: insights into evolution and patterning of the chordate nerve cord and pharynx. *Development* **122**, 1829-1838.

Holland, L. Z., Holland, P. W. H. and Holland, N. D. (1996). Revealing homologies between body parts of distantly related animals by *in situ* hybridization to developmental genes: amphioxus versus vertebrates. In *Molecular Zoology. Advances, Strategies, and Protocols* (eds J. D. Ferraris and S. R. Palumbi), pp. pp. 267-282; 473-483. New York, NY: Wiley-Liss.

Horb, M. E. and Slack, J. M. (2001). Endoderm specification and differentiation in *Xenopus* embryos. *Developmental Biology* **236**, 330-343.

Houle, M., Prinos, P., Iulianella, A., Bouchard, N. and Lohnes, D. (2000). Retinoic acid regulation of Cdx1: an indirect mechanism for retinoids and vertebral specification. *Molecular and Cellular Biology* **20**, 6579-6586.

Houle, M., Sylvestre, J. R. and Lohnes, D. (2003). Retinoic acid regulates a subset of Cdx1 function in vivo. *Development* **130**, 6555-6567.

Iulianella, A., Beckett, B., Petkovich, M. and Lohnes, D. (1999). A molecular basis for retinoic acid-induced axial truncation. *Developmental Biology* **205**, 33-48.

Kinkel, M. D., Eames, S. C., Alonzo, M. R. and Prince, V. E. (2008). Cdx4 is required in the endoderm to localize the pancreas and limit β-cell number. *Development* **135**, 919-929.

Kumar, M., Jordan, N., Melton, D. and Grapin-Botton, A. (2003). Signals from lateral plate mesoderm instruct endoderm toward a pancreatic fate. *Developmental Biology* **259**, 109-122.

Langeland, J. A., Tomsa, J. M., Jackman, W. R. and Kimmel, C. B. (1998). An amphioxus *snail* gene: expression in paraxial mesoderm and neural plate suggests a conserved role in patterning the chordate embryo. *Development Genes and Evolution* **208**, 569-577.

Lickert, H. and Kemler, R. (2002). Functional analysis of *cis*-regulatory elements controlling initiation and maintenance of early *Cdx1* gene expression in the mouse. *Developmental Dynamics* **225**, 216-220.

Maden, M. (2007). Retinoic acid in the development, regeneration and maintenance of the nervous system. *Nature Reviews Neuroscience* **8**, 755-765.

Manzanares, M., Wada, H., Itasaki, N., Trainor, P. A., Krumlauf, R. and Holland, P. W. H. (2000). Conservation and elaboration of *Hox* gene regulation during evolution of the vertebrate head. *Nature* **408**, 854-857.

Mark, M., Ghyselinck, N. B. and Chambon, P. (2006). Function of retinoid nuclear receptors: lessons from genetic and pharmacological dissections of the retinoic acid signaling pathway during mouse embryogenesis. *Annual Review of Pharmacology and Toxicology* **46**, 451-480.

Marletaz, F., Holland, L. Z., Laudet, V. and Schubert, M. (2006). Retinoic acid signaling and the evolution of chordates. *International Journal of Biological Sciences* **2**, 38-47.

Martin, M., Gallego-Llamas, J., Ribes, V., Kedinger, M., Niederreither, K., Chambon, P., Dolle, P. and Gradwohl, G. (2005). Dorsal pancreas agenesis in retinoic acid-deficient *Raldh2* mutant mice. *Developmental Biology* **284**, 399-411.

Micallef, S. J., Janes, M. E., Knezevic, K., Davis, R. P., Elefanty, A. G. and Stanley, E. G. (2005). Retinoic acid induces Pdx1-positive endoderm in differentiating mouse embryonic stem cells. *Diabetes* **54**, 301-305.

Molotkov, A., Molotkova, N. and Duester, G. (2005). Retinoic acid generated by *Raldh2* in mesoderm is required for mouse dorsal Endodermal pancreas development. *Developmental Dynamics* **232**, 950-957.

Mulley, J. F., Chiu, C. H. and Holland, P. W. H. (2006). Breakup of a homeobox cluster after genome duplication in teleosts. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 10369-10372.

Perlmann, T., Rangarajan, P. N., Umesono, K. and Evans, R. M. (1993). Determinants for selective RAR and TR recognition of direct repeat HREs. *Genes & Development* **7**, 1411-1422.

Pilon, N., Oh, K., Sylvestre, J. R., Savory, J. G. and Lohnes, D. (2007). Wnt signaling is a key mediator of *Cdx1* expression in vivo. *Development* **134**, 2315-2323.

Popperl, H., Bienz, M., Studer, M., Chan, S. K., Aparicio, S., Brenner, S., Mann, R. S. and Krumlauf, R. (1995). Segmental expression of *Hoxb-1* is controlled by a highly conserved autoregulatory loop dependent upon *exd/pbx*. *Cell* **81**, 1031-1042.

Prohaska, S. J. and Stadler, P. F. (2006). Evolution of the vertebrate ParaHox clusters. *Journal of Experimental Zoology Part B - Molecular and Developmental Evolution* **306B**, 481-487.

Rebeiz, M. and Posakony, J. W. (2004). GenePalette: a universal software tool for genome sequence visualization and analysis. *Developmental Biology* **271**, 431-438.

Roelen, B. A., de Graaff, W., Forlani, S. and Deschamps, J. (2002). *Hox* cluster polarity in early transcriptional availability: a high order regulatory level of clustered *Hox* genes in the mouse. *Mechanisms of Development* **119**, 81-90.

Sandelin, A. and Wasserman, W. W. (2005). Prediction of nuclear hormone response elements. *Molecular Endocrinology* **19**, 595-606.

Schubert, M., Escriva, H., Xavier-Neto, J. and Laudet, V. (2006a). Amphioxus and tunicates as evolutionary model systems. *Trends in Ecology & Evolution* **21**, 269-277.

Schubert, M., Holland, N. D., Laudet, V. and Holland, L. Z. (2006b). A retinoic acid-*Hox* hierarchy controls both anterior/posterior patterning and neuronal specification in the developing central nervous system of the cephalochordate amphioxus. *Developmental Biology* **296**, 190-202.

Schubert, M., Holland, N. D., Escriva, H., Holland, L. Z. and Laudet, V. (2004). Retinoic acid influences anteroposterior positioning of epidermal sensory neurons and their gene expression in a developing chordate (amphioxus). *Proceedings of the National Academy of Sciences of the United States of America* **101**, 10320-10325.

Schubert, M., Yu, J. K., Holland, N. D., Escriva, H., Laudet, V. and Holland, L. Z. (2005). Retinoic acid signaling acts via *Hox1* to establish the posterior limit of the pharynx in the chordate amphioxus. *Development* **132**, 61-73.

Shiraki, N., Yoshida, T., Araki, K., Umezawa, A., Higuchi, Y., Goto, H., Kume, K. and Kume, S. (2008). Guided Differentiation of ES Cells into Pdx1-Expressing Regional Specific Definitive Endoderm. *Stem Cells* **26**, 874-885.

Simoes-Costa, M. S., Azambuja, A. P. and Xavier-Neto, J. (2008). The search for non-chordate retinoic acid signaling: lessons from chordates. *Journal of Experimental Zoology B – Molecular and Developmental Evolution* **310**, 54-72.

Siegel, N., Hoegg, S., Salzburger, W., Braasch, I. and Meyer, A. (2007). Comparative genomics of ParaHox clusters of teleost fishes: gene cluster breakup and the retention of gene sets following whole genome duplications. *BMC Genomics* **8**, 312.

Stafford, D. and Prince, V. E. (2002). Retinoic acid signaling is required for a critical early step in zebrafish pancreatic development. *Current Biology* **12**, 1215-1220.

Tocchini-Valentini, G. D., Rochel, N., Escriva, H., Germain, P., Peluso-Iltis, C., Paris, M., Sanglier-Cianferani, S., Van Dorsselaer, A., Moras, D. and Laudet, V. (2009). Structural and functional insights into the ligand binding domain of a non-duplicated RXR from the invertebrate chordate amphioxus. *Journal of Biological Chemistry* **284**, 1938-1948.

Toresson, H., Potter, S. S. and Campbell, K. (2000). Genetic control of dorsal-ventral identity in the telencephalon: opposing roles for Pax6 and Gsh2. *Development* **127**, 4361-4371.

Tour, E., Pillemer, G., Gruenbaum, Y. and Fainsod, A. (2002). *Gbx2* interacts with *Otx2* and patterns the anterior-posterior axis during gastrulation in *Xenopus*. *Mechanisms of Development* **112**, 141-151.

Tulachan, S. S., Doi, R., Kawaguchi, Y., Tsuji, S., Nakajima, S., Masui, T., Koizumi, M., Toyoda, E., Mori, T., Ito, D. et al. (2003). All-*Trans* retinoic acid induces differentiation of ducts and endocrine cells by mesenchymal/epithelial interactions in embryonic pancreas. *Diabetes* **52**, 76-84.

Vanacker, J. M., Pettersson, K., Gustafsson, J. A. and Laudet, V. (1999). Transcriptional targets shared by estrogen receptor- related receptors (ERRs) and estrogen receptor (ER) α, but not by ERβ. *EMBO Journal* **18**, 4270-4279.

Wada, H., Garcia-Fernàndez, J. and Holland, P. W. H. (1999). Colinear and segmental expression of amphioxus Hox genes. *Developmental Biology* **213**, 131-141.

Wada, H., Escriva, H., Zhang, S. C. and Laudet, V. (2006). Conserved RARE localization in amphioxus *Hox* clusters and implications for *Hox* code evolution in the vertebrate neural crest. *Developmental Dynamics* **235**, 1522-1531.

Wang, Z., Benoit, G., Liu, J., Prasad, S., Aarnisalo, P., Liu, X., Xu, H., Walker, N. P. and Perlmann, T. (2003). Structure and function of Nurr1 identifies a class of ligand-independent nuclear receptors. *Nature* **423**, 555-560.

Table S1. List of putative retinoic acid response elements (RAREs) identified in the amphioxus ParaHox cluster. The putative binding sites of the amphioxus RAR/RXR heterodimer on the forward (+) or reverse (-) strand (relative to the ParaHox gene cluster sequence as displayed in Fig. 6) are indicated in bold and underlined. The location of the first base of the RARE relative to the start of the indicated PAC (see Materials and Methods for Genbank accession numbers) is stated for the forward strand below each sequence. The 5' overhanging restriction sites added for cloning of the putative RAREs are underlined.

RARE name	Forward	Reverse
DR5a	*CGCG*CCGTTC**AGTTCA**CTCCG**AGGCCA**ATTAACC	*GATC*GGTTAAT**TGGCCT**CGGAG**TGAACT**GAACGG
	283 (PAC36D2)	
DR5b	*CGCG*TCGGTC**AGTTCA**CTCCT**AGGCCA**ATTAACT	*GATC*AGTTAAT**TGGCCT**AGGAG**TGAACT**GACCGA
	3218 (PAC36D2)	
DR5c	*CGCG*ACCTGA**AGGTCA**TTTCA**AGGTGA**AGGCCAA	*GATC*TTGGCCT**TCACCT**TGAAA**TGACCT**TCAGGT
	7724 (PAC36D2)	
DR5d	*CGCG*GGCGCG**TGGTCA**AGTCG**AGGGCA**CGGGGTT	*GATC*AACCCCG**TGCCCT**CGACT**TGACCA**CGCGCC
	25005 (PAC36D2)	
DR5e	*CGCG*AATCAC**TGACCC**ACGCA**AGACCC**CATTAGG	*GATC*CCTAATG**GGGTCT**TGCGT**GGGTCA**GTGATT
	14472 (PAC33B4)	
DR2a	*CGCG*AATGTTAG**AGGTCT**TG**AGTTCA**CATCTGTA	*GATC*TACAGATG**TGAACT**CA**AGACCT**CTAACATT
	25845 (PAC36D2)	
DR2b	*CGCG*TACGGTTG**TGAACC**TG**TGCCCT**TCTACACA	*GATC*TGTGTAGA**AGGGCA**CA**GGTTCA**CAACCGTA
	30221 (PAC36D2)	

Table S2. Number of embryos used in the analysis of *AmphiCdx* treated embryos.

	Retinoic Acid	Control	BMS009
Neurula	7	20	15
Early embryo	5	4	6
Late embryo	4	7	11
Early larvae	12	14	21
Larvae	13	18	16

Table S3. Number of embryos used in the analysis of *AmphiXlox* treated embryos.

	Retinoic Acid	Control	BMS009
Neurula	9	29	8
Early embryo	7	5	2
Late embryo	6	6	6
Early larvae	8	13	18
Larvae	16	10	18

Table S4. Number of embryos used in the analysis of *AmphiGsx* treated embryos.

	Retinoic Acid	Control	BMS009
Early embryo	28	31	27

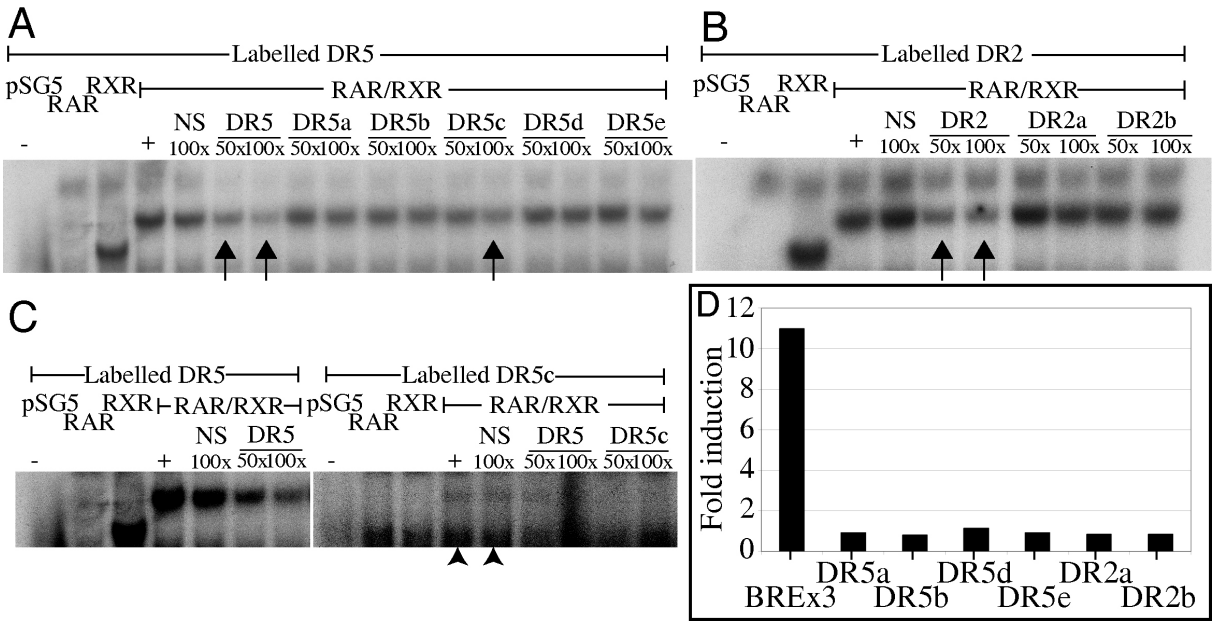


Fig. S1. Electromobility shift assays (A-C) and cell culture transactivation assays (D) to test binding of the amphioxus RAR/RXR heterodimer to putative retinoic acid response elements (RAREs) in the amphioxus ParaHox cluster. Radiolabelled consensus DR5 (A) or DR2 (B) RARE elements were bound to amphioxus RAR/RXR heterodimers and subsequently this binding was competed with non-radiolabelled putative RARE sequences. Only the consensus sequences (DR5 or DR2) and DR5c were able to compete for heterodimer binding (arrows). Weak binding was also observed for radiolabelled DR5c to the amphioxus RAR/RXR heterodimer (arrowheads) (C). Consistent with these gel shift results, the putative RAREs DR5a,b,d,e and DR2a,b do not induce amphioxus RAR/RXR-dependent reporter construct activity upon RA stimulation in cell culture transactivation assays (D). A vector containing three copies of the RARE found in the human *RARβ2* regulatory region (βRE x3) was used as positive control. Error bars are standard error of the mean.

Retinoic Acid Signaling Targets *Hox* Genes during the Amphioxus Gastrula Stage: Insights into Early Anterior-Posterior Patterning of the Chordate Body Plan

Demian Koop, Nicholas D. Holland, Marie Sémon, Susana Alvarez, Angel R. de Lera, Vincent Laudet, Linda Z. Holland, Michael Schubert

Abstract

Previous developmental studies of vertebrate development have shown that retinoic acid (RA) signaling at the gastrula stage strongly influences anterior-posterior (A-P) patterning of the neurula and later stages. However, much less is known about the more immediate effects of RA signaling on gene transcription and developmental patterning at the gastrula stage. To investigate the targets of RA signaling during the gastrula stage, we used the basal chordate amphioxus, in which gastrulation involves very minimal tissue movements. First, we determined the effect of altered RA signaling on expression of 42 genes (encoding transcription factors and components of major signaling cascades) known to be expressed in restricted domains along the A-P axis during the gastrula and early neurula stage. Of these 42 genes, the expression domains during gastrulation of only four (*Hox1, Hox3, HNF3-1* and *Wnt3*) were spatially altered by exposure of the embryos to excess RA or to the RA antagonist BMS009. Moreover, blocking protein synthesis with puromycin before adding RA or BMS009 showed that only three of these genes (*Hox1, Hox3* and *HNF3-1*) are direct RA targets at the gastrula stage. From these results we conclude that in the amphioxus gastrula RA signaling primarily acts via regulation of *Hox* transcription to establish positional identities along the A-P axis and that *Hox1, Hox3*, *HNF3-1* and *Wnt3* constitute a basal module of RA action during chordate gastrulation.

Introduction

Retinoic acid (RA) is a morphogen that acts via heterodimers of the retinoic acid receptor (RAR) and retinoid X receptor (RXR), which bind to retinoic acid response elements (RAREs) in the promoter regions of target genes. The RA signal is mediated by a positive feedback loop involving the direct regulation of *RAR* by RAR/RXR heterodimers (Heyman et al., 1992; Rudert and Gronemeyer, 1993; Blomhoff and Blomhoff, 2006; Campo-Paysaa et al., 2008; Casci, 2008). RA has been extensively studied for its effects on cell cultures, embryonic development, adult growth, regeneration and carcinogenesis in chordates (McCaffery et al., 2003; Mongan and Gudas, 2007; Dann et al., 2008; Niederreither and Dollé, 2008). In embryos of amphioxus and vertebrates, RA has numerous pleiotropic effects. RA signaling is permissive for some tissues—like forelimbs and somites—allowing previously specified structures to complete differentiation, but instructive for other tissues—like hindbrain and foregut—conferring positional information along the anterior-posterior (A-P) axis and specifying tissue identity (Stafford et al., 2006; Duester, 2008).

The earliest known instructive signaling by RA begins at the gastrula stage and mediates patterning of the germ layers along the A-P axis of the embryo (Durston et al., 1989; Sive et al., 1990; Holland and Holland, 1996; Roelen et al., 2002; Grapin-Botton, 2005; White et al., 2007). Even a transitory perturbation of RA signaling in the gastrula can affect A-P patterning during the post-gastrula stages of development. There are few studies on the effects RA signaling on gene transcription at the gastrula stage, and these are largely limited to *Hox* genes and to vertebrates (Kudoh et al., 2002; Roelen et al., 2002; Li et al., 2008a). Such studies are complicated by the mechanics of gastrulation in vertebrates. The germ layers at the gastrula stage are often more than one cell thick (Delarue et al., 1998; Li et al., 2008b), and the constituent cells undergo complex migrations, either individually or in coherent groups during gastrulation (Schoenwolf and Smith, 2000; Kimura et al., 2006). It is, therefore, difficult to establish the A-P limits of the domains of gene expression in vertebrate gastrulae, raising the possibility that there are early immediate targets of RA signaling at the gastrula stage that act in parallel to *Hox* genes in mediating A-P patterning.

To test this hypothesis, we used the invertebrate chordate amphioxus. Amphioxus resembles vertebrates in using RA signaling for axial patterning (Holland and Holland, 1996; Escriva et al., 2002; Marlétaz et al., 2006), but has the advantage of early development that is morphologically uncomplicated: the spherical blastula has a single layer of cells surrounding a hollow blastocoel. Gastrulation begins with invagination from the posterior pole of the embryo. The resulting gastrula has an outer ectoderm and an inner mesendoderm, both only one cell thick. The invaginated mesendoderm largely obliterates the blastocoel and creates a new space, the archenteron, which opens to the exterior via a blastopore. Cell movements during invagination of the mesendoderm are minimal: the cells at the blastoporal lip do not converge toward the dorsal midline, and there is little involution over the lip of the blastopore (Zhang et al., 1997; Holland and Holland, 2007). Amphioxus is also more amenable for such a study than tunicates, where RA signaling of the A-P axis is either highly modified (ascidians) or absent (appendicularians) (Ishibashi et al., 2003; Nagatomo and Fujiwara, 2003; Fujiwara, 2005; Cañestro and Postlethwait, 2007; Imai et al., 2009).

For vertebrates, except for the study of Kudoh et al. (2002), the effects of altering RA signaling during the gastrula stage have not been tallied until the neurula and larval stages (reviewed by Duester, 2008). Similarly, for amphioxus, previous studies altering RA signaling during the gastrula stage only looked at the effects of this treatment at later stages. Therefore, although these studies documented effects of RA on expression of a number of genes (i.e. several *Hox* genes, *Cdx*, *Hedgehog*, *HNF3-1*, *Nodal*, *Notch*, *Otx*, *Pitx*, *Wnt3* and *Wnt5*) at the neurula and larval stages (Holland and Holland, 1996; Schubert et al., 2004, 2005, 2006; Osborne et al., 2009), they could not differentiate direct targets of RA signaling from indirect ones.

In the present study, we administered RA or an RA antagonist (BMS009) continuously to cultures of developing amphioxus from the beginning of the gastrula stage and assayed for effects on expression of 42 genes later in the gastrula stage known to be transcribed in restricted anterior-posterior patterns at the gastrula stage (listed in Supplementary Table S1). In addition, we inhibited protein synthesis with puromycin to distinguish between direct and indirect targets of RA signaling. Our results show that only 3 of these 42 genes (*Hox1*, *Hox3* and *HNF3-1*) are direct targets of RA at the gastrula stage, while only one (*Wnt3*) is an indirect target. Moreover, we present evidence suggesting that, in the amphioxus early neurula, RA may also directly regulate *Hox4* and *Hox6*. Since, in amphioxus, the control regions of at least *Hox1* and *Hox3* include functional RAREs (Manzanares et al., 2000; Wada et al., 2006), we conclude that RA acts primarily through *Hox* genes as direct targets in patterning the A-P body axis at the gastrula stage not only in amphioxus, but probably also in the last invertebrate chordate ancestor of vertebrates.

Materials and Methods

Sexually mature males and females of the Florida amphioxus (*Branchiostoma floridae*) were collected in Tampa Bay Florida, USA, during the summer breeding season. The animals were stimulated to spawn electrically (Holland and Holland, 1993). After fertilization, the embryos were raised in filtered seawater at 28°C and staged according to Holland and Yu (2004) as very early, early, mid and late gastrulae (respectively, 3 h, 4 h, 5 h and 6 h after fertilization) or early neurulae (9h after fertilization).

In a first series of experiments, RA or the RA antagonist BMS009, each dissolved in DMSO (for a final concentration of $1x10^{-6}$ M), were added to cultures of very early gastrulae. DMSO at a 1:1000 dilution alone (Holland and Holland, 1996; Escriva et al., 2002) had no detectable effect. Samples of gastrulae at 4 h, 5 h or 6 h of development were fixed in 4% paraformaldehyde in MOPS buffer (0.1 M MOPS, 0.5 M NaCl, 2 mM MgSO4, 1 mM EGTA, pH 7.4) (Holland et al., 1996). This fixation solution is referred to hereafter simply as PFA. After fixation overnight at 4°C, the specimens were transferred to 70% ethanol and stored at -20°C until subjected to in situ hybridization. For each gene tested, antisense riboprobes were synthesized according to (Holland et al., 1996) from the originally described clones or matching EST clones.

The effects of RA or BMS009 were studied for the following genes expressed at the gastrula stage as well as for two *Hox* genes expressed very shortly thereafter (the normal expression of all of these genes was already known from previous studies, as listed in Supplementary Table S1): *FoxD* (AF512537), *FoxQ2* (AY163864), *HNF3-2* (Y09236), *HNF3-1* (X96519), *Pax3/7* (AF165886), *Hex* (EU296398), *Pitx* (AJ438768), *Otx* (AF043740), *Cdx* (AF052465), *EvxA* (AF374191),

Gbx (DQ416766), *Lim1/5* (DQ399521), *Six1/2* (EF195742), *Six3/6* (EF195743), *Six4/5* (EF195741), *Sox1/2/3* (AF271787), *Blimp1* (EU708968), *Neurogenin* (AF271788), *Brachyury* (X91903), *Eya* (EF195740), *Delta* (BW899056), *Notch* (Y12539), *Nodal* (AY083838), *Lefty* (EST clone bfne107n04), *Fgf8/17/18* (FJ266460), *Hedgehog* (Y13858), *Wnt1* (AF061974), *Wnt3* (AF361013), *Wnt4* (AF061973), *Wnt5* (AF361014), *Wnt6* (AF361015), *Wnt7* (AF061975), *Wnt8* (AF190470), *Wnt11* (AF187553), *Dkk1/2/4* (EST clone bfga017h15), *Dkk3* (EST clone bflv049h10), *sFRP2-like* (EST clone bfga018e02), *sFRP3/4* (EST clone bfad036d02), *Hox1* (AB028206), *Hox3* (X68045), *Hox4* (AB028208), *Hox6* (Z35146). The last two genes in this list, although not conspicuously transcribed at the gastrula stage, have expression domains known to be influenced by RA signaling at the early neurula stage (Schubert et al., 2004, 2005, 2006).

For blocking protein synthesis, puromycin (Sigma-Aldrich, Saint Louis, MO) was added to gastrula cultures to a final concentration of 200 μg/ml at 3 h, 4 h and 5 h of development. After 5 min, RA or BMS009 ($1x10^{-6}$M) or DMSO was added and after 1 hr embryos were fixed for in situ hybridization. The effectiveness of the puromycin concentration was verified by its ability to block synthesis of endogenous alkaline phosphatase in the amphioxus gut endoderm (Supplementary Fig. S1) (Holland et al., 1996).

Results

Altered RA signaling affects gene expression domains in amphioxus gastrulae

To identify potential direct targets of RA signaling during the gastrula stage in amphioxus, we first determined the effect of altered RA signaling on 40 genes with limited domains of expression along the A-P axis of the gastrula (Fig. 1; Supplementary Table S1) plus *Hox4* and *Hox6*, although expression of the last two only begins at the very end of (*Hox4*) or shortly after (*Hox6*) gastrulation (Schubert et al., 2004, 2005, 2006). Expression of *Hox2* at the gastrula stage was too weak to allow proper interpretation and was therefore excluded from our analysis. The other sampled genes fall into several categories: Notch, Nodal, Wnt and FGF signaling, forkhead and homeobox genes as well as transcription factors (*Neurogenin*, *Sox1/2/3*). About half have a single expression domain in a single tissue layer (either the outer ectoderm or the inner mesendoderm), and about half have domains in both tissue layers and/or two domains in a given tissue layer. Expression of thirteen of these 42 genes (*Cdx*, *Hedgehog*, *HNF3-1*, *Hox1*, *Hox3*, *Hox4*, *Hox6*, *Nodal*, *Notch*, *Otx*, *Pitx*, *Wnt3* and *Wnt5*) at the neurula and later stages was already known to be affected by RA applied during the gastrula stage, but it was not known if their expression in the gastrula itself was also affected (Holland and Holland, 1996; Schubert et al., 2004, 2005, 2006; Osborne et al., 2009). Therefore, it was not known whether any or all are direct targets of RA signaling. Administration of RA or the RA antagonist BMS009 from the onset of gastrulation did not alter the expression patterns of *Pitx*, *Otx*, *Cdx*, *Notch*, *Nodal*, *Hedgehog* and *Wnt5* during the gastrula stage (Fig. 1I-K,X,Y,B',F'), even though their expression was affected by altered RA signaling in later embryos and larvae (Schubert et al., 2005; 2006; Osborne et al., 2009). *Nodal*, *Hedgehog* and *Wnt5* encode secreted signaling proteins involved in axial patterning in early development, and these results suggest that they are acting in parallel to RA signaling at the gastrula stage, with any crosstalk occurring only later in development. Moreover, none of the other 29 other genes we tested was affected by RA at the gastrula stage, excluding them as potential direct targets (Fig.

1). However, expression of seven of these was affected at the neurula stage, indicating that they are probably indirect targets of RA signaling (data not shown). Thus, only six of the 42 genes expressed in limited patterns in the early amphioxus embryo, namely *HNF3-1*, *Hox1*, *Hox3*, *Hox4*, *Hox6* and *Wnt3* responded to the administration of RA or BMS009 at the start of the gastrula stage (Fig. 1D-F,D'; Fig. 2A-O), and were, therefore, potential early immediate targets of RA signaling in the amphioxus gastrula.

Genes with expression domains responding rapidly to RA signaling at the gastrula stage

In untreated embryos, *HNF3-1* expression is detectable from the mid gastrula stage (Fig. 2A) in two domains, the posterior-dorsal and anterior-ventral mesendoderm. By the end of gastrulation (Fig. 2B), *HNF3-1* is expressed throughout both the dorsal and ventral mesendoderm, except in an anterior gap (between the arrowheads in Fig. 2A). In RA-treated embryos, both domains are expanded at the mid-gastrula stage with the anterior gap between them (between the arrowheads in Fig. 2A) being markedly reduced compared to the controls. At the end of gastrulation, the gap is eliminated as *HNF3-1* becomes expressed throughout the mesendoderm of RA-treated embryos (Fig. 2B). Conversely, BMS009 treatments result in a posterior shift in expression, resulting in an increase in the anterior gap between the two domains at the mid-gastrula stage (Fig. 2A) and a posterior restriction of ventral mesendodermal expression compared to controls at the end of gastrulation (Fig. 2B).

Hox1 expression is first detectable in the early gastrula of controls as a small patch of weak staining in the dorsal blastoporal lip (Fig. 2C). As gastrulation proceeds, expression expands first throughout the blastopore lip and then spreads anteriorly in both the ectoderm and mesendoderm. Treatment with RA upregulates *Hox* expression. The *Hox1* domain in the dorsal blastopore lip is considerably enlarged at the early gastrula, and by the mid-gastrula stage it is expanded anteriorly compared to controls (Fig. 2D). By the late gastrula and early neurula, the *Hox1* domain is expanded almost to the anterior tip of the embryo (Fig. 2E,F). Conversely, in BMS009-treated embryos, *Hox1* expression is undetectable at the early gastrula and is restricted to the dorsal lip of the blastopore at the mid and late gastrula stages and to the tail bud, posterior-most nerve cord and somites at the neurula stage (Fig. 2E,F).

Hox3 turns on later than *Hox1* with a more posterior anterior limit. In controls, *Hox3* is first detectable in the dorsal blastopore lip of the late gastrula (Fig. 2G). RA expands the *Hox3* domain, while BMS009 completely downregulates expression. In the late gastrula treated with RA (Fig. 2H), *Hox3* expression is expanded anteriorly compared to controls, in BMS009-treated embryos expression of *Hox3* is restricted to the dorsal blastopore lip. By the early neurula, *Hox3* is expressed throughout the nerve cord of RA treated embryos, and the mesodermal and endodermal domains are also expanded anteriorly (Fig. 2I). Conversely, BMS009 treatment restricts *Hox3* expression to the posterior third of the embryo.

The effect of altered RA signaling on *Hox4* expression is similar to that on *Hox1* and *Hox3*. In embryos treated with RA, *Hox4* expression is first detectable in the dorsal blastopore lip of the mid-gastrula (Fig. 2J). However in controls, expression is only first apparent in a few dorsal cells at the very end of gastrulation (Fig. 2K). At this stage and at the early neurula stage, the *Hox4* domain is expanded anteriorly in RA-treated embryos, while expression is undetectable in BMS009-treated ones (Fig. 2K,L). In RA-treated embryos, *Hox6* expression is first detectable

at the late gastrula stage (Fig. 2M), but not in controls until the neurula stage (Fig. 2N). At this stage, expression in RA-treated embryos is expanded anteriorly, especially in the nerve cord, compared to controls, while expression is only weakly detectable in BMS009-treated amphioxus embryos.

The only other gene we found to be affected by altered RA signaling at the gastrula is *Wnt3*. However, the effects were minimal. In normal embryos, expression of *Wnt3* is detectable around the blastopore from the early gastrula stage onwards. We saw no effect of RA or BMS009 until the late gastrula stage (Fig. 2O). At this stage in controls, *Wnt3* is expressed posteriorly in the ectoderm and ventral endoderm. RA treatment leads to an anterior expansion of the ventral endodermal expression domain, while treatment with BMS009 results in a posterior restriction. There was no effect on expression dorsally in the embryo.

Direct RA target genes in early amphioxus embryos

To determine which of the genes affected by RA treatments at the gastrula stage are direct targets of RA signaling, and which are indirect, we used puromycin to block protein synthesis before adding RA or BMS009. We did not test *Hox4* and *Hox6*, because they are not normally expressed until the very late gastrula or early neurula, respectively. In the presence of puromycin, of the four genes tested, only *Hox1*, *Hox3* and *HNF3-1* were still affected by altered RA signaling levels (Fig. 3A-E). However, when embryos were treated with puromycin and either RA or BMS009, the expression pattern of *Wnt3* in the ventral mesendoderm was not noticeably different from that in controls at the late gastrula stage (Fig. 3F). This is in contrast to the effects of treating solely with RA or BMS009, which resulted, respectively, in anterior and posterior shifts in expression. Thus, new protein synthesis is not necessary for altered RA signaling to affect expression of, but is required for *Wnt3* expression to be affected. These results indicate that *Hox1*, *Hox3* and *HNF3-1* are direct targets of RA signaling at the gastrula stage, while *Wnt3* is likely an indirect target.

Discussion

RA-targeted Hox *genes compared between amphioxus and vertebrate embryos*

In both amphioxus and vertebrates, A-P patterning is most sensitive to perturbations in RA signaling at the gastrula stage. However, the effects of perturbations have usually been assayed as alterations in gene expression, especially in the central nervous system (CNS) and endoderm, at developmental stages *subsequent to* the gastrula (Holland and Holland, 1996; Balmer and Blomhoff, 2002; Escriva et al., 2002; Schubert et al., 2005, 2006; Duester, 2008; Osborne et al., 2009). To date, very few studies in vertebrates have considered the effects of altered RA signaling on gene expression during the gastrula stage itself. Furthermore, these studies have focused on the effects of altered RA signaling on only a few genes, such as anterior *Hox, Cdx*, *Otx*, *Raldh2*, *Cyp26*, *Chordin*, *Iroquois*, *Meis* or *Sizzled* (Kudoh et al., 2002; Roelen et al., 2002; Li et al., 2008a; Ribes et al., 2009). In the present study, we considered the effects of altered RA signaling on 42 genes and covered both temporal and spatial changes in transcription (Supplementary Table S1). From our data and the relatively incomplete data for vertebrates, we conclude that the anterior *Hox* genes are the major components of the RA-sensitive gene network involved in the initial A-P patterning of amphioxus and vertebrate embryos (Fig. 4). The importance of the *Hox* genes is emphasized in our

study by failure of altered RA signaling to change the expression of several dozen non-*Hox* genes (Supplementary Table S1).

Our puromycin experiments showed that *Hox1* and *Hox3* are direct targets of RA. This result is not surprising in the light of the presence of RAREs in the control regions of *Hox1* and *Hox3* in amphioxus (Manzanares et al., 2000; Wada et al., 2006; Amemiya et al., 2008) and vertebrates (Balmer and Blomhoff, 2002; Mainguy et al., 2003; Oosterveen et al. 2003; Glover et al., 2006; Wada et al., 2006; Su and Gudas, 2008). The RAREs near amphioxus *Hox1* and *Hox3* are functional and, when linked to a reporter construct, can direct expression in chicken and mice in a RA-dependent manner (Manzanares et al., 2000; Wada et al., 2006). Moreover, the region of the amphioxus Hox cluster comprising *Hox1* through *Hox6* contains several potential RAREs (Supplementary Fig. S2) (Amemiya et al., 2008). Combined with the effects of RA or RA antagonist treatment on initial *Hox4* and *Hox6* expression, this suggests that, like *Hox1* and *Hox3*, *Hox4* and *Hox6* might also be directly regulated by RA. Likewise, in vertebrates, several *Hox* genes (i.e. *Hoxa1*, *Hoxb1*, *Hoxa3*, *Hoxa4*, *Hoxb4, Hoxd4, Hoxb5* and *Hoxb8*) are directly regulated by RA signaling (Balmer and Blomhoff, 2002; Mainguy et al., 2003; Oosterveen et al., 2003; Glover et al., 2006; Wada et al., 2006; Su and Gudas, 2008). Since both RA signaling and *Hox* genes were probably present in the last common ancestor of bilaterians (Deschamps, 2007; Campo-Paysaa et al., 2008; De Robertis, 2008), it will be interesting to assess, when in evolution the direct regulation of *Hox* genes by RA signaling first appeared.

Genes other than Hox *that may be involved in RA signaling during A-P patterning*

The present study included nine non-*Hox* genes (i.e. *Cdx*, *Hedgehog*, *HNF3-1*, *Nodal*, *Notch*, *Otx*, *Pitx*, *Wnt3* and *Wnt5*) known to have their expression altered in the amphioxus neurula and later stages by perturbations of RA signaling at the gastrula stage (Schubert et al., 2005, 2006; Osborne et al., 2009). However, we found that expression of only two of these genes (*HNF3-1* and *Wnt3*) was influenced by RA signaling at the gastrula stage (Fig. 4).

Our puromycin experiments indicate that, although amphioxus *Wnt3* is probably involved in early A-P patterning of the embryo, it is an indirect, not a direct, target of RA signaling. During gastrulation, the *Wnt3* domain overlaps that of *Hox1*, with *Hox1* being expressed earlier and more broadly than *Wnt3*. Similar patterns have previously been reported in the posterior foregut endoderm of amphioxus neurulae (Schubert et al., 2005) suggesting that *Wnt3* might act downstream of *Hox* genes, such as *Hox1* (Fig. 4). The expression patterns of seven other amphioxus *Wnt* genes (*Wnt1*, *Wnt4*, *Wnt5*, *Wnt6*, *Wnt7*, *Wnt8* and *Wnt11*), which signal through both the canonical and the Wnt/Ca^{++} pathways (Holland et al., 2000; Schubert et al., 2000a, 2000b, 2000c, 2001), as well as those of several Wnt antagonists (*Dkk1/2/4*, *Dkk3*, *sFRP2-like*, *sFRP3/4*) were unaffected by changing RA levels. In contrast, in vertebrates RA is known to affect *Wnt* gene expression and to suppress Wnt signaling possibly by competition of RAR and TCF for binding to b-catenin (Easwaran et al., 1999; Shum et al., 1999; Balmer and Blomhoff 2002; Halilagic et al., 2007; Li et al. 2008a). Taken together, our results support the hypothesis that the complex interactions during early development between RA and *Wnt* genes may be vertebrate innovations (Onai et al., 2009).

Our puromycin experiments showed that *HNF3-1*, unlike *Wnt3*, is a likely direct target of RA signaling—a conclusion that is supported by the presence of three putative RAREs in the amphioxus *HNF3* locus (Supplementary Fig. S2). Amphioxus

has a second *HNF3* gene due to an independent duplication, *HNF3-2* (Shimeld, 1997; Wang et al., 2007), but its expression is not affected by RA signaling at the gastrula stage (data not shown). As in amphioxus, one of the *HNF3* genes in vertebrates (*HNF3a*) is a direct target of RA signaling, at least in cell culture assays (Jacob et al., 1999).

Chordate *HNF3* genes are expressed in the endoderm during gastrulation and are crucial for proper expression of some genes in the anterior visceral endoderm in the mouse and for anterior endoderm development in tunicates (Olsen and Jeffrey, 1997; Kimura-Yoshida et al., 2007). Furthermore, regulation of *Otx* and of several Wnt antagonists (e.g. *Dkk1* and *Cerl*) by *HNF3* in both vertebrates and tunicates suggests a conserved mechanism for *HNF3*-dependent specification of anterior identity (Lamy et al., 2006; Kimura-Yoshida et al., 2007). While it is not known whether *HNF3* genes regulate *Otx* expression in amphioxus, their domains overlap in the endoderm. In amphioxus, the effect of RA on *HNF3-1* expression is limited to the anterior mesendoderm in the gastrula and to the pharyngeal endoderm in the neurula and larval stages—domains where *Hox* genes are not expressed (Schubert et al., 2005). This suggests that RA regulation of *HNF3* in the amphioxus gastrula may be important in establishing the anterior endoderm, a feature that could be common to all chordates.

RA regulation of Hox *genes in chordates: implications for evolution and mechanisms of development*

Recent phylogenetic studies have placed amphioxus as the most basal chordate clade, with tunicates as sister group to the vertebrates (Phillipe et al., 2005; Delsuc et al., 2008; Putnam et al., 2008). This arrangement makes amphioxus a favorable model for deciphering the evolutionary history of chordate development. The utility of amphioxus is also enhanced both by its morphological and genomic simplicity (Koop and Holland, 2008), with the unduplicated genome containing a single cluster of 15 *Hox* genes (Holland et al., 2008). Whether this *Hox15* gene arose by lineage-specific duplication or represents an ancestral chordate condition is unknown. RA signaling influences the spatial expression domains of these *Hox* genes as soon as their transcription has been initiated. In amphioxus, RA treatment results in *Hox* gene expression becoming detectable at earlier developmental stages, while maintaining temporal collinearity. Whether this is due to premature initiation of *Hox* gene expression or to upregulation of *Hox* gene expression is unclear. By the larval stage, expression of anterior *Hox* genes is expanded to the rostral tip in the CNS of RA-treated amphioxus (Schubert et al. 2004, 2006). RA sequentially induces *Hox* gene expression in human embryonal carcinoma cells (Simeone et al., 1990) and precociously initiates expression of *Hoxb1* and *Hoxb2* in mice and of *Hoxd4a* in zebrafish, while in *Raldh2* mutant mice, expression of *Hox1b* is both reduced and delayed (Niederreither et al., 1999; Roelen et al., 2002; Maves and Kimmel, 2005). Although prematurely activated *Hoxb1* expression initially followed the normal spatiotemporal pattern, a *Hoxb1* transgene did not activate prematurely in response to RA suggesting the existence of global, Hox cluster-specific regulatory mechanisms controlling the initiation of *Hox* transcription (Roelen et al., 2002).

Hox genes are activated in a temporally collinear manner in amphioxus (Wada et al., 1999; Schubert et al., 2006). Expression of *Hox1*, *Hox3*, *Hox4* and *Hox6* is first detectable in the dorsal blastopore lip, most prominently in presumptive neuroectoderm. Their domains subsequently spread to the mesendoderm and expand anteriorly. In vertebrates, *Hox* gene expression is initiated in the mesodermal

component of the blastopore margin (or equivalent) (Wacker et al., 2004) and then progresses anteriorly along the A-P axis. The temporal sequence of *Hox* initiation thus translates into a spatial sequence of *Hox* expression along the A-P body axis (Forlani et al., 2003; Wacker et al., 2004; Deschamps and van Nes, 2005; Jansen et al., 2007). Subsequent modification of *Hox* expression in mesoderm and neuroectoderm leads to the establishment of the definitive *Hox* codes (Forlani et al., 2003; Deschamps and van Nes, 2005). At least in the mouse, the initial temporal activation in the primitive streak is under the control of two antagonistic global regulators at either end of the *Hox* cluster, while the second phase is under the control of regulatory cues located within the *Hox* cluster (Tschopp et al., 2009).

Taken together, these results suggest that in amphioxus, as in vertebrates, temporal collinearity of *Hox* genes is initiated under the control of a global regulator, possibly sensitive to RA, in the dorsal blastopore margin, from which expression subsequently spreads anteriorly. Spatial collinearity seems to result from a temporally coordinated activation of *Hox* genes in a "*Hox* induction field" (Deschamps et al., 1999) with final spatial expression of *Hox* genes depending on regulatory cues located within the *Hox* cluster. These local cues include RAREs and hence direct regulation by RA signaling.

In sum, our data suggest that, while RA may play a role in the initiation of *Hox* expression in amphioxus, it is probably not required for the establishment and maintenance of *temporal* collinearity. In contrast, RA signaling is clearly important in controlling *spatial* collinear *Hox* expression in amphioxus after initiation of expression. It thus seems very likely that in the chordate ancestor, a "*Hox* induction field" was established in the posterior ectoderm at the onset of gastrulation, with RA signaling regulating *Hox* gene expression to establish positional identity along the developing A-P axis. RA regulation of *Hox* initiation in this ancestral chordate might have been important for the tight coordination of temporal and spatial collinearity in A-P patterning of new body segments as they are sequentially added during the posterior growth of the larva.

Acknowledgements

The authors would like to thank J. M. Lawrence and S. Bell at the University of South Florida in Tampa, USA, for providing laboratory space during the amphioxus spawning season. We are indebted to A. O'Rourke for assistance in larval culturing and puromycin treatments. This research was supported by funds from ANR (ANR-07-BLAN-0038 and ANR-09-BLAN-0262-02) and CNRS (to M.S.), by CRESCENDO (a European Union Integrated Project of FP6) and MENRT (to V.L. and M.S.), by March of Dimes I-FY05-108 (to L.Z.H.), and by NSF IOB 0416292 (to N.D.H. and L.Z.H.).

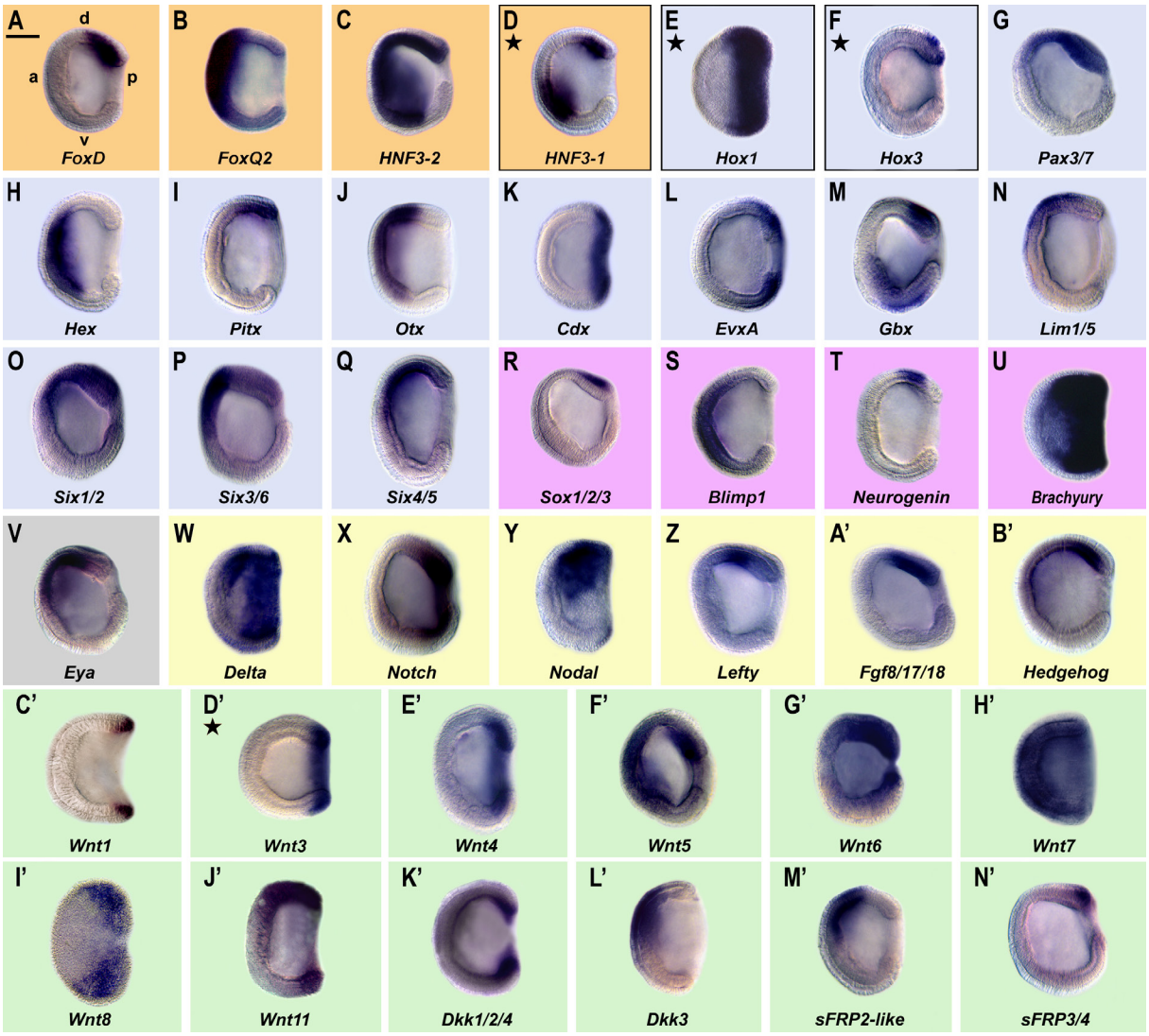


Fig. 1. Genes assayed for regulation by retinoic acid (RA) in amphioxus gastrulae. Side views of 40 whole mounts of embryos of the Florida amphioxus (*Branchiostoma floridae*) at the mid to late gastrula stage. In A, the anterior, posterior, dorsal, and ventral sides of the embryo are indicated, respectively by a, p, d and v. The 50 μm scale line in A is applicable to the entire figure. Each embryo shows the normal expression pattern of a developmental gene. These genes were selected, because they are transcribed in spatially discrete patterns at the gastrula stage. Gene families are indicated by the background color as follows: Forkhead genes (orange); Homeobox genes (blue); Sox domain, zinc finger, bHLH and T-box transcription factor genes (purple); Notch/Delta, Tgfβ, Fgf and Hedgehog signaling genes (yellow); Eya, a protein tyrosine phosphatase (grey); Wnt signaling and Wnt modulator genes (green). Expression of four genes (indicated by stars) are affected in gastrulae when exogenous RA or the RA antagonist BMS009 is administered from very early gastrulation (D-F and D'). Only three of them (boxed) are direct targets of RA signaling (D-F).

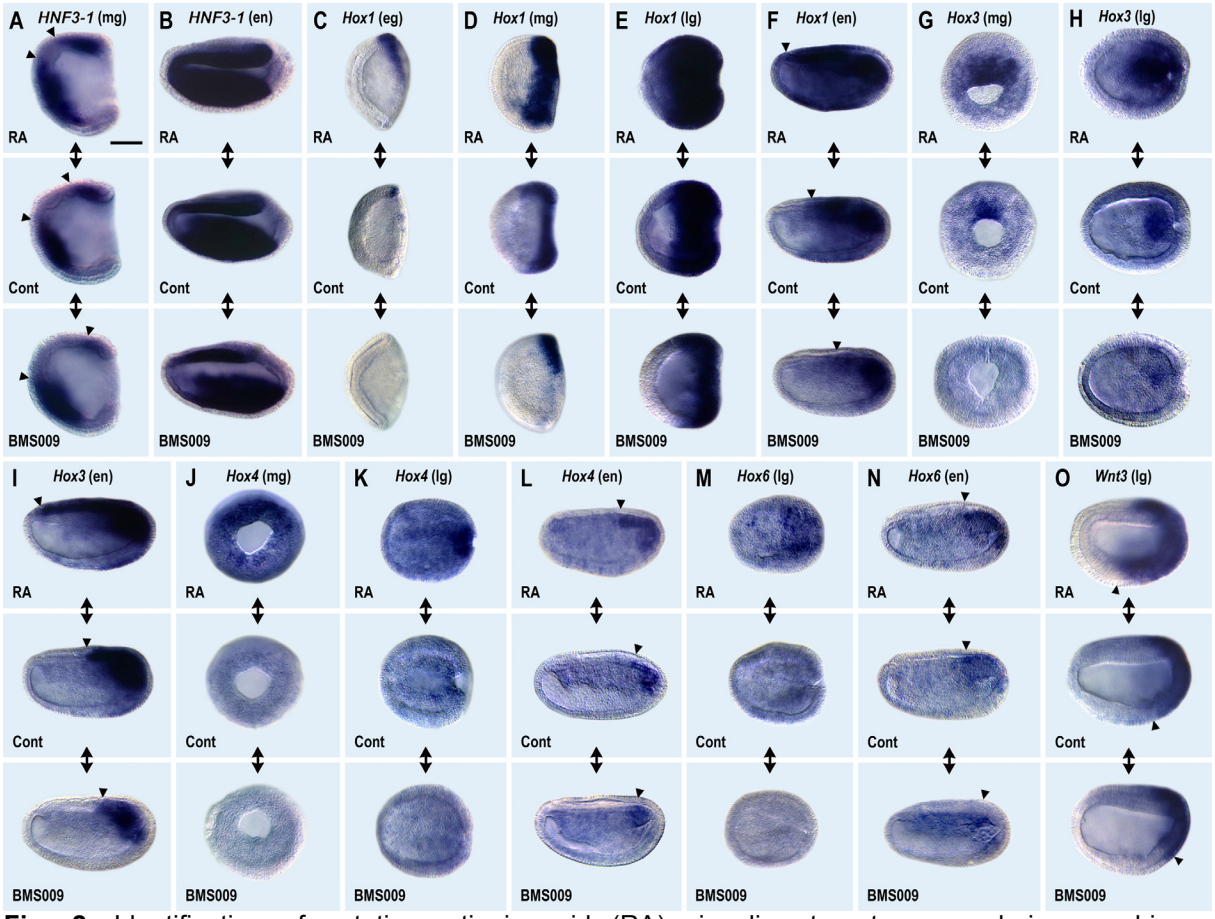


Fig. 2. Identification of putative retinoic acid (RA) signaling target genes during amphioxus gastrulation. Whole mounts of amphioxus embryos in side view (A-F, I, L, N and O) with anterior to the left and dorsal up, in dorsal view (H, K and M) with anterior to the left and in blastopore view (G and J) with dorsal up. The 50 μm scale line in A is applicable to the entire figure. The figure includes four (*HNF3-1*, *Hox1*, *Hox3* and *Wnt3*) genes, whose expression domains during the gastrula stage are altered by exposure to RA or to the RA antagonist BMS009. In addition the effects of RA and BMS009 treatments on *Hox4* and *Hox6*, expressed at the very end of (*Hox4*) or just after (*Hox6*) gastrulation, show that all *Hox* genes tested are affected by altered RA signaling from the onset of expression. The arrowheads indicate the anterior limits of gene expression in mesendoderm (A and O) or nerve cord (F, I, L and N). For each gene, the vertical double-headed arrows facilitate aligning the control and experimental embryos. Abbreviations: (eg) = early gastrula; (mg) = mid gastrula; (lg) = late gastrula; (en) = early neurula; Cont = control; RA = retinoic acid; BMS009 = RA antagonist.

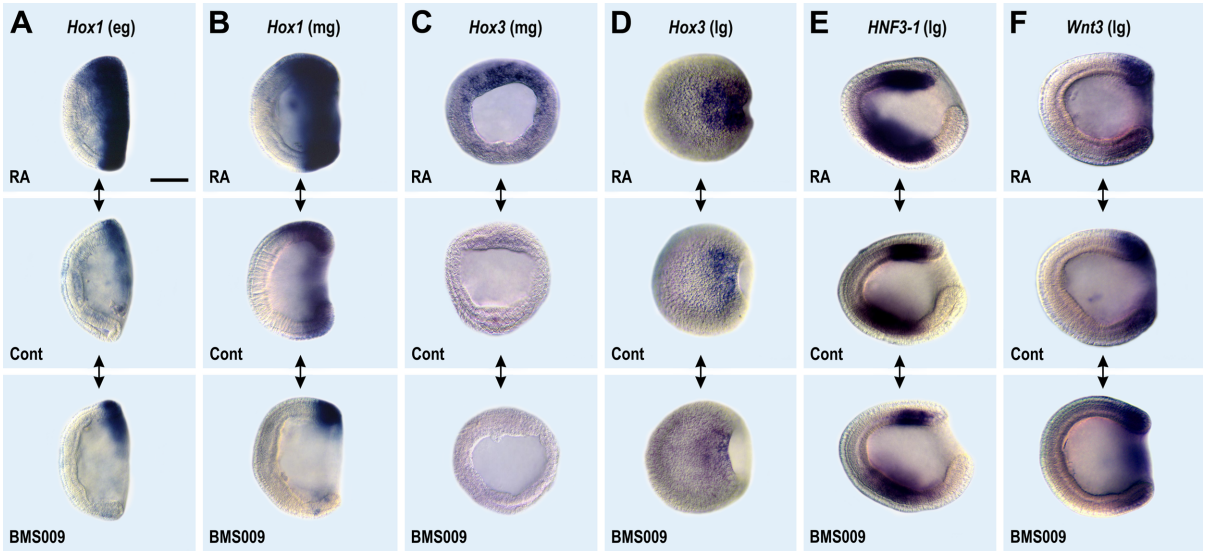


Fig. 3. Identification of direct retinoic acid (RA) signaling target genes in the amphioxus gastrula using puromycin assays. To block translation, amphioxus embryos were treated with puromycin from the start of gastrulation—prior to exposure with RA or RA antagonist (BMS009). For side views (A, B and E, F), anterior is to the left and dorsal is up, for blastopore views (C), dorsal is up and for dorsal views (D), anterior is to the left. This figure shows the expression of *Hox1* (A, B), *Hox3* (C, D), *HNF3-1* (E) and *Wnt3* (F) in controls as well as in embryos treated with either RA or BMS009 after exposure to puromycin. *Hox4* and *Hox6* were not appropriate for this experiment, because their expression begins at very end of (*Hox4*) or just after (*Hox6*) the gastrula stage. Puromycin does not block the effects of RA or BMS009 on expression of *Hox1*, *Hox3* and *HNF3-1*. In contrast, puromycin blocks the effects RA or BMS009 on *Wnt3* expression. For each gene, the vertical double-headed arrows facilitate aligning the control and experimental embryos. The 50 μm scale line in A is applicable to the entire figure. Abbreviations: (eg) = early gastrula; (mg) = mid gastrula; (lg) = late gastrula; Cont = control; RA = retinoic acid; BMS009 = RA antagonist.

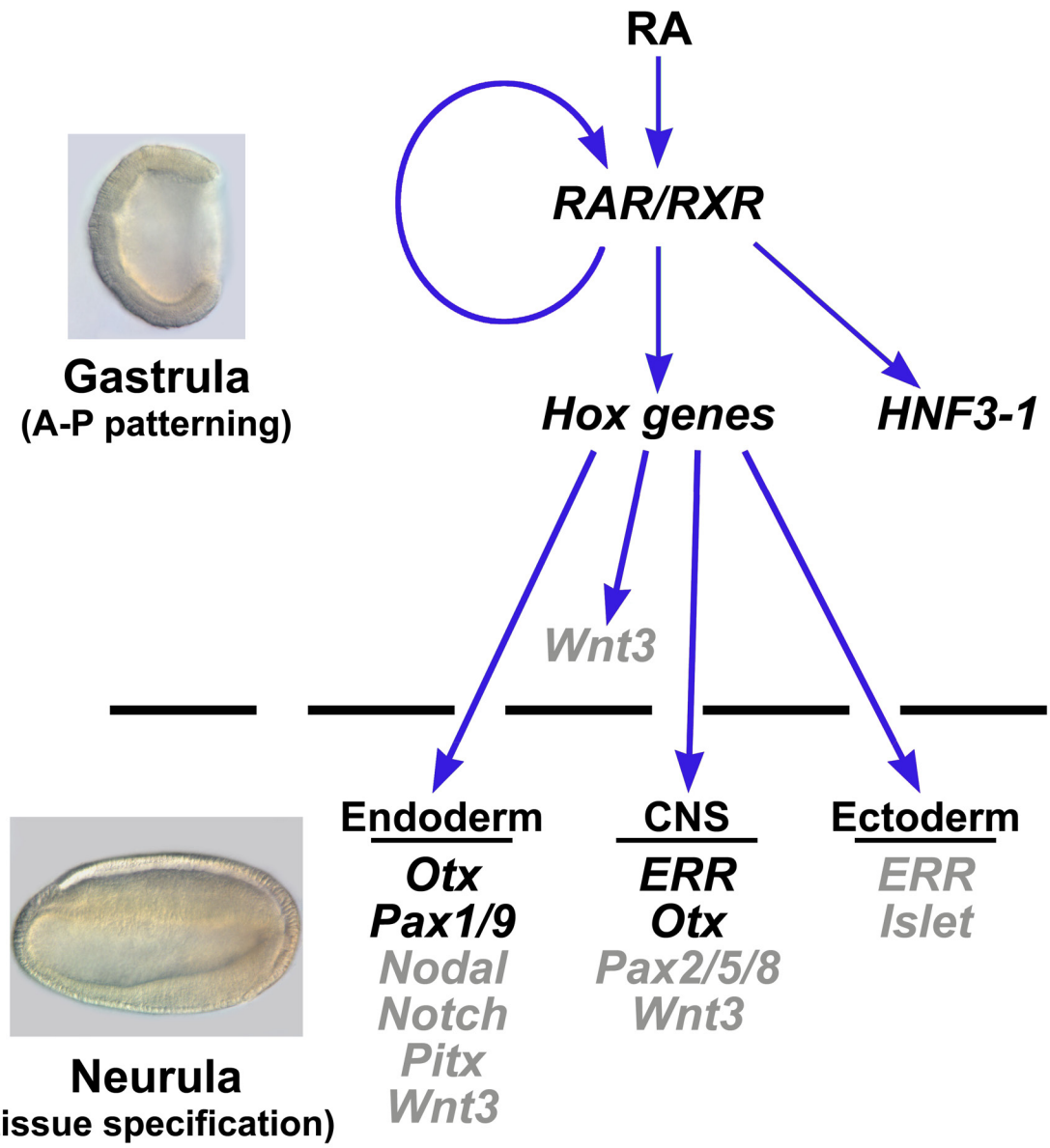


Fig. 4. The retinoic acid (RA) signaling network in the developing amphioxus embryo. Experimentally confirmed regulatory interactions are indicated in black and putative ones in grey. RA binds to RAR, whose expression is controlled by an autoregulatory loop (data not shown). RA patterns the anterior-posterior (A-P) axis during gastrulation by directly regulating expression of *Hox* genes and of *HNF3-1*. These genes are in turn involved in conferring positional identity during the neurula stage with *Hox* genes regulating genes, like *ERR*, *Otx* and *Pax2/5/8* in the central nervous system (CNS) or *ERR* and *Islet* in the general ectoderm (Schubert et al., 2004, 2006). In the endoderm, *Hox1* is involved in establishing the posterior limit of the pharynx by downregulating expression of pharyngeal genes, such as *Otx*, *Pax1/9*, *Nodal*, *Notch* and *Pitx*, in posterior foregut endoderm (Schubert et al., 2005). *Wnt3*, an indirect RA signaling target, might act downstream of *Hox* genes, such as *Hox1*, both in the gastrula and during later development in the endoderm and the CNS (Schubert et al., 2005, 2006).

References

Amemiya, C.T., Prohaska, S.J., Hill-Force, A., Cook, A., Wasserscheid, J., Ferrier, D.E.K., Pascual-Anaya, J., Garcia-Fernàndez, J., Dewar, K., Stadler, P. F., 2008. The amphioxus *Hox* cluster: characterization, comparative genomics, and evolution. J. Exp. Zool. 310B, 465-477.

Balmer, J.E., Blomhoff, R., 2002. Gene expression regulation by retinoic acid. J. Lipid Res. 43, 1773-1808.

Blomhoff, R., Blomhoff, H.K., 2006. Overview of retinoid metabolism and function. J. Neurobiol. 66, 606-630.

Campo-Paysaa, F., Marlétaz, F., Laudet, V., Schubert, M., 2008. Retinoic acid signaling in development: tissue-specific functions and evolutionary origins. Genesis 46**,** 640-656.

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

Casci, T., 2008. Retinoic acid passes the morphogen test. Nat. Rev. Genet. 9, 7.

Dann, C.T., Alvarado, A.L., Malyneux, L.A., Denard, B.S., Garbers, D.L., Porteur, M.H., 2008. Spermatogonial stem cell self-renewal requires OCT4, a factor downregulated during retinoic acid-induced differentiation. Stem Cells 26, 2928-2937.

Delarue, M., Johnson, K.E., Boucaut, J.C., 1998. Anteroposterior segregation of superficial and deep cells during gastrulation in *Pleurodeles waltl* and *Rana pipiens* embryos. J. Exp. Zool. 276A, 345-360.

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

De Robertis, E. M., 2008. Evo-Devo: variations on ancestral themes. Cell 132, 185-195.

Deschamps, J., van den Akker, E., Forlani, S., De Graaff, W., Oosterveen, T., Roelen, B., Roelfsema, J., 1999. Initiation, establishment and maintenance of *Hox* gene expression patterns in the mouse. Int. J. Dev. Biol. 43, 635-650.

Deschamps, J., van Nes, J., 2005. Developmental regulation of the Hox genes during axial morphogenesis in the mouse. Development 132, 2931-2942.

Deschamps, J., 2007. Ancestral and recently recruited global control of the *Hox* genes in development. Curr. Opin. Genet. Dev. 17, 422-427.

Duester, G., 2008. Retinoic acid synthesis and signaling during early organogenesis. Cell 134, 921-931.

Durston, A.J., Timmermans, J.P.M., Hage, W.J., Hendriks, H.F.J., de Vries, N.J., Heideveld, M., Nieuwkoop, P.D., 1989. Retinoic acid causes an anteroposterior transformation in the developing central nervous system. Nature 340**,** 141-144.

Easwaran, B., Pishvaian, M., Salimuddin, Byers, S., 1999. Cross-regulation of b-catenin-LEF/TCF and retinoid signaling pathways. Curr. Biol. 9, 1415-1418.

Escriva, H., Holland, N.D., Gronemeyer, H., Laudet, V., 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, 2905-2916.

Forlani, S., Lawson, K.A., Deschamps, J., 2003. Acquisition of Hox codes during gastrulation and axial elongation in the mouse embryo. Development 130, 3807-3819.

Fujiwara, S., 2005. Promoter activity of the retinoic acid receptor gene in the *Ciona intestinalis* embryo. Dev. Dyn. 232, 1124-1130.

Glover, J.C., Renaud, J.S., Rijli, F.M., 2006. Retinoic acid and hindbrain patterning. J.Neurobiol. 66, 705-725.

Grapin-Botton, A., 2005. Antero-posterior patterning of the vertebrate digestive tract: 40 years after Nicole Le Douarin's PhD thesis. Int. J. Dev. Biol. 49, 335-347.

Halilagic, A., Ribes, V., Ghyselinck, N.B., Zile, M.H., Dollé, P., Studer, M., 2007. Retinoids control anterior and dorsal properties in the developing forebrain. Dev. Biol. 303**,** 362-375.

Heyman, R.A., Mangelsdorf, D.J., Dyck, J.A., Stein, R.B., Eichele, G., Evans, R.M., Thaller, C., 1992. 9-*cis* retinoic acid is a high affinity ligand for the retinoid X receptor. Cell 68, 397-406.

Holland, L.Z., Holland, N.D. 1996. Expression of *AmphiHox-1* and *AmphiPax-1* in amphioxus embryos treated with retinoic acid: insights into evolution and patterning of the chordate nerve cord and pharynx. Development 122, 1829-1838.

Holland, L.Z., Holland, P.W.H., Holland, N.D. 1996. Revealing homologies between body parts of distantly related animals by in situ hybridization to developmental genes: amphioxus versus vertebrates. In: Palumbi, S. and Ferraris, J.D. (Eds), Molecular Approaches to Zoology and Evolution. John Wiley, New York, pp. 267-282; 473-483.

Holland, L.Z., Holland, N.D., Schubert, M., 2000. Developmental expression of *AmphiWnt1*, an amphioxus gene in the *Wnt1/wingless* subfamily. Dev. Genes Evol. 210, 522-524.

Holland, L.Z., Yu, J.K., 2004. Cephalochordate (amphioxus) embryos: procurement, culture, basic methods. Meth. Cell Biol. 74, 195-215.

Holland, L.Z., Holland, N.D., 2007. A revised fate map for amphioxus and the evolution of axial patterning in chordates. Int. Comp. Biol. 47, 360-372.

Holland, L.Z., Albalat, R., Azumi, K., Benito-Gutiérrez, E., Blow, M.J., Bronner-Fraser, M., Brunet, F., Butts, T., Candiani, S., Dishaw, L.J., Ferrier, D.E.K., Garcia-Fernàndez, J., Gibson-Brown, J.J., Gissi, C., Godzik, A., Hallböök, F., Hirose, D., Hosomichi, K., Ikuta, T., Inoko, H., Kasahara, M., Kasamatsu, J., Kawashima, T., Kimura, A., Kobayashi, M., Kozmik, Z., Kubokawa, K., Laudet, V., Litman, G.W., McHardy, A.C., Meulemans, D., Nonaka, M., Olinski, R.P., Pancer, Z., Pennacchio, L.A., Pestarino, M., Rast, J.P., Rigoutsos, I., Robinson-Rechavi, M., Roch, G., Saiga, H., Sasakura, Y., Satake, M., Satou, Y., Schubert, M., Sherwood, N., Shiina, T., Takatori, N., Tello, J., Vopalensky, P., Wada, S., Xu, A., Ye, Y., Yoshida, K., Yoshizaki, F., Yu, J.K., Zhang, Q., Zmasek, C.M., de Jong, P.J., Osoegawa, K., Putnam, N.H., Rokhsar, D.S., Satoh, N., Holland, P.W.H., 2008. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. Genome Res. 18, 1100-1111.

Holland, N.D., Holland, L.Z. Embryos and larvae of invertebrate deuterostomes. In: Stern, C.S. and Holland, P.W.H. (Eds.), Essential Developmental Biology, a Practical Approach. IRL Press, Oxford, 1993, pp. 21-32.

Imai, K.S., Stolfi, A., Levine, M., Satou, Y., 2009. Gene regulatory networks underlying the compartmentalization of the *Ciona* central nervous system. Development 136, 285-293.

Ishibashi, T., Nakazawa, M., Ono, H., Satoh, N., Gojobori, T., Fujiwara, S., 2003. Microarray analysis of embryonic retinoic acid target genes in the ascidian *Ciona intestinalis*. Dev. Growth Different. 45, 249-259.

Jacob, A., Budhiraja, S., Reichel, R.R., 1999. The HNF-3α transcription factor is a primary target for retinoic acid action. Exp. Cell Res. 250, 1-9.
Jansen, H.J., Wacker, S.A., Bardine, N., Durston, A.J., 2007. The role of the Spemann organizer in anterior-posterior patterning of the trunk. Mech. Dev. 124, 668-681.
Kimura, W., Yasugi, S., Stern, C.D., Fukuda, K., 2006. Fate and plasticity of the endoderm in the early chick embryo. Dev. Biol. 289, 283-295.
Kimura-Yoshida, C., Tian, E., Nakano, H., Amazaki, S., Shimokawa, K., Rossant, J., Aizawa, S., Matsuo, I., 2007. Crucial roles of Foxa2 in mouse anterior-posterior axis polarization via regulation of anterior visceral endoderm-specific genes. Proc. Natl. Acad. Sci. U.S.A. 104, 5915-5924.
Koop, D., Holland, L.Z., 2008. The basal chordate amphioxus as a simple model for elucidating developmental mechanisms in vertebrates. Birth Defects Res. C Embryo Today 84, 175-187.
Kudoh, T., Wilson, S.W., Dawid, I.B., 2002. Distinct roles for Fgf, Wnt and retinoic acid in posteriorizing the neural ectoderm. Development 129, 4335-4346.
Lamy, C., Rothbächer, U., Caillol, D., Lemaire, P., 2006. *Ci-FoxA-a* is the earliest zygotic determinant of the ascidian anterior ectoderm and directly activates *Ci-sFRP1/5*. Development 133, 2835-2844.
Li, S., Lou, X., Wang, J., Liu, B., Ma, L., Su, Z., Ding, X., 2008a. Retinoid signaling can repress blastula Wnt signaling and impair dorsal development in *Xenopus* embryo. Differentiation 76, 897-907.
Li, Y., Rankin, S.A., Sinner, D., Kenny, A.P., Krieg, P.A., Zorn, A.M., 2008b. Sfrp5 coordinates foregut specification and morphogenesis by antagonizing both canonical and noncanonical Wnt11 signaling. Genes Dev. 22, 3050-3063.
Mainguy, G., Reiden, P., Berezikov, E., Woltering, J.M., Plasterk, R.H.A., Durston, A.J., 2003. A position-dependent organisation of retinoid response elements is conserved in the vertebrate *Hox* clusters. Trends Genet. 19, 476-479.
Manzanares, M., Wada, H., Hasaki, N., Trainor, P.A., Krumlauf, R., Holland, P.W.H., 2000. Conservation and elaboration of Hox gene regulation during evolution of the vertebrate head. Nature 408, 854-857.
Marlétaz, F., Holland, L.Z., Laudet, V., Schubert, M., 2006. Retinoic acid signaling and the evolution of chordates. Int. J. Biol. Sci. 2, 38-47.
Maves, L., Kimmel, C.B., 2005. Dynamic and sequential patterning of the zebrafish posterior hindbrain by retinoic acid. Dev. Biol. 285, 593-605.
McCaffery, P.J., Adams, J., Maden, M., Rosa-Molinar, E., 2003. Too much of a good thing: retinoic acid as an endogenous regulator of neural differentiation and exogenous teratogen. Eur. J. Neurosci. 18, 457-472.
Mongan, N.P., Gudas, L.J., 2007. Diverse actions of retinoid receptors in cancer prevention and treatment. Differentiation 75, 853-870.
Nagatomo, K., Fujiwara, S., 2003. Expression of *Raldh2*, *Cyp26* and *Hox-1* in normal and retinoic acid-treated *Ciona intestinalis* embryos. Gene Exp. Patt. 3, 273-277.
Niederreither, K., Subbarayan, V., Dollé, P., Chambon, P., 1999. Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. Nature Genet. 21, 444-448.
Niederreither, K., Dollé, P., 2008. Retinoic acid in development: towards an integrated view. Nat. Rev. Genet. 9, 541-553.

Olsen, C.L., Jeffery, W.R., 1997. A forkhead gene related to HNF-3β is required for gastrulation and axis formation in the ascidian embryo. Development 124, 3609-3619.

Onai, T., Lin, H.C., Schubert, M., Koop, D., Osborne, P.W., Alvarez, S., Alvarez, R., Holland, N.D., 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.

Oosterveen, T., Niederreither, K., Dollé, P., Chambon, P., Meijlink, F., Deschamps, J., 2003. Retinoids regulate the anterior expression boundaries of 5' *Hoxb* genes in posterior hindbrain. EMBO J. 22, 262-269.

Osborne, P.W., Benoit, G., Laudet, V., Schubert, M., Ferrier, D.E.K., 2009. Differential regulation of ParaHox genes by retinoic acid in the invertebrate chordate amphioxus (*Branchiostoma floridae*). Dev. Biol. 327, 252-262.

Philippe, H., Lartilliot, N., Brinkman, H., 2005. Multigene analyses of bilaterian animals chorroborate the monophyly of Ecdysozoa, Lophotrochozoa and Protostomia. Mol. Biol. Evol. 22, 1246-1253.

Putnam, N.H., Butts, T., Ferrier, D.E.K., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.K., Benito-Gutiérrez, E.L., Dubchak, I., Garcia-Fernàndez, J., Gibson-Brown, J.J., Grigoriev, I.V., Horton, A.C., de Jong, P.J., Jurka, J., Kapitonov, V.V., Kohara, Y., Kuroki, Y., Lindquist, E., Lucas, S., Osoegawa, K., Pennacchio, L.A., Salamov, A.A., Satou, Y., Sauka-Spengler, T., Schmutz, J., Shin-I, T., Toyoda, A., Bronner-Fraser, M., Fujiyama, A., Holland, L.Z., Holland, P.W.H., Satoh, N., Rokhsar, D.S., 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature 453, 1064-1071.

Ribes, V., Le Roux, I., Rhinn, M., Schuhbaur, B., Dollé, P., 2009. Early mouse caudal development relies on crosstalk between retinoic acid, Shh and Fgf signalling pathways. Development 136, 665-676.

Roelen, B.A., De Graaff, W., Forlani, S., Deschamps, J., 2002. *Hox* cluster polarity in early transcriptional availability: a high order regulatory level of clustered *Hox* genes in the mouse. Mech. Dev. 119, 81-90.

Rudert, F., Gronemeyer, H., 1993. Retinoic acid-response elements with a highly repetitive structure isolated by immuno-selection from genomic DNA. J. Steroid Biochem. Mol. Biol. 46, 121-133.

Schoenwolf, G.C., Smith, J.L., 2000. Gastrulation and early mesodermal patterning in vertebrates. Meth. Mol. Biol. 135, 69-109.

Schubert, M., Holland, L.Z., Holland, N.D., 2000a. Characterization of two amphioxus *Wnt* genes (*AmphiWnt4* and *AmphiWnt7b*) with early expression in the developing central nervous system. Dev. Dyn. 217, 205-215.

Schubert, M., Holland, L.Z., Holland, N.D., 2000b. Characterization of an amphioxus *Wnt* gene, *AmphiWnt11*, with possible roles in myogenesis and tail outgrowth. Genesis 27, 1-5.

Schubert, M., Holland, L.Z., Panopoulou, G.D., Lehrach, H, Holland, N.D., 2000c. Characterization of amphioxus *AmphiWnt8*: insights into the evolution of patterning of the embryonic dorsoventral axis. Evol. Dev. 2, 85-92.

Schubert, M., Holland, L.Z., Stokes, M.D., Holland, N.D., 2001. Three amphioxus *Wnt* genes (*AmphiWnt3*, *AmphiWnt5*, and *AmphiWnt6*) associated with the tail bud: the evolution of somitogenesis in chordates. Dev. Biol. 240, 262-273.

Schubert, M., Holland, N.D., Escriva, H., Holland, L.Z., Laudet, V., 2004. Retinoic acid influences anteroposterior positioning of epidermal sensory neurons and

their gene expression in a developing chordate (amphioxus). Proc. Natl. Acad. Sci. U.S.A. 101, 10320-10325.
Schubert, M., Yu, J.K., Holland, N.D., Escriva, H., Laudet, V., Holland, L.Z., 2005. Retinoic acid signaling acts via Hox1 to establish the posterior limit of the pharynx in the chordate amphioxus. Development 132, 61-73.
Schubert, M., Holland, N.D., Laudet, V., Holland, L.Z., 2006. A retinoic acid-*Hox* hierarchy controls both anterior/posterior patterning and neuronal specification in the developing central nervous system of the cephalochordate amphioxus. Dev. Biol. 296, 190-202.
Shimeld, S.M., 1997. Characterisation of amphioxus HNF-3 genes: conserved expression in the notochord and floor plate. Dev. Biol. 183, 74-85.
Shum, A.S.W., Poon, L.L.M., Tang, W.W.T., Koide, T., Chan, B.W.H., Leung, Y.C.G., Shiroishi, T., Copp, A.J., 1999. Retinoic acid induces down-regulation of Wnt-3a, apoptosis and diversion of tail bud cells to a neural fate in the mouse embryo. Mech. Dev. 84, 17-30.
Simeone, A., Acampora, D., Arcioni, L., Andrews, P.W., Boncinelli, E., Mavilio, F., 1990. Sequential activation of *HOX2* homeobox genes by retinoic acid in human embryonal carcinoma cells. Nature 346, 763-766.
Sive, H.L., Draper, B.W., Harland, R.M., Weintraub, H., 1990. Identification of a retinoid acid-sensitive period during primary axis formation in *Xenopus laevis*. Genes Dev. 4, 932-942.
Stafford, D., White, R.J., Kinkel, M.D., Linville, A., Schilling, T.F., Prince, V.E., 2006. Retinoids signal directly to zebrafish endoderm to specify *insulin*-expressing β-cells. Development 133, 949-956.
Su, D., Gudas, L. J., 2008. Gene expression profiling elucidates a specific role for RARγ in the retinoic acid-induced differentiation of F9 teratocarcinoma stem cells. Biochem. Pharmacol. 75, 1129-1160.
Tschopp, P., Tarchini, B., Spitz, F., Zakany, J., Duboule, D., 2009. Uncoupling time and space in the collinear regulation of *Hox* genes. PLoS Genet. 5, e1000398.
Wacker, S.A., Jansen, H.J., McNulty, C.L., Houtzager, E., Durston, A.J., 2004. Timed interactions between the *Hox* expressing non-organiser mesoderm and the Spemann organiser generate positional information during vertebrate gastrulation. Dev. Biol. 268, 207-219.
Wada, H., Garcia-Fernàndez, J., Holland, P.W.H., 1999. Colinear and segmental expression of amphioxus Hox genes. Dev. Biol. 213, 131-141.
Wada, H., Escriva, H., Zhang, S., Laudet, V., 2006. Conserved RARE localization in amphioxus Hox clusters and implications for Hox code evolution in the vertebrate neural crest. Dev. Dyn. 235, 1522-1531.
Wang, W., Zhong, J., Su, B., Zhou, Y., Wang, Y.J., 2007. Comparison of *Pax1/9* locus reveals 500-Myr-old syntenic block and evolutionary conserved noncoding regions. Mol. Biol. Evol. 24, 784-791.
White, R.J., Nie, Q., Lander, A.D., Schilling, T.F., 2007. Complex regulation of cyp26a1 creates a robust retinoic acid gradient in the zebrafish embryo. PLoS Biol. 5, e304.
Zhang, S.C., Holland, N.D., Holland, L.Z., 1997. Topographic changes in nascent and early mesoderm in amphioxus embryos studied by DiI labeling and by in situ hybridization for a *Brachyury* gene. Dev. Genes. Evol. 206, 532-535.

Table S1. Genes examined as potential targets of retinoic acid (RA) signaling at the gastrula stage of the Florida amphioxus (*Branchiostoma floridae*).

Gene name	Reference
FoxD	Yu et al., 2002b
FoxQ2	Yu et al., 2003
HNF3-2	Shimeld, 1997
HNF3-1	Shimeld, 1997
Hox1	Wada et al., 1999
Hox3	Wada et al., 1999
Hox4	Wada et al., 1999
Hox6	Cohn, 2002
Pax3/7	Holland et al., 1999
Hex	Yu et al., 2007
Pitx	Boorman and Shimeld, 2002
Otx	Williams and Holland, 1998
Cdx	Brooke et al., 1998
EvxA	Ferrier et al., 2001
Gbx	Castro et al., 2006
Lim1/5	Langeland et al., 2006
Six1/2	Kozmik et al., 2007
Six3/6	Kozmik et al., 2007
Six4/5	Kozmik et al., 2007
Sox1/2/3	Holland et al., 2000b
Blimp1	T. Onai, personal communication
Neurogenin	Holland et al., 2000b
Brachyury	Holland et al., 1995
Eya	Kozmik et al., 2007
Delta	Rasmussen et al., 2007
Notch	Holland et al., 2001
Nodal	Yu et al., 2002a
Lefty	Yu et al., 2007
Fgf8/17/18	Meulemans and Bronner-Fraser, 2007
Hedgehog	Shimeld, 1999
Wnt1	Holland et al., 2000a
Wnt3	Schubert et al., 2001
Wnt4	Schubert et al., 2000a
Wnt5	Schubert et al., 2001
Wnt6	Schubert et al., 2001
Wnt7	Schubert et al., 2000a
Wnt8	Schubert et al., 2000c
Wnt11	Schubert et al., 2000b
Dkk1/2/4	Yu et al., 2007
Dkk3	Yu et al., 2007
sFRP2-like	Yu et al., 2007
sFRP3/4	Yu et al., 2007

Supplementary References

Boorman, C.J., Shimeld, S.M. 2002. Pitx homeobox genes in *Ciona* and amphioxus show left-right asymmetry is a conserved chordate character and define the ascidian adenohypophysis. Evol. Dev. 4, 354-365.

Brooke, N.M., Garcia-Fernàndez, J., Holland, P.W.H., 1998. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. Nature 392, 920-922.

Castro, L.F.C., Rasmussen, S.L.K., Holland, P.W.H., Holland, N.D., Holland, L.Z., 2006. A Gbx homeobox gene in amphioxus: insights into ancestry of the ANTP class and evolution of the midbrain/hindbrain boundary. Dev. Biol. 295, 40-51.

Cohn, M.J., 2002. Lamprey *Hox* genes and the origin of jaws. Nature. 416, 386-387.

Ferrier, D.E.K., Minguillón, C., Cebrián, C., Garcia-Fernàndez, J., 2001. Amphioxus Evx genes: implications for the evolution of the Midbrain-Hindbrain Boundary and the chordate tailbud. Dev. Biol. 237, 270-281.

Holland, L.Z., Schubert, M., Kozmik, Z., Holland, N.D., 1999. *AmphiPax3/7*, an amphioxus paired box gene: insights into chordate myogenesis, neurogenesis, and the possible evolutionary precursor of definitive vertebrate neural crest. Evol. Dev. 1, 153-165.

Holland, L.Z., Holland, N.D., Schubert, M., 2000a. Developmental expression of *AmphiWnt1*, an amphioxus gene in the *Wnt1/wingless* subfamily. Dev. Genes Evol. 210, 522-524.

Holland, L.Z., Schubert, M., Holland, N.D., Neuman, T., 2000b. Evolutionary conservation of the presumptive neural plate markers *AmphiSox1/2/3* and *AmphiNeurogenin* in the invertebrate chordate amphioxus. Dev. Biol. 226, 18-33.

Holland, L.Z., Rached, L.A., Tamme, R., Holland, N.D., Kortschak, D., Inoko, H., Shiina, T., Burgtorf, C., Lardelli, M., 2001. Characterization and developmental expression of the amphioxus homolog of Notch (*AmphiNotch*): evolutionary conservation of multiple expression domains in amphioxus and vertebrates. Dev. Biol. 232, 493-507.

Holland, P.W.H., Koschorz, B., Holland, L.Z., Herrmann, B.G., 1995 Conservation of *Brachyury* (*T*) genes in amphioxus and vertebrates: developmental and evolutionary implications. Development 121, 4283-4291.

Kozmik, Z., Holland, N.D., Kreslova, J., Olivieri, D., Schubert, M., Jonasova, K., Holland, L.Z., Pestarino, M., Benes, V., Candiani, S., 2007. *Pax-Six-Eya-Dach* network during amphioxus development: conservation *in vitro* but context specificity *in vivo*. Dev. Biol. 306, 143-159.

Langeland, J.A., Holland, L.Z., Chastain, R.A., Holland, N.D., 2006. An amphioxus LIM-homeobox gene, *AmphiLim1/5*, expressed early in the invaginating organizer region and later in differentiating cells of the kidney and central nervous system. Int. J. Biol. Sci. 2, 110-116.

Meulemans, D., Bronner-Fraser, M., 2007. Insights from amphioxus into the evolution of vertebrate cartilage. PLoS ONE. 2, e787.

Rasmussen, S.L.K., Holland, L.Z., Schubert, M., Beaster-Jones, L., 2007. Amphioxus *AmphiDelta*: evolution of Delta protein structure, segmentation, and neurogenesis. Genesis 45:113-122.

Schubert, M., Holland, L.Z., Holland, N.D., 2000a. Characterization of two amphioxus *Wnt* genes (*AmphiWnt4* and *AmphiWnt7b*) with early expression in the developing central nervous system. Dev. Dyn. 217, 205-215.

Schubert, M., Holland, L.Z., Holland, N.D., 2000b. Characterization of an amphioxus *Wnt* gene, *AmphiWnt11*, with possible roles in myogenesis and tail outgrowth. Genesis 27, 1-5.

Schubert, M., Holland, L.Z., Panopoulou, G.D., Lehrach, H, Holland, N.D., 2000c. Characterization of amphioxus *AmphiWnt8*: insights into the evolution of patterning of the embryonic dorsoventral axis. Evol. Dev. 2, 85-92.

Schubert, M., Holland, L.Z., Stokes, M.D., Holland, N.D., 2001. Three amphioxus *Wnt* genes (*AmphiWnt3*, *AmphiWnt5*, and *AmphiWnt6*) associated with the tail bud: the evolution of somitogenesis in chordates. Dev. Biol. 240, 262-273.

Shimeld, S.M., 1997. Characterization of amphioxus HNF-3 genes: conserved expression in the notochord and floor plate. Dev. Biol. 183, 74-85.

Shimeld, S.M., 1999. The evolution of the hedgehog gene family in chordates: insights from amphioxus hedgehog. Dev. Genes Evol. 209, 40-47.

Wada, H., Garcia-Fernàndez, J., Holland, P.W.H., 1999. Colinear and segmental expression of amphioxus Hox genes. Dev. Biol. 213, 131-141.

Williams, N.A., Holland, P.W.H., 1998. Molecular evolution of the brain of chordates. Brain Behav. Evol. 52, 177-185.

Yu, J.K., Holland, L.Z., Holland, N.D., 2002a. An amphioxus *nodal* gene (*AmphiNodal*) with early symmetrical expression in the organizer and mesoderm and later asymmetrical expression associated with left-right axis formation. Evol. Dev. 4, 418-425.

Yu, J.K., Holland, N.D., Holland, L.Z., 2002b. An amphioxus winged helix/forkhead gene, *AmphiFoxD*: insights into vertebrate neural crest evolution. Dev. Dyn. 225, 289-297.

Yu, J.K., Holland, N.D., Holland, L.Z., 2003. *AmphiFoxQ2*, a novel winged helix/forkhead gene, exclusively marks the anterior end of the amphioxus embryo. Dev. Genes Evol. 213, 102-105.

Yu, J.K., Satou, Y., Holland, N.D., Shin-I, T., Kohara, Y., Satoh, N., Bronner-Fraser, M., Holland, L.Z., 2007. Axial patterning in cephalochordates and the evolution of the organizer. Nature 445, 613-617.

Fig. S1. Effects of puromycin on alkaline phosphatase activity in amphioxus embryos. Activity of the protein synthesis inhibitor puromycin in amphioxus was assessed by determining endogenous alkaline phosphatase levels. At the neurula stage (16 h), control embryos (A) show endogenous alkaline phosphatase activity in the midgut endoderm (arrow), while puromycin-treated embryos (B) do not. The scale line is 50 μm.

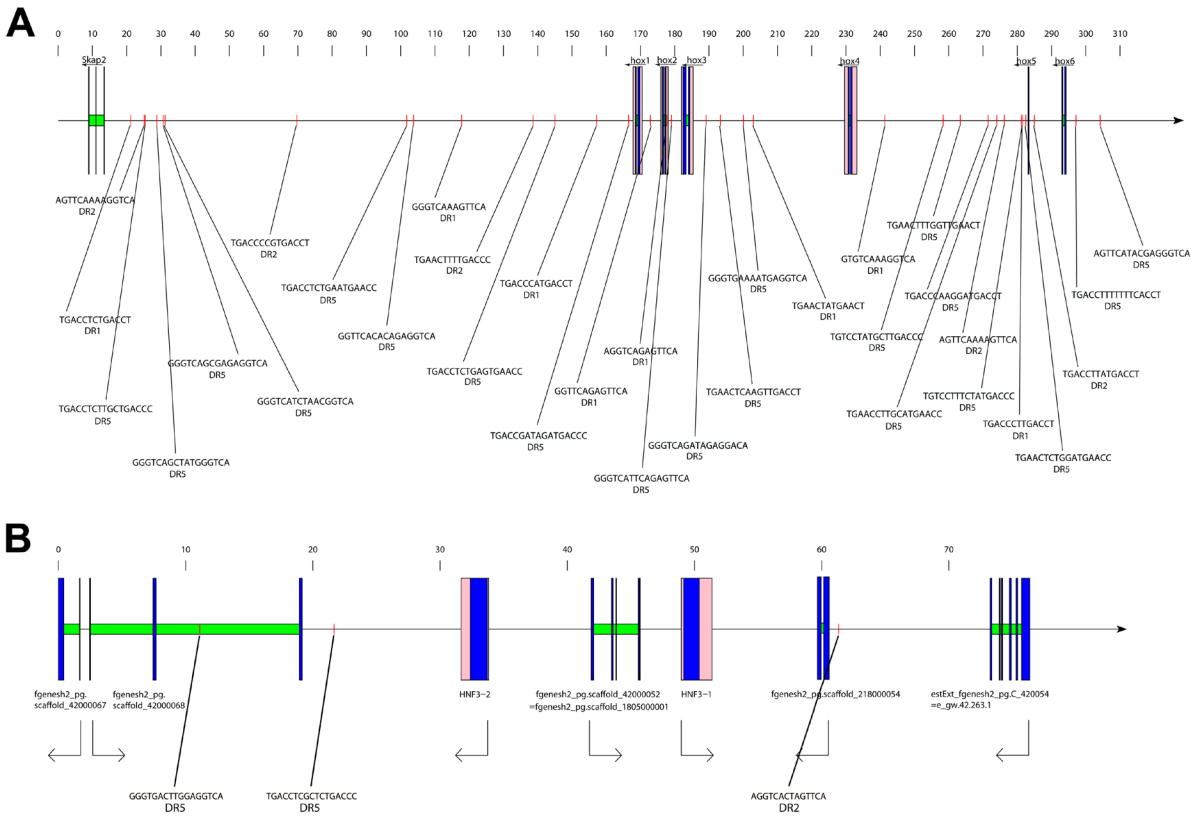


Fig. S2. Distribution of putative retinoic acid response elements (RAREs) in the amphioxus *Hox* (A) and *HNF3* (B) loci. For the *Hox* locus, only the anterior cluster (*Hox1* through *Hox6*) has been analyzed. The position of putative RAREs within a given genomic locus is indicated together with the sequence and the general structure of the element: two direct repeats with a one-nucleotide spacer (DR1), two direct repeats with a two-nucleotide spacer (DR2) or two direct repeats with a five-nucleotide spacer (DR5). Predicted exons are in blue (coding sequence) and pink (untranslated region), while predicted introns are in green. The numbers given above each locus indicate the size of the genomic fragment in kilobases (kb).

Evolution of Retinoid and Steroid Signaling: Vertebrate Diversification from an Amphioxus Perspective

Ricard Albalat[†], Frédéric Brunet[†], Vincent Laudet, Michael Schubert
[†]These authors contributed equally to this work.

Abstract

Although the physiological relevance of retinoids and steroids in vertebrates is very well established, the origin and evolution of the genetic machineries implicated in their metabolic pathways is still very poorly understood. We investigated the evolution of these genetic networks by conducting an exhaustive survey of components of the retinoid and steroid pathways in the genome of the invertebrate chordate amphioxus (*Branchiostoma floridae*). Due to its phylogenetic position at the base of chordates, amphioxus is a very useful model to identify and study chordate versus vertebrate innovations, both on a morphological and a genomic level. We have characterized more than 220 amphioxus genes evolutionarily related to vertebrate components of the retinoid and steroid pathways and found that, globally, amphioxus has orthologs of most of the vertebrate components of both the retinoid and steroid pathways, with some very important exceptions. For example, we failed to identify a vertebrate-like machinery for retinoid storage, transport and delivery in amphioxus and were also unable to characterize components of the adrenal steroid pathway in this invertebrate chordate. The absence of these genes from the amphioxus genome suggests that both an elaboration and a refinement of the retinoid and steroid pathways took place at the base of the vertebrate lineage. In stark contrast, we also identified massive amplifications in some amphioxus gene families, most extensively in the SDR superfamily, which, based on phylogenetic and genomic linkage analyses, were likely the result of duplications specific to the amphioxus lineage. In sum, this detailed characterization of genes implicated in retinoid and steroid signaling in amphioxus not only allows us to reconstruct an outline of these pathways in the ancestral chordate, but also to discuss functional innovations in retinoid homeostasis and steroid-dependent regulation in both cephalochordate and vertebrate evolution.

Introduction

Acting as intercellular messengers, retinoids and steroids contribute to the coordination of developmental processes and cellular functions in vertebrates. At the signaling level, both signaling systems rely on members of the nuclear hormone receptor family to regulate transcription of target genes, and, classically, the evolution of the retinoid and steroid pathways has been inferred by identification and characterization of nuclear hormone receptors in different animal lineages (Mangelsdorf et al. 1995; Escriva et al. 1997; Escriva, Delaunay, and Laudet 2000; Bertrand et al. 2004; Escriva, Bertrand, and Laudet 2004; Markov et al. 2008). In contrast, although retinoids and steroids use related machineries and similar strategies to regulate their physiological activity, the evolution of these metabolic pathways is still poorly understood.

In vertebrates, retinoid metabolism, storage and transport requires no less than 7 enzymatic activities, three binding capacities, one cell-surface receptor and one plasma protein (fig. 1*A*) (for recent reviews on retinoid metabolism see Albalat 2009; Theodosiou, Laudet, and Schubert 2010). In brief, retinoid metabolism begins when dietary β-carotenes, the main source of retinoids, are cleaved by β,β-carotene-15,15'-oxygenase (Bcmo1; also known as Bco or Bcox), which generates two molecules of retinal. This retinal is subsequently reduced to retinol (vitamin A) by retinaldehyde reductases, a heterogeneous group of enzymes of diverse protein families (Dalfó, Marqués, and Albalat 2007 and references therein). The retinol is esterified to retinyl esters by lecithin-retinol acyltransferase (Lrat) or acyl-CoA-retinol acyltransferase (Dgat1) enzymes and stored in the liver. Retinyl esters can be mobilized from the liver by retinyl ester hydrolases (including Rpe65) that transform retinyl esters back to retinol. In the bloodstream, this retinol is transported by retinol-binding protein 4 (Rbp4) and transthyretin (Ttr) and the subsequent cellular uptake of retinol is mediated by Stra6 (stimulated by retinoic acid protein 6) (Kawaguchi et al. 2007; Theodosiou, Laudet, and Schubert 2010).

Within the cell, retinol is oxidized to retinal and, in a second step, to retinoic acid (RA), the main physiologically active retinoid. In this process, retinol oxidation is the rate-limiting step and a key reaction for the physiological control of RA availability. Several dehydrogenases of the medium-chain dehydrogenase/reductases (MDR-Adh) superfamily, including Adh1, Adh3 and Adh4, and of the short-chain dehydrogenase/reductases (SDR) superfamily, such as Rdh16 (also known as Rodh4) and Rdh10, catalyze this rate-limiting reaction *in vitro* (Gough et al. 1998; Belyaeva, Johnson, and Kedishvili 2008). Retinal produced by these enzymes is then transformed to RA through the action of retinaldehyde dehydrogenases (Aldh1a and Aldh8a), which, in close coordination with RA-degrading enzymes of the cytochrome P450 subfamily 26 (Cyp26) class, regulate the spatiotemporal levels of RA (Niederreither et al. 2002; Reijntjes et al. 2005). Inside the cell, retinol, retinal and RA are generally bound to proteins that solubilize and stabilize retinoids in the aqueous environment of the cell. In vertebrates, several cellular retinol binding proteins (Crbp) and cellular RA binding proteins (Crabp) have been identified (reviewed in Noy 2000), although it is still unclear whether the main physiological task of these binding proteins is to deliver retinoids to or to sequester them from the enzymes.

Similar to retinoid metabolism, steroidogenesis (fig. 1*B*) is also a complex biochemical pathway that involves numerous cytochrome P450 (Cyp) and hydroxysteroid dehydrogenases (Hsd) enzymes (Payne and Hales 2004). In

vertebrates, steroidogenesis is initiated by Cyp11a enzymes that produce pregnenolone from cholesterol transferred from the outer to the inner membrane of mitochondria by Star (for steroidogenic acute regulatory protein, also known as Stard1) or by Stard3 (for Star-related lipid transfer domain containing 3, also known as Mln64) proteins (Clark et al. 1994; Watari et al. 1997; Stocco 2001). Pregnenolone is transformed into progesterone by 3β-hydroxysteroid dehydrogenase/Δ^5-Δ^4-isomerase (Hsd3β) or to dehydroepiandrosterone (DHEA) by Cyp17. Progesterone can be metabolized to aldosterone, cortisol or androstenedione through different pathways involving Cyp11b1, Cyp11b2, Cyp17 and Cyp21 enzymes, whereas DHEA can be transformed to androstanediol and to dihydrotestosterone (5αDHT), a potent androgen. 5αDHT can also be synthesized from androstenedione via testosterone by the action of type 1, type 2 and type 3 5α-reductases (Srd5α1, Srd5α2 and Srd5α3) (Russell and Wilson 1994; Uemura et al. 2008). Finally, androstenedione can be transformed to estrone by the aromatase enzyme Cyp19 and further to estradiol by Hsd enzymes of the SDR and aldo-keto reductases (Akr) superfamilies (reviewed in Payne and Hales 2004).

In vertebrates, pairs of Hsd enzymes participate in a pre-receptor regulatory system of steroid action through the interconversion of weak to potent hormones (Penning 2003). This pre-receptor regulation was first documented for mineralocorticoid receptors and Hsd11β enzymes (Funder et al. 1988; Oppermann, Persson, and Jörnvall 1997). Hsd11β type 1 enzymes convert inactive cortisone into the potent glucocorticoid cortisol, while Hsd11β type 2 enzymes catalyze the opposite reaction, hence neutralizing the hormone. Other pairs of activating/deactivating Hsd enzymes are: Hsd17β1 (and Hsd17β12 and Akr1c3) versus Hsd17β2 (and Hsd17β8) enzymes for estrone (weak) to estradiol (potent) conversion (and vice versa), Hsd17β3 (and Akr1c3) and Hsd17β2 (and Hsd17β8) for androstenedione (weak) to testosterone (potent) conversion (and vice versa), and Akr1c2 (and Hsd17β6, Rdh5 and Dhrs9) and Akr1c3 (and Akr1c2 and Hsd17β8) for androstanediol (weak) to 5αDHT (potent) conversion (and vice versa) (Biswas and Russell 1997; Fomitcheva et al. 1998; Penning et al. 2000; Chetyrkin et al. 2001a; Chetyrkin et al. 2001b; Huang and Luu-The 2001).

Although the retinoid and steroid metabolic pathways have extensively been studied in vertebrates, the evolutionary origins of these signaling systems are not fully understood. To investigate the evolution of these signaling systems in vertebrates, we have analyzed the genome of the invertebrate chordate amphioxus (*Branchiostoma floridae*). Phylogenetically positioned at the base of the chordate phylum, amphioxus is the extant invertebrate most closely resembling the last invertebrate ancestor of vertebrates, both in terms of morphology and genome (Schubert et al. 2006a; Holland et al. 2008; Putnam et al. 2008). We have identified more than 220 amphioxus genes evolutionarily related to vertebrate components of the retinoid and steroid metabolisms and our phylogenetic analyses allowed us to reconstruct the retinoid and steroid machineries of the ancestral chordate. Interestingly, we also found compelling evidence for massive expansions of some components of the retinoid and steroid pathways in amphioxus (most distinctively within the SDR superfamily). Based on phylogenetic and genomic linkage data, we hypothesize that these expansions arose by extensive lineage-specific gene duplications, suggesting that the retinoid and steroid pathways in amphioxus are much more complex than initially expected. Taken together, our work demonstrates

that functional diversification upon gene duplication in chordates is by no means limited to the vertebrate lineage.

Materials and Methods

Genome analyses, gene identification and sequence alignment

We have carried out exhaustive searches of the amphioxus (*B. floridae*) genome (versions 1.0 and 2.0) (www.genome.jgi-psf.org/Brafl1/Brafl1.home.html) for components of retinoid and steroid metabolism. Version 1.0 of the *B. floridae* genome was preferentially used, because it is more inclusive than version 2.0, from which some sequences have been removed (Putnam et al. 2008). Moreover, for the genomic linkage survey of the *B. floridae* sequences, we needed genome position data, which is unavailable for version 2.0 (Putnam et al. 2008).

As recommended by Kuzniar et al. (2008), we searched the amphioxus genome for orthologous sequences using precomputed sets of homologous genes. In brief, for each component of retinoid and steroid signaling, we downloaded all related sequences from the protein families defined in the Ensembl database (version 52) (www.ensembl.org) (Flicek et al. 2010), to which we added sequences from Uniprot (www.uniprot.org) and from the non-redundant protein database at NCBI (www.ncbi.nlm.nih.gov). Whenever possible, we added other families to the analysis to properly outgroup our phylogenies. Sequences identified in the genome of the cnidarian *Nematostella vectensis* (version 1.0) (www.genome.jgi-psf.org/Nemve1/Nemve1.home.html) were also included in the analyses and were used wherever possible to root the phylogenetic trees. Protein alignments were calculated with ClustalW (version 2.0) (Larkin et al. 2007) and Muscle (version 3.6) (Edgar 2004). The results of both alignment tools were merged and subsequently manually refined using SeaView (version 3.2) (Galtier, Gouy, and Gautier 1996) to exclude redundant and improperly annotated sequences. In contrast to our initial survey of the amphioxus genome, where only human sequences were used for BLAST searches (Holland et al. 2008), in the present analysis all members of a given protein family were used for BLAST searches of the complete predicted protein datasets of the amphioxus genome. This comprehensive methodology yielded an inclusive and reproducable set of amphioxus sequences potentially homologous to the different protein families of interest.

All amphioxus sequences obtained by these database searches were retrieved, sorted and allelic polymorphs were removed using nucleotide and amino acid alignments. In order to identify allelic polymorphs, nucleotide alignments were analyzed by eye and two sequences were considered as alleles, when, with the exception of a few polymorphic changes at third codon positions, large stretches of the compared sequences were identical. Intronic sequences were also aligned and conservation of intronic nucleotidic sequences stretches were used as additional criterion for the identification of allelic polymorphs. Of the different allelic polymorphs, the sequence with the longest overall alignment with the gene family in question was retained. The sequence identification numbers of all retrieved amphioxus sequences are given in the Supplementary Material online. The genomic position of each amphioxus gene was obtained from version 1.0 of the *B. floridae* genome (Putnam et al. 2008) and used as a basis for genomic linkage analyses. Automatically annotated genes were manually revised taking into consideration the data available in both version 1.0 and version 2.0 of the *B. floridae* genome and errors were corrected to maximize the similarity with known proteins. This additional verification step involving

both versions of the *B. floridae* genome yielded highly consistent results. The treated amphioxus data were subsequently incorporated into the protein family alignments used for the initial BLAST analysis of the amphioxus genome. Alignments were obtained using ClustalW (version 2.0) (Larkin et al. 2007) and Muscle (version 3.6) (Edgar 2004). The results of both alignment tools were merged and subsequently subjected to manual refinement. All alignment files are available from the authors.

Phylogenetic analyses

All sequences retrieved from the amphioxus genome were subjected to phylogenetic analyses in the context of their proper protein families. For each family, phylogenies were calculated with sequences from humans, zebrafish and amphioxus using *N. vectensis* sequences as outgroup. The SDR tree was calculated without *N. vectensis* representatives and each family hence served as outgroup for the other families, which resulted in a reliable phylogenetic reconstruction of this superfamily. The iLbp tree was also calculated without *N. vectensis* representatives, because we failed to identify iLbp-like sequences in the *N. vectensis* genome. Phylogenetic tree reconstructions were carried out using both the neighbor joining (NJ) and the maximum likelihood (ML) method. For both NJ and ML analyses, robustness of the obtained tree topologies was assessed with 1000 Bootstrap replicates. NJ trees were calculated with Phylo_Win (version 2.0) using a Poisson correction and pairwise gap removal (Galtier, Gouy, and Gautier 1996), while ML trees were obtained with PhyML (version 2.4.4) using a JTT model, a discrete gamma model with four categories, a gamma shape parameter of two and a ML estimate of the proportion of invariable sites (Guindon and Gascuel 2003). Phylogenetic trees were rooted with NJPlot (version 1.0) (pbil.univ-lyon1.fr/software/njplot.html) (Perrière and Gouy 1996) and formatted using FigTree (version 1.3.1) (tree.bio.ed.ac.uk).

Results

We have carried out an exhaustive search of the amphioxus (*B. floridae*) genome for components of retinoid and steroid metabolism. To our great surprise, we found evidence for amphioxus-specific duplications in most of the families analyzed. The best example for this tendency, with more than 100 representatives in amphioxus (compared to only 37 in humans), is the SDR superfamily. We will hence first describe our analysis of amphioxus homologs of vertebrate dehydrogenases/reductases with activity against retinol and/or hydroxysteroids, before presenting other elements implicated in retinoid and steroid metabolism.

Identification of retinol and hydroxysteroid dehydrogenases/reductases in amphioxus

SDR-Rdh and SDR-Hsd sequences

The SDR superfamily includes 73 human members with a substrate spectrum ranging from polyols, retinoids, steroids and fatty acid derivatives to xenobiotics. We have focused our analysis on those SDR families that include enzymes active against retinoids and steroids Belyaeva and Kedishvili 2006; Dalfó, Marqués, and Albalat 2007; Wu et al. 2007). According to the recent reclassification of the SDR superfamily, based on sequence analyses and predicted secondary structure (Bray, Marsden, and Oppermann 2009) and hidden Markov models (Kallberg, Oppermann, and Persson 2010), we have therefore analyzed the SDR-Rdh and SDR-Hsd enzyme families 5C, 7C, 8C, 9C, 12C, 16C, 25C, 26C, 28C, 30C, 32C, 34C, 37C and 47C, with a total of 37 members in humans. A survey of the *B. floridae* genome identified

107 genes encoding SDR enzymes belonging to these families and, hence, with putative functions in amphioxus retinoid or steroid metabolism. Of these 107 amphioxus SDRs, 6 partial sequences were excluded, because they were too short for reliable phylogenetic tree reconstruction. Phylogenetic analysis of the retained amphioxus, zebrafish and human SDR representatives combined with genomic linkage analysis of the amphioxus genes revealed the evolutionary history of the chordate Rdh/Hsd enzymes. These results will be described following the order of the individual SDR families within the overall SDR tree (fig. 2; supplementary fig. S1):

Family SDR-26C: Hsd11β1 – Hsd11β1l enzymes

Hsd11β type 1 enzymes convert the inactive glucocorticoid cortisone into the potent cortisol, and, in concert with Hsd11β type 2, regulate cortisol levels (Funder et al. 1988). Hsd11β1 enzymes are closely related to the Hsd11β1l proteins (also named Sdr10b or Hsd11β3 in chicken and fish), whose function remains unknown (Baker 2004b). The amphioxus genome contains 23 family SDR-26C genes (fig. 2; supplementary fig. S1; NJ, support value 97 of 100; ML, support value 1000 of 1000), most of them are found as tandem duplicates located on four scaffolds of the amphioxus genome (Table 1). Our phylogenetic analysis suggests that the amphioxus enzymes are orthologous to both vertebrate Hsd11β1 and Hsd11β1l, and we hence named the sequences Bf_Hsd11β1/1l_1 to 23. The genomic organization together with the tree topology suggest that the multiplicity of family SDR-26C members in amphioxus arose by lineage-specific duplication and further supports the notion that Hsd11β1 and Hsd11β1l diverged early in vertebrate evolution (Baker 2004b).

Family SDR-34C: Dhrs7 enzymes

The human *DHRS7* gene was identified from database searches of ESTs from retina and pineal gland (Haeseleer and Palczewski 2000) and, although no biochemical activity has been detected for DHRS7 so far, is expressed in the retina as well as in skeletal muscle and heart. Three amphioxus sequences, Bf_Dhrs7_1 to 3, grouped with the vertebrate Dhrs7 enzymes (fig. 2; supplementary fig. S1; NJ, 99; ML, 997). The tree topology indicates that the three amphioxus *Dhrs7* genes duplicated specifically in the cephalochordate lineage.

Family SDR-32C: Dhrs7b – Dhrs7c enzymes

Human *DHRS7b* and *DHRS7c* are predicted genes located on chromosome 17 that encode SDR proteins of unknown function. A single amphioxus sequence, called Bf_Dhrs7b/c, reliably clustered with the vertebrate Dhrs7b–Dhrs7c family (fig. 2; supplementary fig. S1; NJ, 99; ML, 996) indicating the presence of this SDR family in the last common ancestor of amphioxus and vertebrates.

Family SDR-16C: Rdh10 – RdhE2 – RdhE2l – Hsd17β11 – Hsd17β13 – Dhrs3 enzymes

Family SDR-16C includes vertebrate Rdh10, RdhE2 (for epidermal retinaldehyde dehydrogenase 2), RdhE2l (for RdhE2-like), Hsd17β11, Hsd17β13 and Dhrs3 enzymes, which have diverse activities against retinoids and steroids. Vertebrate Rdh10, RdhE2, RdhE2l and Dhrs3 have been involved in retinoid metabolism during differentiation and development, in pathological states and in the visual cycle. Hsd17β11 has been related to androgen metabolism during

steroidogenesis, and the function of Hsd17β13, a hepatic enzyme with a high degree of sequence identity with Hsd17β11, is still unknown. A total of 8 amphioxus sequences group with vertebrate family SDR-16C enzymes (fig. 2; supplementary fig. S1; NJ, 100; ML, 1000). Our phylogenetic analysis shows that one amphioxus sequence, Bf_Rdh10, is similar to Rdh10, three sequences, Bf_RdhE2_1 to 3, all of which located on the same scaffold (Table 1), are related to both RdhE2 and RdhE2l, two sequences, Bf_Hsd17β11/13_1 and 2, are associated with the vertebrate Hsd17β11–Hsd17β13 clade, and two sequences, Bf_Dhrs3_1 and 2, cluster with the vertebrate Dhrs3 enzymes, hence establishing an almost complete repertoire of amphioxus family SDR-16C members. The phylogenetic analysis thus indicates that the diversification of family SDR-16C took place before the cephalochordate-vertebrate split.

Family SDR-12C: Hsd17β3 – Hsd17β12 – HsdL1 enzymes

This family includes several enzymes involved in the production of sexual hormones and in the metabolism of fatty acids (Geissler et al. 1994; Huang et al. 2001; Moon and Horton 2003; Luu-The, Tremblay, and Labrie 2006). Vertebrate Hsd17β3 catalyzes the conversion of androstenedione to testosterone, Hsd17β12 transforms estrone into estradiol and participates in the microsomal elongation of fatty acids, and HsdL1 is highly expressed in steroidogenic tissues and its function has been related to the progression of prostate cancer. Four amphioxus enzymes belong to family SDR-12C (fig. 2; supplementary fig. S1; NJ, 100; ML, 1000). Three of them, Bf_ Hsd17β12_1 to 3, group with Hsd17β12 and one, Bf_HsdL1, clusters with HsdL1. Of these, Bf_ Hsd17β12_2 and Bf_Hsd17β12_3 are located on the same scaffold (Table 1). No sequences orthologous to vertebrate Hsd17β3 could be identified in the amphioxus genome.

Family SDR-9C: Rdh cluster – Bdh1 – Hsd11β2 – Hsd17β2 enzymes

Family SDR-9C includes three groups of vertebrate enzymes: the Rdh cluster, the Hsd type 2 group, and the 3-hydroxybutyrate dehydrogenase type 1 (Bdh1) group. Altogether, 27 amphioxus sequences group within this family (fig. 2; supplementary fig. S1; NJ, 98; ML, 999). The vertebrate Rdh cluster includes 6 human genes, *RDH5*, *RDH16*, *similar to RDH16*, *HSD17β6*, *SDR9C7* and *DHRS9.* Human RDH5 and RDH16 are involved in visual and general retinol oxidation, respectively (Gough et al. 1998; Yamamoto et al. 1999), although both of them also show some activity towards steroids. In addition, HSD17β6 and DHRS9 are also active against both retinoids and against steroids (Biswas and Russell 1997; Chetyrkin et al. 2001a; Chetyrkin et al. 2001b; Soref et al. 2001), while SDR9C7 does not have a significant activity against steroids and shows only very weak retinaldehyde reductase activity (Kowalik et al. 2009). A total of 12 amphioxus sequences, named Bf_Rdh_1 to 12, group with the vertebrate Rdh cluster (fig. 2; supplementary fig. S1), including two previously characterized retinaldehyde reductases of amphioxus (Bf_Rdh_1 and Bf_Rdh_2) (Dalfó et al. 2001; Dalfó, Marqués, and Albalat 2007). The genomic linkage data indicate that at least 7 of the 12 amphioxus Rdh genes cluster together on three distinct scaffolds (Table 1). In our phylogenetic tree, vertebrate and amphioxus Rdh sequences group separately, indicating that the extensive expansion of this family in amphioxus and the diversification of the Rdh cluster in vertebrates were evolutionary independent events.

Hsd type 2 enzymes function as NAD^+-dependent oxidases to convert active steroid hormones to their inactive forms. Hsd11β2 oxidizes glucocorticoids, transforming potent cortisol to weak cortisone, while Hsd17β2 oxidizes potent sex hormones, such as estradiol and testosterone, into weak estrone and androstenedione, respectively (Wu et al. 1993; Albiston et al. 1994). One amphioxus sequence branches outside a clade that comprises both vertebrate type 2 enzymes (fig. 2; supplementary fig. S1). The amphioxus enzyme orthologous to both vertebrate forms was therefore named Bf_Hsd11/17β2. The tree topology indicates that the 11β- and 17β-forms diverged from an ancestral Hsd-type 2 enzyme duplicated in the lineage leading to extant vertebrates. This divergence in vertebrates of the 11β- and 17β-forms, has been supported by the phylogenetic analysis of Hsd-type 2 sequences from sea urchins and acorn worms, which also cluster at the base of the vertebrate Hsd-type 2 clade (Baker 2010a).

We also identified 14 amphioxus sequences that clustered with vertebrate Bdh1 (fig. 2; supplementary fig. S1), a mitochondrial lipid-requiring enzyme that serves to interconvert ketone bodies (Marks et al. 1992). Two of the 14 amphioxus sequences, named Bf_Bdh1_1 and 2, stably group with the vertebrate enzymes, while the other 12 sequences, Bf_Bhd1_3 to 12, fall outside this cluster. Interestingly, Bf_Bdh1_1 and 2, together with Bf_Bdh1_3, are linked in the amphioxus genome, with the other 11 amphioxus Bdh1 genes clustering on two different scaffolds (Table 1). Together, the tree topology and linkage data suggest an initial duplication of a Bdh1 ancestor predating the cephalochordate-vertebrate split. While only one of these ancestral Bdh1 duplicates was preserved in vertebrates, both of them were retained and independently expanded in the amphioxus lineage.

Family SDR-28C: Hsd17β1 – Rdh8 enzymes

Vertebrate Hsd17β1 is predominantly expressed in ovaries, breast tissues and placenta, and has been considered as a major determinant of peripheral and gonadal estradiol synthesis (Peltoketo et al. 1988; Peltoketo et al. 1999). Rdh8, also known as photoreceptor retinol dehydrogenase (prRDH), localizes specifically to the outer segments of photoreceptor cells reducing retinal to retinol in the visual cycle (Rattner, Smallwood, and Nathans 2000; Maeda et al. 2005; Maeda et al. 2007). Our survey of the amphioxus genome did not reveal any amphioxus sequences clustering with this family. Nevertheless, the identification of several putative Rdh8/Hsd17β1 sequences in cnidarians (Tarrant et al. 2009) and sea urchins (Mindnich and Adamski 2009) suggests that the members of this family might specifically have been lost in the amphioxus lineage.

Family SDR-7C: Rdh11 – Rdh12 – Rdh13 – Rdh14 enzymes

Vertebrate Rdh11, Rdh12, Rdh13 and Rdh14 are retinaldehyde reductases predominantly involved in retinoid metabolism of the eye (Haeseleer et al. 2002; Kim et al. 2005; Lee et al. 2007; Maeda et al. 2007). In addition, human RDH11 is also involved in the regulation of retinoid homeostasis (Lin et al. 2001; Kedishvili et al. 2002; Belyaeva et al. 2005), RDH12 in retinal reduction of extraocular tissues (Belyaeva et al. 2005), RDH13 in protection against retinal produced from dietary β-carotenes (Belyaeva et al. 2008), and RDH14 in retinal reduction in most human tissues (Belyaeva and Kedishvili 2002). Phylogenetically, Rdh11 and Rdh12 are closely related enzymes that arose from an Rdh11/12 precursor likely during mammalian evolution, while Rdh13 and Rdh14 have diverged from the Rdh11/12 ancestor early during evolution. In amphioxus, we found 24 sequences that belonged

to the SDR family SDR-7C (fig. 2; supplementary fig. S1; NJ, 100; ML, 1000). One amphioxus sequence, Bf_Rdh13, is closely related to vertebrate Rdh13 and one, Bf_Rdh14, branches with vertebrate Rdh14, while the 22 remaining sequences, named Bf_Rdh11/12_1 to 22, are not clearly associated with any of the vertebrate SDR-7C family members. A total of 20 of the 22 Bf_Rdh11/12 sequences (Bf_Rdh11/12_3 to 22) cluster together in the phylogenetic tree, indicating a lineage-specific amplification of these amphioxus Rdh11/12 sequences. Genomic linkage of 16 of the 22 amphioxus Rdh11/12 genes on 5 distinct scaffolds (Table 1) strongly supports this notion. The internal tree topology does not unambiguously resolve the phylogenetic relationship of the amphioxus and vertebrate enzymes. The phylogenetic analysis nonetheless indicates that the chordate ancestor already had multiple genes encoding SDR-7C family enzymes, including Rdh11/12, Rdh13 and Rdh14, and that some of these ancestral sequences experienced a significant expansion in the amphioxus lineage.

Family SDR-37C: Hsd17β7 enzymes

Hsd17β7 enzymes show the typical SDR signature at the cofactor-binding motif and at the active site, but their overall sequence identity with the other SDR families is low. Hsd17β7, initially described as prolactin receptor-associated protein (PRAP) (Duan, Linzer, and Gibori 1996), are cholesterogenic enzymes that participate in cholesterol biosynthesis (Mindnich, Moller, and Adamski 2004; Shehu et al. 2008). Hsd17β7 have been lost in cholesterol auxotrophic animals that also lost squalene epoxidase (Sqle), a crucial enzyme for cholesterol biosynthesis (Marijanovic et al. 2003). There is a single Hsd17β7 ortholog in the amphioxus genome (fig. 2; supplementary fig. S1; NJ, 100; ML, 1000). The identification of Bf_Hsd17β7, together with the presence of a putative Sqle enzyme in amphioxus (GenBank accession number XP_002594656), is of evolutionary relevance, because these findings suggest that cephalochordates are capable of the *de novo* synthesis of cholesterol and hence imply that this metabolic pathway was already present in the last common ancestor of all chordates.

Family SDR-30C: Hsd17β8 enzymes

Hsd17β8 is an oxidative enzyme inactivating estradiol, testosterone and dihydrotestosterone (Fomitcheva et al. 1998). Whereas in humans and zebrafish, there is only a single Hsd17β8, we found 6 Hsd17β8 genes encoded in the amphioxus genome, Bf_ Hsd17β8_1 to 6 (fig. 2; supplementary fig. S1; NJ, 100; ML, 1000). Two of these amphioxus sequences, Bf_ Hsd17β8_3 and 4, are linked in the genome (Table 1). Altogether, these data suggest that members of the Hsd17β8 family were already present in the genome of the chordate ancestor and that the amphioxus Hsd17β8 gene complement was expanded by lineage-specific duplication.

Family SDR-25C: Dhrs4 enzymes

The vertebrate members of this SDR family have been assigned very different enzymatic activities: whereas human DHRS4 has been shown to contribute to the reduction of 3-ketosteroids into 3β-hydroxysteroids (Matsunaga et al. 2008), the mouse Dhrs4 enzyme has been described as a retinaldehyde reductase (Lei et al. 2003). Our analysis identified one amphioxus representative of this SDR family, Bf_Dhrs4 (fig. 2; supplementary fig. S1; NJ, 100; ML, 1000). Given the presence of

an amphioxus representative in this SDR family, the evolutionary origin of Dhrs4 enzymes evidently predated chordate divergence.

Family SDR-5C: Hsd17β10 enzymes

Vertebrate Hsd17β10 is a multifunctional enzyme that, in addition to a 17β-Hsd activity with sex steroids, acts as 3α-, 7α-, 7β-, 17β-, 20β- and 21-oxidase of several different substrates (Shafqat et al. 2003). A single amphioxus sequences, named Bf_ Hsd17β10, branched at the base of the vertebrate representatives of this SDR family (fig. 2; supplementary fig. S1; NJ, 100; ML, 1000) suggesting that Hsd17β10 enzymes were already present in the last common ancestor of amphioxus and vertebrates.

Family SDR-47C: Hsd17β14 enzymes

Hsd17β14 was initially suggested to function in retinol metabolism (Haeseleer and Palczewski 2000), but recent work suggests that the human enzyme acts as a 17β-Hsd with sex steroids in brain, liver and placenta (Lukacik et al. 2007). In amphioxus, there is a single member of this SDR family, Bf_Hsd17β14 (fig. 2; supplementary fig. S1; NJ, 100; ML, 1000). Hsd17β14 enzymes were thus very likely already present in the last common ancestor of all chordates suggesting that the evolutionary origin of this SDR family predates the diversification of the chordate lineage.

Family SDR-8C: Hsd17β4 enzymes

The vertebrate Hsd17β4 enzyme (also called MFE-2, MFP-2 and DBP) is involved in peroxisomal fatty acid β-oxidation and its deficiency causes severe abnormalities in several organs, but particularly in the brain (reviewed in Huyghe et al. 2006). In the amphioxus genome, we identified one sequence closely related to vertebrate Hsd17β4. This sequence, Bf_Hsd17β4, marks the base of the vertebrate Hsd17β4 enzymes (fig. 2; supplementary fig. S1; NJ, 100; ML, 1000) hence indicating that this SDR family might have already been present at the dawn of chordate diversification.

Adh and Akr sequences in amphioxus

In vertebrates, retinol dehydrogenase activity has also been associated with some members of the Adh family, mainly with the Adh1, Adh3 and Adh4 enzymes. Our survey of the *B. floridae* genome supports previous work on Adh evolution showing that amphioxus had a single Adh enzyme (Bf_Adh3) that groups with the members of the vertebrate Adh family (fig. 3*A*, supplementary fig. S2) (Cañestro et al. 2002; reviewed in Gonzalez-Duarte and Albalat 2005). Biochemically, amphioxus Adh3 is similar to vertebrate Adh3 (Cañestro et al. 2000; Godoy, Gonzalez-Duarte, and Albalat 2006) suggesting that the enzyme activities of Adh1 and Adh4 were vertebrate innovations (Cañestro et al. 2003; Cañestro, Albalat, and Postlethwait 2010).

Several members of the Akr1 group in the Akr superfamily show activity against retinoids and steroids. Akr1b enzymes that convert glucose in sorbitol can also reduce retinal to retinol (Ruiz et al. 2009), and some Akr1c enzymes can act as Hsd enzymes controlling the interconversion of weak steroid hormones to potent hormones (Dufort et al. 1999). There are 9 Akr1 sequences in humans classified into 5 subgroups (Akr1A to E). The survey of the amphioxus genome revealed 19 Akr1

sequences in amphioxus (Bf_Akr1_1 to 19) and the phylogenetic analysis grouped 18 of these 19 amphioxus sequences into four well-supported clades (fig. 3*B*, supplementary fig. S3). Genomically linked amphioxus Akr1 genes are found only within a given clade: linkage groups Akr1_2, 4 and 6 and Akr1_3 and 5 branch within the same clade; Akr1_7, 8, 9, 10 and 12 are also grouped in the tree; Akr1_14 and 16 as well as Akr1_15 and 17 together establish one of the four well-supported clades (Table 1).

The tree does not efficiently resolve the phylogenetic relationship between the 19 amphioxus Akr1 sequences and their vertebrate orthologs, but given that amphioxus Akr1s within the same clade have the tendency to be linked on the genome, the expansion of Akr1 enzymes in amphioxus was likely independent of the diversification of the vertebrate Akr1 proteins. To further investigate the lineage-specific Akr1 duplicates in amphioxus, we carried out structural analyses of their substrate binding and processing domains. In vertebrate Akr1 enzymes, in addition to the invariant residues Y50 and H113 within the active site, the amino acids at positions 47, 49, 82, 114, 116, 124, 125, 133, 135, 299 and 301 (numbering relative to human AKR1A1) have been associated with substrate binding in the binding pocket (Jez et al. 1997; Penning 1997). In amphioxus Akr1s, the conservation at these positions is highest at residues involved in general substrate recognition and binding (e.g. 100% conservation at positions 47, 50 and 113), is intermediate at residues determining sugar versus steroid specificity (e.g. 63.2% conservation at position 49 and 52.6% conservation at position 114), and is lowest at residues contributing to the definition of steroid specificity (e.g. 26.3% of conservation at position 299 and 15.8% of conservation at position 301). Altogether, this pattern is compatible with the hypothesis that at least some of the amphioxus Akr1s exhibit the capacity to metabolize steroids.

Amphioxus retinoid metabolism

Bcmo1, Bco2 and Rpe65 family

In vertebrates, the Bcmo1 enzyme (also known as Bco) cleaves β-carotene into two molecules of retinal (reviewed in Ross et al. 2000 and Gottesman, Quadro, and Blaner 2001). Bcmo1 is closely related to Bco2 (for β,β-carotene-9',10'-oxygenase) and Rpe65 (for retinal pigment epithelium protein of 65 kDa) enzymes. Bco2 catalyzes the asymmetric cleavage of β-carotenes (Kiefer et al. 2001), which has been proposed as an alternative pathway of RA synthesis (Simões-Costa, Azambuja, and Xavier-Neto 2008), while Rpe65 converts retinyl esters to retinol for photopigment regeneration in the visual cycle (Jin et al. 2005). From an evolutionary point of view, the identification of Bcmo1/Bco2/Rpe65 enzymes in urochordates (Takimoto et al. 2006) and in *Drosophila* (von Lintig and Vogt 2000) indicates that the cleavage of β-carotene is an ancient biochemical activity characterizing the bilaterian lineage.

In total, 5 amphioxus sequences showed similarity with vertebrate Bcmo1/Bco2/Rpe65 enzymes and grouped with the vertebrate enzymes in our phylogenetic analysis (fig. 3*C*, supplementary fig. S4; NJ, 94; ML, 993). The amphioxus sequences did not group with any one of the three vertebrate enzymes and are not linked in the amphioxus genome (Table 1), which is compatible with the notion that the Bcmo1, Bco2 and Rpe65 enzymes underwent independent duplication and functional specialization in the cephalochordate and the vertebrate lineages.

Aldh1a, Aldh8a and Cyp26 enzymes

In vertebrates, retinal is mainly oxidized to RA by three Aldh1a enzymes (Aldh1a1, Aldh1a2 and Aldh1a3) and endogenous RA is degraded to biologically inactive forms by Cyp26 enzymes (Cyp26a, Cyp26b and Cyp26c). Vertebrate Aldh8a1 (also known as Raldh4 and Aldh12) is also capable of oxidizing retinal (Lin and Napoli 2000; Lin et al. 2003; Liang et al. 2008), although the contribution of Aldh8a1 to RA signaling is not yet fully understood. Previous work has identified 6 Aldh1a and three Cyp26 in amphioxus, and phylogenetic analyses have suggested that this diversity of Aldh1a and Cyp26 originated by independent duplications in the amphioxus and vertebrate lineages (Cañestro et al. 2006; Marlétaz et al. 2006). Genomic linkage data provide additional evidence for this hypothesis: 5 of the 6 amphioxus Aldh1a genes are located on two genome scaffolds and all three amphioxus Cyp26 genes are clustered on a single scaffold (Table 1).

We have extended the previous analyses by searching for Aldh8a1 orthologs in amphioxus and identified one amphioxus sequence, Bf_Aldh8a. Phylogenetic analysis showed that Bf_Aldh8a is orthologous to vertebrate Aldh8a1 (fig. 3*D*, supplementary fig. S5; NJ, 100; ML, 1000). The identification of Bf_Aldh8a completes the evolutionary analysis in amphioxus of the basic genetic machinery controlling the spatiotemporal levels of endogenous RA (i.e. Aldh1a, Aldh8a and Cyp26 enzymes) and further strengthens the notion that this machinery was already present in the chordate ancestor (Cañestro et al. 2006; Marlétaz et al. 2006; Campo-Paysaa et al. 2008; Albalat and Cañestro 2009; Theodosiou, Laudet, and Schubert 2010).

Cellular retinoid binding proteins: Crbp and Crabp

Intracellular retinoid binding proteins, such as Crbp and Crabp, bind retinol, retinal and RA to solubilize and stabilize the retinoids in the aqueous environment of the cell. Vertebrate Crbp and Crabp belong to the family of intracellular lipid binding proteins (iLbp), together with three groups of fatty acid-binding proteins (Fabps), namely Fabp1/6, Fabp2 and Fabp3/4/5/7/8/9 (Schaap, van der Vusse, and Glatz 2002). In humans, there are four Crbp, two Crabp, one Fabp2, two Fabp1/6 and 6 Fabp3/4/5/7/8/9, each one of these proteins being characterized by different ligand binding properties. In amphioxus, we identified 11 iLbp. While one of these amphioxus iLbp, Bf_Crbp, strongly groups with vertebrate Crbp (fig. 3*E*, supplementary fig. S6; NJ, 97; ML, 998) and one, Bf_Fabp1/6, groups with vertebrate Fab1/6 (fig. 3*E*, supplementary fig. S6; NJ, 74; ML, 958), the other 9 amphioxus sequences, named Bf_iLbp_1 to 9, are not clearly associated with any vertebrate iLbp group. Since 7 of the 9 amphioxus iLbp genes are clustered on two genome scaffolds (Table 1), it is very likely that these amphioxus genes originated by lineage-specific duplication.

Unlike for Crbp, we were hence unable to identify a clear amphioxus ortholog of vertebrate Crabp. Several iLbp family members with retinoid-binding capacity have been described in protostomes, such as insects and crustaceans (Mansfield et al. 1998; Gu et al. 2002; Folli et al. 2005; Söderhäll et al. 2006), but sequence analyses and 3D structural modeling showed that these protostome iLbps are not orthologous to vertebrate Crbp and Crabp (Gu et al. 2002; Folli et al. 2005; Söderhäll et al. 2006). It is therefore very likely that the capacity to bind retinoids has evolved independently in different iLbp families in the course of bilaterian diversification. Biochemical analyses will hence be required to determine the exact physiological function of each amphioxus member of the iLbp family.

Lrat and Dgat1 enzymes

The Lrat enzyme catalyzes retinol esterification into retinyl esters, which are accumulated in the vertebrate liver where they can be mobilized to maintain retinoid homeostasis. Lrats hence participate in retinoid storage mechanisms. Lrat enzymes belong to the Lrat-like family of the complex NlpC/P60 superfamily, which unites Lrat with the Hrasl (for Hras-like suppressor, also known as H-rev107-like protein) and Fam84 (for family with sequence similarity 84) subfamilies (Anantharaman and Aravind 2003). We found 13 amphioxus sequences with similarities to vertebrate Lrat-like proteins. Only one of these sequences grouped with one of the three vertebrate subfamilies: Bf_Hrasls, which is weakly associated with the vertebrate Hrasls (fig. 3*F*, supplementary fig. S7; NJ, 38; ML, 773). The other 12 amphioxus sequences are located on only three scaffolds of the genome (Table 1) and cluster together in the phylogenetic tree (fig. 3*F*, supplementary fig. S7), indicating that they are the result of an extensive lineage-specific duplication. These 12 amphioxus sequences were called Bf_ Fam84/Hrasls/Lrat_1 to 12.

Because phylogenetic analysis did not properly resolve the relationships between amphioxus and vertebrate enzymes, we analyzed the amino acid signatures of the different Lrat-like subfamilies. Lrat-like proteins share two catalytic residues, H60 and C161 (numbers relative to the human LRAT enzyme) (Mondal et al. 2000; Jahng, Xue, and Rando 2003). C161 is the central cysteine of the highly conserved NCE-box of Lrat and Hrasls proteins, whereas in Fam84 members this cysteine has been replaced by a serine (resulting in a NSE-box). None of the 13 amphioxus sequences exhibits a NSE-box, suggesting that this amino acid replacement in the Fam84 subfamily took place only after the split of the amphioxus and vertebrate lineages. For H60, the conserved amino acid context is HYGIY in Lrat enzymes, $HW^{A}/_{G}{}^{L}/I/_{V}Y$ in Hrasls, and $HWA^{V}/_{I}{}^{F}/Y/_{C}$ in Fam84. For this signature, the 13 amphioxus sequences all show a Hrasls $HW^{A}/_{G}{}^{L}/I/_{V}Y$ context, suggesting that the activity of the amphioxus Lrat-like proteins might be most similar to that of vertebrate Hrasls. In this context it is interesting to note that the Lrat-like protein EGL-26 of the nematode worm *Caenorhabditis elegans* has been functionally implicated in lipid metabolism (Estes, Kalamegham, and Hanna-Rose 2007), a role more similar to that of vertebrate Hrasls (Jin et al. 2007; Nazarenko, Schafer, and Sers 2007) than to that of Lrats. Taken together, this suggests a Hrasls-like activity might be ancestral in this family and that Lrat- and Fam84-like activities arose specifically in the vertebrate lineage after duplication of an ancestral Hrasls-like enzyme.

In addition to Lrat proteins, enzymes involved in the synthesis of triacylglycerol, such as acyl coenzyme A:diacylglycerol acyltransferase (Dgat1), are capable of contributing to retinol esterification by carrying out the acyl coenzyme A:retinol acyltransferase reaction (Ross 1982; O'Byrne et al. 2005; Orland et al. 2005). Although the function of vertebrate Dgat1 in the hepatic control of retinoid homeostasis has been questioned (Batten et al. 2004), in the absence of a clear amphioxus Lrat ortholog, the contribution of Dgat1 enzymes in amphioxus to the esterification of retinol for retinyl ester storage deserves further investigation. Dgat1 enzymes are related to sterol O-acyltransferases (Soat) that catalyze the formation of cholesterol esters from cholesterol. In humans, there are one Dgat1 and two Soat enzymes. We found two sequences similar to vertebrate Dgat1 and Soat enzymes in the amphioxus genome and phylogenetic analysis showed that one of theses sequences, named Bf_Dgat1, groups with vertebrate Dgat1, while the other one, called Bf_Soat1/2, branches with vertebrate Soat1 and Soat2 enzymes (fig. 3*G*, supplementary fig. S8; NJ, 100; ML, 980). Taken together, these data suggest that, in

the absence of a clear amphioxus Lrat ortholog, retinol esterification in amphioxus might be catalyzed by Dgat, an enzyme that might have hence played a key role in the biosynthesis of retinyl esters in the ancestral chordate.

Rbp4, Ttr and Stra6 proteins

In vertebrates, hepatic retinyl esters can be hydrolyzed back to retinol, which subsequently enters the blood stream in association with Rbp4 and Ttr proteins. Rbp4 belongs to the lipocalin family, a large group of extracellular proteins that bind and transport small hydrophobic molecules. While the evolutionary origins of Rbp4 remain unclear, another member of this family, apolipoprotein D (ApoD), is considered the most ancient metazoan lipocalin. Vertebrate Rbp4 is evolutionary related to the prostaglandin D2 synthases (Ptgds)-like group that includes Ptgds, alpha-1-microglobulin/bikunin preproprotein (Ambp) and apolipoprotein M (ApoM) (Flower 1996; Ganfornina et al. 2000; Gutiérrez, Ganfornina, and Sánchez 2000).

In amphioxus, we found 6 lipocalin-like sequences. A total of 5 of those, Bf_ApoD_1 to 5, are similar to the vertebrate ApoD proteins, while one sequence is more closely related to the vertebrate Ptgds-like group (fig. 3*H*, supplementary fig. S9; NJ, 72; ML, 761). The amphioxus sequence was named Bf_Ambp/ApoM/Ptgds/Rbp4 to reflect this phylogenetic relationship. Since the exon-intron organization of lipocalin genes contains information about the evolutionary origins of this gene family (Sánchez et al. 2003), we decided to analyze the genomic locus of *Bf_Ambp/ApoM/Ptgds/Rbp4*. The amphioxus *Ambp/ApoM/Ptgds/Rbp4* gene has 6 coding exons and 5 introns with a 0,2,1,1,1 pattern of intron phases, which is identical to vertebrate *ApoM* and *Ptgds* genes, but different from vertebrate *Rbp4* genes (Sánchez et al. 2003). Therefore, it seems that the lipocalin gene of the ancestral chordate was structurally more similar to extant *ApoM* and *Ptgds* genes than to *Rbp4* genes.

Rbp4 interacts with transthyretin (Ttr) to transport and distribute retinol to peripheral tissues. Vertebrate Ttr proteins are evolutionary related to 5-hydroxyisourate hydrolases (Urah) enzymes (Zanotti et al. 2006). An amphioxus sequence similar to vertebrate Ttr and Urah was identified in our genomic survey. The phylogenetic analysis was unable to resolve, whether the amphioxus sequence is orthologous to either vertebrate Ttr or Urah, or to both, and we hence named the sequence Bf_Ttr/Urah (fig. 3*I*, supplementary fig. S10). Vertebrate Ttr and Urah, however, exhibit different amino acids at key positions for substrate accommodation and catalysis: H11, D49, R51, H102, Y115 and S118 in Urah (numbers refer to the mouse Urah sequence, because no human Urah has so far been identified) versus K35, S72, E74, T126, T139 and V142 in Ttr (numbers refer to the human TTR sequence). In the amphioxus Ttr/Urah, the amino acids at these key positions correspond more closely to those of vertebrate Urah, which suggests a 5-hydroxyisourate hydrolase activity of the amphioxus protein. These data therefore support the notion that the hormone transport protein Ttr arose during vertebrate evolution by duplication from an ancestral Urah-like enzyme (Zanotti et al. 2006).

After transport to the target tissue, retinol uptake into the cell is mediated by Stra6, a cell-surface receptor for Rbp4 (Kawaguchi et al. 2007; Theodosiou, Laudet, and Schubert 2010). Stra6 proteins show similarity with a group of vertebrate proteins of unknown function, called Stra6l (for Stra6-like). We identified four amphioxus sequences with similarity to vertebrate Stra6 in the amphioxus genome, but phylogenetic reconstruction groups the amphioxus sequences with the vertebrate Stra6l proteins (fig. 3*J*, supplementary fig. S11; NJ, 74; ML, 982). We concluded,

therefore, that there is no Stra6 ortholog in amphioxus, suggesting that Stra6 was either lost in cephalochordates or originated specifically in the vertebrate lineage.

In summary, our searches for Rbp4, Ttr and Stra6 orthologs in the amphioxus genome suggest that amphioxus probably lacks a vertebrate-like system of retinoid transport and uptake. Functional studies are needed to determine, whether some of the sequences we have identified in the genome might participate in an alternative system in amphioxus for the mobilization and delivery of retinoids from storage reservoirs and hence whether some kind of retinoid storage, mobilization and delivery system was active in the ancestral chordate.

Steroidogenesis enzymes in amphioxus

Star and Cyp11a

The first reaction in the steroidogenesis pathway is the synthesis of pregnenolone from cholesterol in the inner membrane of mitochondria. In vertebrates, cholesterol is delivered to mitochondria by members of the Star family (Clark et al. 1994; Watari et al. 1997; Stocco 2001), all of which contain a START domain required for intracellular lipid transport, lipid metabolism and cell signaling processes. Members of the Star family include multiple Stard (for START domain) proteins, Dlc1 (for deleted in liver cancer 1), Col4a3bp (for collagen, type IV, alpha 3 binding protein) and Pctp (for phosphatidylcholine transfer protein). We searched the *B. floridae* genome for Star family members and found 6 amphioxus sequences similar to vertebrate Stard proteins that, based on their phylogenetic relationships, were named Bf_Star/Stard3, Bf_Stard7, Bd_Stard10, Bf_Dlc1/Stard8/13, Bf_Col4a3bp and Bf_Pctp (fig. 3*K*, supplementary fig. S12). In our phylogenetic analysis, Bf_Star/Stard3 reliably groups with vertebrate Star/Stard3 proteins (fig. 3*K*, supplementary fig. S12; NJ, 85; ML, 999). This amphioxus sequence is therefore a good candidate to mediate the delivery of cholesterol to the inner mitochondrial membrane of amphioxus.

Once in the mitochondrion, synthesis of pregnenolone from cholesterol is carried out by Cyp11a enzymes, which belong to the mitochondrial clan of the cytochrome P450 enzymes. The mitochondrial Cyp clan includes vertebrate Cyp11, Cyp24 and Cyp27 as well as some invertebrate Cyp enzymes (Nelson 2009). Our survey of the amphioxus genome identified 12 sequences grouping within the mitochondrial Cyp clan (fig. 3*L*, supplementary fig. S13; NJ, 97; ML, 1000). The phylogenetic analyses do not resolve the evolutionary relationships between the amphioxus and vertebrate sequences. In total, 8 sequences, named Bf_Cyp11/24/27_1 to 8, grouped at the base of the vertebrate mitochondrial Cyp enzymes, three sequences, Bf_Cyp24/27_1 to 3, seem closely related to vertebrate Cyp24 and Cyp27, and one amphioxus sequence, Bf_Cyp27, stably groups with vertebrate Cyp27 (fig. 3*L*, supplementary fig. S13; NJ, 46; ML, 988). Of the amphioxus mitochondrial Cyp clan members, 6 of the 8 Bf_Cyp11/24/27 genes are genomically linked, clustering on a single genome scaffold (Table 1).

Given the absence of crystal structures for mitochondrial Cyp enzymes, only very limited structural information of these proteins is available (Storbeck, Swart, and Swart 2007). We were hence not able to carry out meaningful structural comparisons between the amphioxus and vertebrate sequences to assist with the inference of evolutionary relationships between the different members of this Cyp clan. In conclusion, numerous amphioxus mitochondrial Cyp sequences were identified, but

the physiological contribution of these enzymes to the synthesis of pregnenolone from cholesterol remains to be determined.

Hsd3β enzymes

Pregnenolone is metabolized to progesterone by Hsd3β enzymes. Hsd3β enzymes are unconventional SDR enzymes that belong to the extended category of SDR proteins (Bray, Marsden, and Oppermann 2009). For the sake of clarity, they were not included in the global SDR analysis, but have been studied separately. The Hsd3β clade includes three human enzymes, namely HSD3β1, HSD3β2 and HSD3β7, which are evolutionarily related to Nsdhl (for NAD(P)-dependent steroid dehydrogenase-like) and Sdr42e1 (for short chain dehydrogenase/reductase family 42E, member 1) enzymes (Bray, Marsden, and Oppermann 2009). We identified 9 sequences similar to these vertebrate enzymes in the amphioxus genome. Phylogenetic analysis shows that 7 sequences, Bf_Hsd3β_1 to 7, are similar to vertebrate Hsd3β (fig. 3*M*, supplementary fig. S14; NJ, 100; ML, 1000), one sequence, Bf_Nsdhl, groups with vertebrate Nsdhl (fig. 3*M*, supplementary fig. S14; NJ, 100; ML, 1000) and one sequence, Bf_Sdr42e1, is associated with vertebrate Sdr42e1 (fig. 3*M*, supplementary fig. S14; NJ, 100; ML, 1000). Although the 7 amphioxus Hsd3β genes do not seem to be linked in the genome (Table 1), tree topology nonetheless clearly indicates that the amphioxus genome contains sequences orthologous to vertebrate Hsd3β and that these multiple amphioxus Hsd3β arose by linage-specific duplication, independent of the Hsd3β diversification of vertebrates.

Cyp17, Cyp19 and Cyp21 enzymes

In steroidogenesis, vertebrate Cyp17 converts progesterone to 17α-OH-progesterone and to androstenedione, a precursor of active androgens and estrogens. Cyp17 is also capable to transform pregnenolone to 17α-hydroxypregnenolone and to dehydroepiandrosterone (DHEA). Cyp17 belongs to clan 2 of the cytochrome P450 superfamily, which also includes Cyp1, Cyp2 and Cyp21. We found 12 *B. floridae* sequences belonging to this Cyp clan 2 (fig. 3*L*, supplementary fig. S13; NJ, 95; ML, 1000). Two of these, Bf_Cyp17_1 and 2, are orthologous to vertebrate Cyp17, while the other 10, Bf_Cyp2_1 to 10, group with vertebrate Cyp2. Interestingly, our phylogenetic analysis suggests that there is no amphioxus ortholog of vertebrate Cyp21, which is an essential enzyme for the synthesis of corticosteroids in vertebrates.

In vertebrates, androstenedione produced by Cyp17 enzymes is transformed by Cyp19, the so-called aromatase, to estrone. Cyp19 enzymes have also been involved in the synthesis of estradiol from testosterone. In the *B. floridae* genome, there are two sequences, Bf_Cyp19_1 and 2, that are orthologous to vertebrate Cyp19 (fig. 3*L*, supplementary fig. S13; NJ, 100; ML, 1000). These two amphioxus Cyp19 sequences are most likely the result of a lineage-specific duplication.

Srd5α enzymes

Testosterone can be transformed to 5α-dihydrotestosterone (5αDHT), a potent sex steroid, by the action of type 1, type 2 and type 3 Srd5α enzymes (Russell and Wilson 1994; Uemura et al. 2008). Srd5α enzymes are found in invertebrates and vertebrates and are phylogenetically related to the Tecr enzymes (for trans-2,3-enoyl-CoA reductase, also known as glycoprotein synaptic 2 or Gpsn2) (Markov et al.

2009). We identified two *B. floridae* sequences, named Bf_Srd5α1/2_1 and 2, that group within the vertebrate Srd5α1/2 clade (fig. 3*N*, supplementary fig. S15; NJ, 97; ML, 1000), one sequence, Bf_Srd5α3, that branches with the vertebrate Srd5α3 enzymes (fig. 3*N*, supplementary fig. S15; NJ, 99; ML, 999) and one sequence, Bf_Tecr, that clusters with the vertebrate Tecr proteins (fig. 3*N*, supplementary fig. S15; NJ, 100; ML, 1000). These phylogenetic data support the notion that the amphioxus genome encodes enzymes with Srd5α activity and that some of these amphioxus genes arose by lineage-specific duplication.

Discussion

In this study, we have investigated the evolutionary history of the chordate retinoid and steroid signaling systems by carrying out an exhaustive analysis of the amphioxus genome. Our evolutionary inferences are based on the premises that orthologs typically have similar functions (Tatusov, Koonin, and Lipman 1997; Eisen 1998; Gabaldón and Huynen 2004) and that integrative functional predictions combining phylogenetic information, genomic linkage data, concerted gene gains or losses and structural analyses are a powerful tool to predict the biological processes an orthologous protein participates in (Gabaldón and Huynen 2004). Using this combinatorial approach, we have thus evaluated the presence of the genetic machinery for retinoid and steroid signaling in amphioxus (fig. 4). We will now explore the implications of our findings on the evolutionary history of retinoid and steroid signaling in chordates, paying special attention on the evolution of new physiological capabilities in the vertebrate lineage.

Evolution of retinoid signaling

In the vertebrate RA signaling cascade, several enzymes, binding proteins and transport elements constitute complex machineries that control the physiological levels of RA. Our global analysis of the genome of the cephalochordate amphioxus, the extant species most closely resembling the last invertebrate ancestor of vertebrates (Schubert et al. 2006a; Holland et al. 2008; Putnam et al. 2008), gives us the unique opportunity to attempt the reconstruction of the RA pathway in the invertebrate ancestor and hence to evaluate the roles RA signaling might have played in the evolutionary diversification of vertebrates (fig. 4).

We searched the genome of the amphioxus *B. floridae* for sequences similar to all vertebrate proteins involved in the retinoid pathway (fig. 1). Phylogenetic, genomic and structural analyses indicate that amphioxus has (i) SDR-retinol dehydrogenases putatively capable to catalyze retinol oxidation and retinal reduction (Bf_Rdh_1 to 12, Bf_Rdh10, Bf_Rdh11/12_1 to 22, Bf_Rdh13, Bf_Rdh14, Bf_Dhrs3_1 and 2, and Bf_Dhrs4), (ii) enzymes for RA synthesis and degradation (Bf_Aldh1a_1 to 6, Bf_Aldh8a and Bf_Cyp26_1 to 3), (iii) enzymes likely involved in β-carotene cleavage (Bf_Bcmo1/Bco2/Rpe65_1 to 5) and (iv) candidates for lipid binding proteins possibly involved in retinoid stabilization and protection within the cell (Bf_Crbp and Bf_iLbp_1 to 9). Our screening of the amphioxus genome is consistent with the results of previous analyses demonstrating the presence in amphioxus of single copies of retinoic acid and retinoid X receptors (RAR and RXR, respectively) (Holland and Holland 1996; Cañestro et al. 2001; Escriva et al. 2002; Schubert et al. 2004; Schubert et al. 2005; Schubert et al. 2006b; Koop et al. 2010) and of active retinoid metabolism (Dalfó et al. 2001; Dalfó et al. 2002). Moreover, the lineage-specific duplications of some key components of the retinoid pathway might

actually reflect very intricate control mechanisms for retinoid metabolism throughout the amphioxus life cycle.

In contrast, our searches for Lrat, Rbp4, Ttr and Stra6 suggest that amphioxus might lack counterparts of these components of vertebrate retinoid signaling. The amphioxus genome contains sequences related to these proteins, but structural analyses indicate that the amphioxus representatives of these families more closely resemble vertebrate proteins with functions that are not directly linked to the retinoid pathway. This concerted absence of the four main components involved in storage, transport and cellular uptake of retinoids suggests that this system might be a functional novelty of vertebrates. Elaboration of this axis of retinoid metabolism might have enabled vertebrates to improve the control of retinoid homeostasis and to compensate retinoid fluctuations in natural environments.

Evolution of steroid signaling

Comparative analyses of steroidogenic enzymes has led to the conclusion that some enzymatic activities arose independently in arthropods and vertebrates and, therefore, that steroidogenesis has been elaborated in parallel in the three main bilaterian lineages (Markov et al. 2009). The origin of the enzymatic activities within the chordate phylum, however, remains uncertain and the survey of steroidogenic enzymes (fig. 1) in amphioxus is crucial to obtain insights into the evolution of the steroid pathway in chordates and vertebrates (fig. 4).

Our analysis reveals that, with some minor and very intriguing exceptions, the amphioxus genome contains most of the genetic machinery for steroid metabolism. The repertoire of steroidogenic enzymes is even significantly larger than previously anticipated (Baker 2004a). Amphioxus has orthologs of vertebrate Cyp17 (Bf_Cyp17_1 and 2), Cyp19 (Bf_Cyp19_1 and 2), Star (Bf_Star/Stard3), Hsd3β (Bf_Hsd3β_1 to 7), Srd5α (Bf_Srd5α1/2_1 and 2, and Bf_Srd5α3) and of mitochondrial Cyp enzymes evolutionarily related to vertebrate Cyp11a (Bf_Cyp11/24/27_1 to 8). Although Cyp19 and Hsd3β activities have been detected in amphioxus gonadal tissues (Callard et al. 1984; Mizuta et al. 2008) and although docking studies predict, for example, that Cyp19 can bind androgens (Callard et al. 2011), the biochemical activity of most of these amphioxus enzymes still requires experimental confirmation. Nevertheless, based on this collection of amphioxus enzymes and supported by the presence of sex steroids in amphioxus (Chang, Liu, and Zhu 1985; Fang, Weng, and Hu 2001; Takeda, Kubokawa, and Matsumoto 2003; Mizuta and Kubokawa 2007), we hypothesize that, while a vertebrate-like sex steroid metabolism exists in cephalochordates. In contrast, amphioxus probably lacks Cyp21, which in vertebrates is an enzyme instrumental for corticosteroid synthesis. The absence of Cyp21 and the fact that C21 hydroxylated steroids could not be detected after incubation of amphioxus tissues with pregnenolone (Mizuta et al. 2008) together suggest that cephalochordates might not be capable of synthesizing corticosteroids and therefore that adrenal steroid hormones were a vertebrate innovation.

In amphioxus, sex steroids, including progesterone, estrone, estradiol and testosterone, have been detected by RIA and immunohistochemical methods (Chang, Liu, and Zhu 1985; Fang, Weng, and Hu 2001; Takeda, Kubokawa, and Matsumoto 2003; Mizuta and Kubokawa 2007), and two steroid receptors, ER and SR, have been characterized (Bridgham et al. 2008; Paris et al. 2008; Katsu et al. 2010; Callard et al. 2011). Amphioxus ER is orthologous to vertebrate estrogen receptor (ER) and amphioxus SR is orthologous to the vertebrate steroid receptors

(SR), including androgen receptor (AR), glucocorticoid receptor (GR), mineralocorticoid receptor (MR) and progesterone receptor (PR) (Bridgham et al. 2008; Paris et al. 2008; Schubert et al. 2008; Katsu et al. 2010; Callard et al. 2011). While amphioxus ER does not activate transcription and seems to be an estrogen insensitive inhibitor of SR activity (Bridgham et al. 2008; Katsu et al. 2010; Callard et al. 2011), amphioxus SR is activated by estrogens and mediates transcriptional activation through ER- and AR-like DNA response elements (Bridgham et al. 2008; Katsu et al. 2010). Although the binding of amphioxus ER and SR to other sex steroids, such as progesterone or testosterone, has not been demonstrated, 3D modeling of amphioxus SR suggests that it may also bind androgen steroids (Baker and Chang 2009). In sum, given the relative complexity of the amphioxus steroidogenic system due to various lineage-specific duplications, the study of the enzymatic activities of the different steroidogenic enzymes in amphioxus, the complete analysis of the binding capacities of the two amphioxus steroid receptors, and the detailed assessment of the physiological roles of steroids in amphioxus are crucial for understanding the elaboration of steroid-dependent signaling not only in cephalochordates, but also in vertebrates.

Evolution of pre-receptor regulation

In the steroid pathway, hydroxysteroid dehydrogenases regulate the steroid response in vertebrates by controlling hormone activation and inactivation, a mechanism known as pre-receptor regulation of steroid action (reviewed in Penning 2003). Vertebrate Hsd17β control sex steroids, while Hsd11β regulate adrenal hormones. Amphioxus has many Hsd17β enzymes: one amphioxus Hsd11/17β2, 6 Hsd17β8, one Hsd17β10, and two Hsd17β11/13 are candidates to catalyze oxidation of sex steroids, while one Hsd17β7, three Hsd17β12 and one Hsd17β14 might be responsible of steroid reduction. Amphioxus therefore has the genetic machinery necessary for enzymatically activating and inactivating sex steroids, pushing back in the evolutionary origins for pre-receptor control of sex steroid response to the origin of chordates. In contrast, given the likely absence of the corticosteroidogenic Cyp21 enzyme in amphioxus for synthesis of adrenal steroids, a pre-receptor control for corticosteroid action is probably not required in amphioxus. Even though the amphioxus genome contains 23 Hsd11β1/1l-like and one Hsd11/17β2-like sequences, evolutionary and functional studies are compatible with these enzymes metabolizing steroids other than corticosteroids (Kusakabe, Nakamura, and Young 2003; Baker 2004b; Baker 2010b). The evolution of Hsd11β1 and Hsd11β2 activities against corticosteroids would thus be functional innovations of vertebrates that allowed the elaboration of a pre-receptor control of the adrenal hormone response in the vertebrate lineage.

In vertebrates, retinoid action is tightly controlled spatiotemporally by the coordinated activity of enzymes synthesizing (i.e. Aldh1a and Aldh8a) and degrading (i.e. Cyp26) endogenous RA (Niederreither et al. 2002; Reijntjes et al. 2005; Campo-Paysaa et al. 2008; Theodosiou, Laudet, and Schubert 2010). This metabolic system controlling the physiological levels of retinoids available for receptor occupancy is reminiscent of the pre-receptor system for steroid action. Our phylogenetic data indicate that this pre-receptor regulation is also functioning in amphioxus and was hence already present in the last common ancestor of amphioxus and vertebrates. Moreover, since in amphioxus the components of retinoid pre-receptor regulation have been subjected to lineage-specific duplications (there are 6 Aldh1a, one Aldh8a and three Cyp26 enzymes encoded in the amphioxus genome), it is possible that this

system has secondarily been elaborated in this lineage, creating a rather complex metabolic regulation of retinoid signaling, which contrasts the simplicity of the retinoid receptor system, which in cephalochordates is based on a single RAR and a unique RXR.

Acknowledgements

The authors would like to thank Cristian Cañestro and Christophe Tiffoche for fruitful discussions and critical reading of the manuscript. We are also grateful to Diego Sánchez for his advice on the evolution of the lipocalin family. This research was supported by funds from ANR (ANR-07-BLAN-0038 and ANR-09-BLAN-0262-02) and CNRS to MS, by CRESCENDO, a European Union Integrated Project of FP6, to VL and MS, by a grant from the Ministerio de Ciencia e Innovación (BFU2010-14875) to RA and by the GDRE-RA Comparative Genomics to RA and FB.

Table 1. Genomic linkage in amphioxus of retinoid and steroid metabolism genes.

Superfamily name	Amphioxus family name	Linkage groups ([scaffold of the amphioxus genome]: amphioxus sequence name)
SDR	Hsd11β1/1l	[32]:12,19,21 [43]:13,14,15,16,18,20,22 [45]:23 [66]:2,3,4 [87]:1 [134]:6 [157]:7 [227]:17 [295]:8,9,10,11 [405]:5
	Dhrs7	[34]:1 [42]:3 [46]:2
	Dhrs7b/c	[5]
	Rdh10	[181]
	RdhE2	[1]:1,2,3
	Hsd17β11/13	[191]:2 [282]:1
	Dhrs3	[469]:2 [496]:1
	Hsd17β12	[44]:2,3 [208]:1
	HsdL1	[34]
	Rdh cluster	[34]:10,11 [44]:4 [72]:6 [110]:7 [237]:8,9 [311]:12 [373]:5 [452]:1,2,3
	Bdh1	[40]:5,7,9,11,14 [79]:1,2,3 [159]:4,6,8,10,12,13
	Hsd11/17β2	[167]
	Rdh11/12	[7]:5 [8]:13,15,16 [9]:1 [18]:6,9,10,11,12 [52]:7 [74]:2 [87]:8 [180]:14 [189]:3,4 [196]:17,18 [376]:19,20,21,22
	Rdh13	[18]
	Rdh14	[84]
	Hsd17β7	[214]
	Hsd17β8	[5]:3,4 [69]:2 [190]:5 [238]:1 [723]:6
	Dhrs4	[390]
	Hsd17β10	[10]
	Hsd17β14	[3]
	Hsd17β4	[193]
Adh	Adh3	[151]
Akr1	Akr1	[3]:13 [59]:7,8,9,10,12 [89]:14,16 [229]:2,4,6 [236]:18 [258]:3,5 [281]:15,17 [374]:19 [391]:1 [551]:11
Bco	Bcmo1/Bco2/Rpe65	[12]:1 [44]:2 [74]:5 [161]:3 [555]:4
Aldh	Aldh2	[118]
	Aldh1a	[21]:2 [31]:1,5,6 [155]:3,4
	Aldh8a	[1]
	Aldh9a	[152]
iLbp	Crbp	[9]
	iLbp	[25]:2,4,5,6 [46]:7 [55]:1,8,9 [104]:3
	Fabp1/6	[12]
NlpC/P60	Hrasls	[103]
	Fam84/Hrasls/Lrat	[26]:1,2,4,6,10 [135]:7,12 [265]:3,5,8,9,11
Dgat	Dgat1	[76]
	Soat1/2	[217]
Lipocalin	Ambp/ApoM/Ptgds/Rbp4	[11]
	Apod	[30]:2 [87]:1 [168]:3,4,5
Ttr/Urah	Ttr/Urah	[60]
Stra6	Stra6l	[4]:1,3 [271]:2 [350]:4
Star	Star/Stard3	[53]
	Col4a3bp	[15]
	Dlc1/Stard8/13	[2]
	Stard10	[2]
	Stard7	[338]
	Pctp	[166]
Cyp	Cyp27	[25]
	Cyp24/27	[3]:3 [28]:2 [140]:1
	Cyp11/24/27	[44]:2 [191]:8 [214]:1,3,4,5,6,7
	Cyp3	[7]:1 [7]:2
	Cyp2	[44]:4,6,7,10 [61]:1,3 [101]:9 [438]:8 [682]:2 [736]:5
	Cyp17	[157]:2 [248]:1
	Cyp26	[164]:1,2,3
	Cyp19	[9]:1 [484]:2
Hsd3β	Hsd3β	[1]:3 [7]:1 [31]:2 [33]:5 [37]:6 [68]:4 [89]:7
	Sdr42e1	[5]
	Nsdhl	[2]
Srd5α	Srd5α1/2	[1]:1 [81]:2
	Srd5α3	[107]
	Tecr	[41]

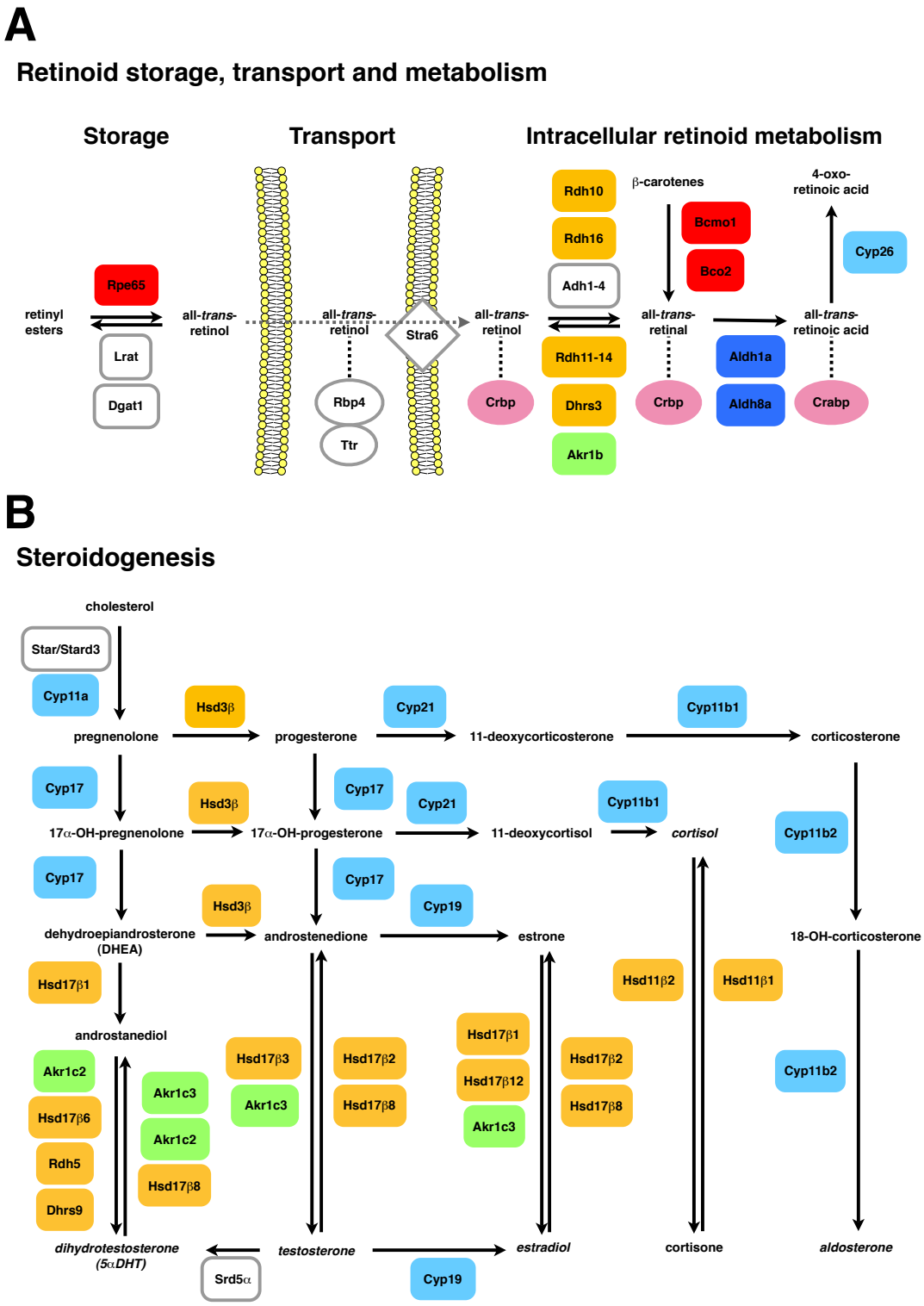


Fig. 1. Vertebrate retinoid (*A*) and steroid (*B*) metabolisms. (*A*) The main components of retinoid metabolism, storage and transport in vertebrates are shown. Proteins are boxed and retinoids are in black. Proteins of the same family are shown in identical colors. Enzymatic reactions are shown with arrows. Dotted arrow indicates retinoid transport and dotted lines highlight retinoid binding to proteins. (*B*) The main enzymatic reactions of steroidogenesis in vertebrates are indicated. Enzymatic reactions are shown with arrows. Proteins are boxed and steroids are in black. Proteins of the same family are shown in identical colors.

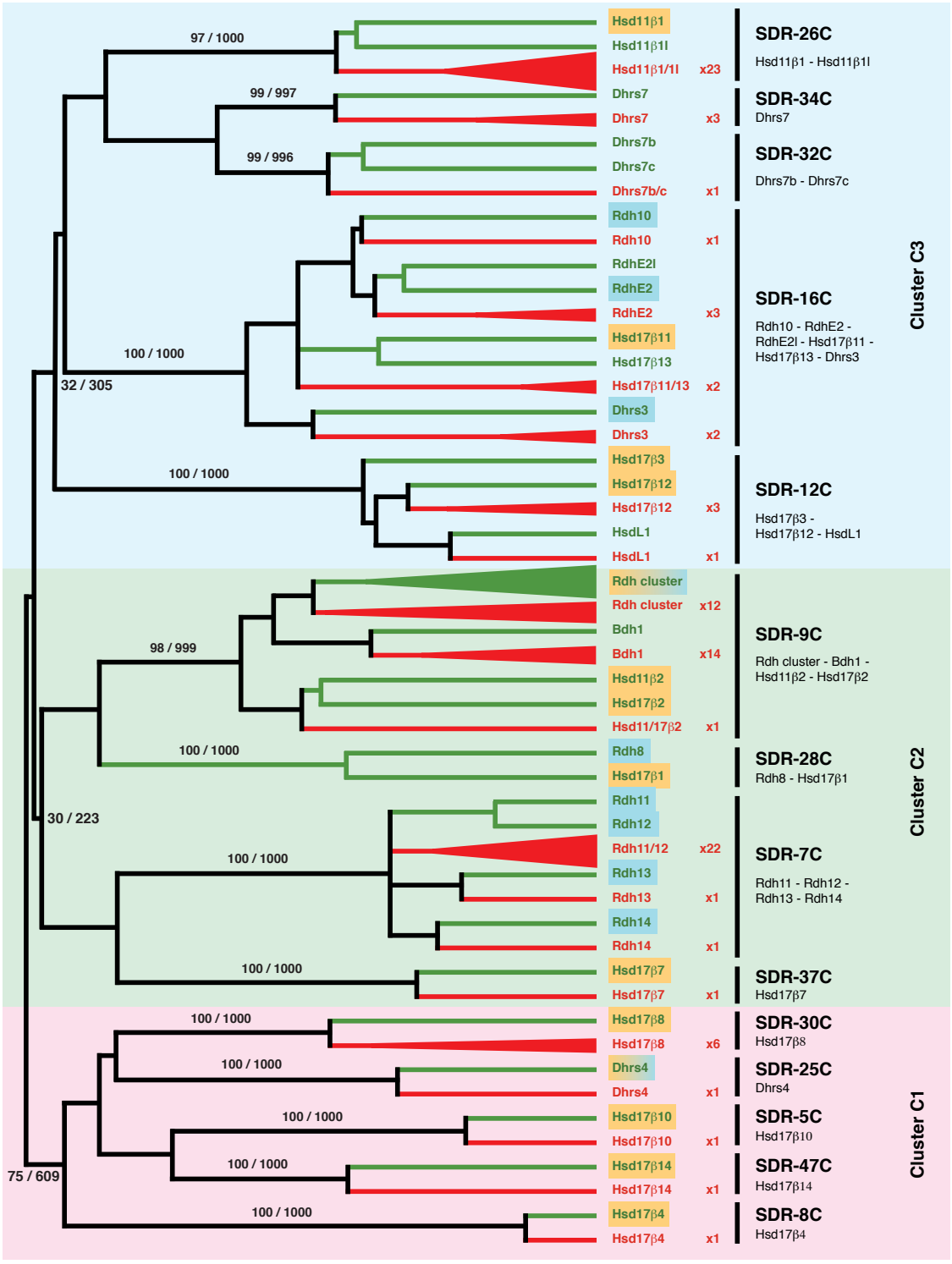


Fig. 2. Schematic phylogeny of members of the SDR superfamily with activity against retinoids and/or steroids. The nomenclature of the three major SDR clusters (C1, C2 and C3) as well as of the individual SDR families is based on the recent reclassification of the SDR superfamily (Bray, Marsden, and Oppermann 2009; Kallberg, Oppermann, and Persson 2010). Bootstrap values (Neighbor Joining/Maximum Likelihood) supporting the three major SDR clusters (C1, C2 and C3) and each of the 14 SDR families are indicated. Retinoid (blue) and steroid (orange) activities of the vertebrate enzymes are indicated, illustrating that both enzymatic activities are commonly found within the SDR superfamily, sometimes even within a single SDR family. Vertebrate branches are in green, while amphioxus branches are in red. The number of amphioxus sequences in each branch is indicated.

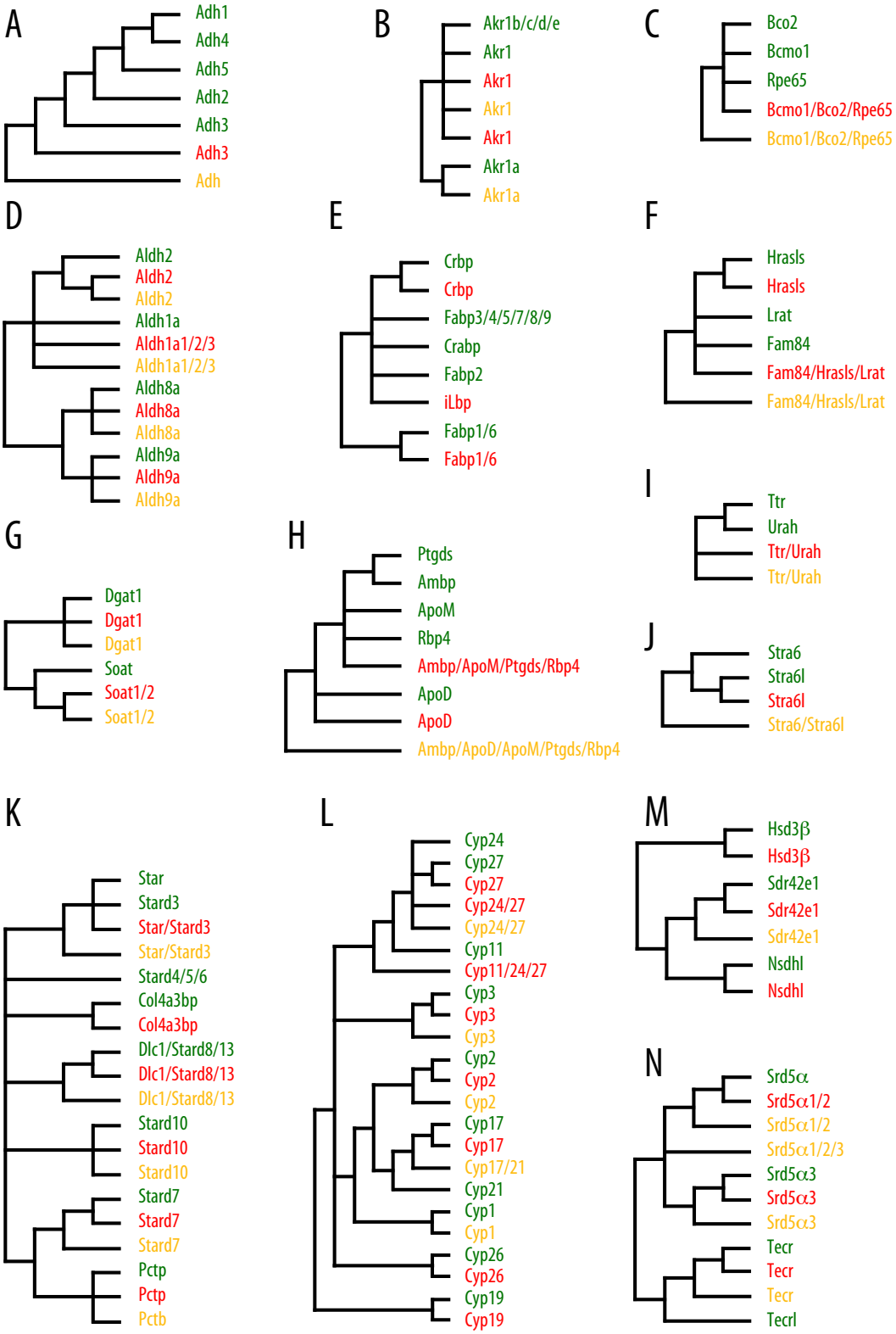


Fig. 3. Diagrammatic tree topologies of the (*A*) Adh, (*B*) Akr1, (*C*) Bco, (*D*) Aldh, (*E*) iLbp, (*F*) NlpC/P60, (*G*) Dgat, (*H*) Lipocalin, (*I*) Ttr/Urah, (*J*) Stra6, (*K*) Star, (*L*) Cyp, (*M*) Hsd3β and (*N*) Srd5α families. Simplified phylogenies of the main components of retinoid and steroid metabolism are shown. Vertebrate branches are in green, amphioxus branches are in red and branches containing sequences from the cnidarian *Nematostella vectensis* are in yellow.

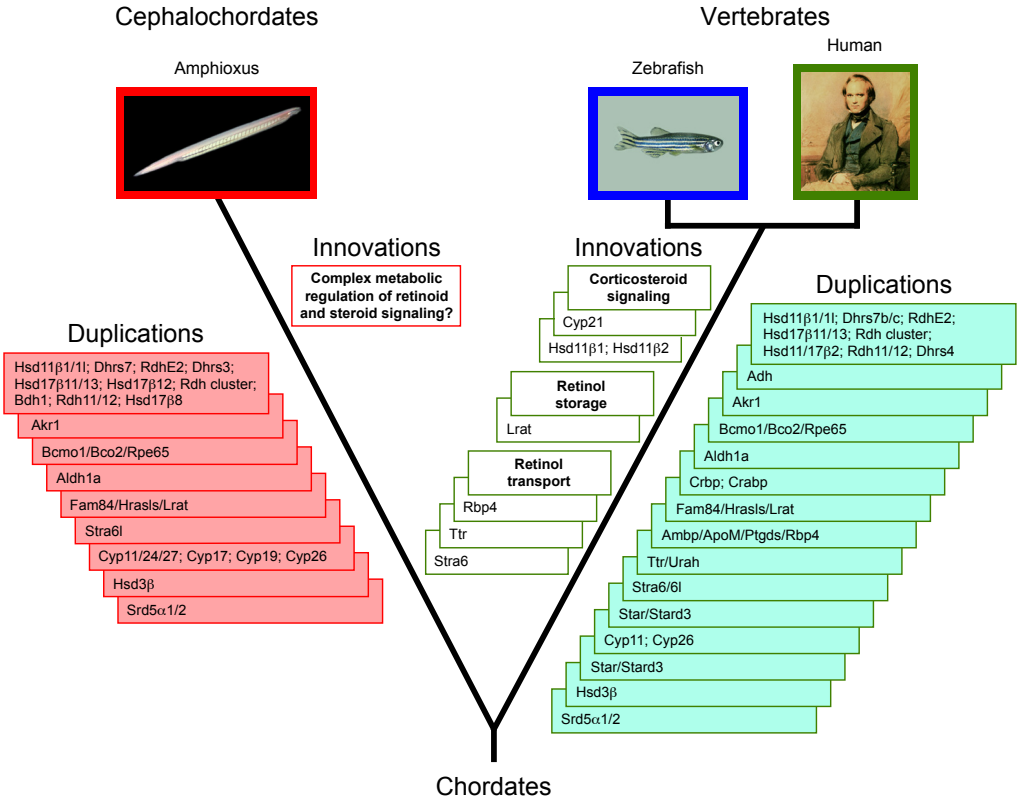


Fig. 4. The evolutionary diversification of the retinoid and steroid genetic machineries in chordates. Gene duplications and functional innovations are shown for both the cephalochordate and the vertebrate lineages. The names of duplicated genes and gene families are indicated for both cephalochordates and vertebrates and several functional innovations resulting from the evolution of novel protein functions are proposed. The "?" highlights that additional studies assessing the biochemical and functional properties of the cephalochordate retinoid and steroid machineries are required to provide experimental support for the hypotheses derived from our evolutionary analyses.

References

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

Albalat R, 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.

Albiston AL, Obeyesekere VR, Smith RE, Krozowski ZS. 1994. Cloning and tissue distribution of the human 11 β-hydroxysteroid dehydrogenase type 2 enzyme. Mol. Cell. Endocrinol. 105:R11-R17.

Anantharaman V, Aravind L. 2003. Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. Genome Biol. 4:R11.

Baker ME. 2004a. Co-evolution of steroidogenic and steroid-inactivating enzymes and adrenal and sex steroid receptors. Mol. Cell. Endocrinol. 215:55-62.

Baker ME. 2004b. Evolutionary analysis of 11β-hydroxysteroid dehydrogenase-type 1, -type 2, -type 3 and 17β-hydroxysteroid dehydrogenase-type 2 in fish. FEBS Lett. 574:167-170.

Baker ME. 2010a. 11β-hydroxysteroid dehydrogenase-type 2 evolved from an ancestral 17β-hydroxysteroid dehydrogenase-type 2. Biochem. Biophys. Res. Comm. 399:215-220.

Baker ME. 2010b. Evolution of 11β-hydroxysteroid dehydrogenase-type 1 and 11β-hydroxysteroid dehydrogenase-type 3. FEBS Lett. 584:2279-2284.

Baker ME, Chang DJ. 2009. 3D model of amphioxus steroid receptor complexed with estradiol. Biochem. Biophys. Res. Commun. 386:516-520.

Batten ML, Imanishi Y, Maeda T, Tu DC, Moise AR, Bronson D, Possin D, Van Gelder RN, Baehr W, Palczewski K. 2004. Lecithin-retinol acyltransferase is essential for accumulation of all-*trans*-retinyl esters in the eye and in the liver. J. Biol. Chem. 279:10422-10432.

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. 2002. Human pancreas protein 2 (PAN2) has a retinal reductase activity and is ubiquitously expressed in human tissues. FEBS Lett. 531:489-493.

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, Korkina OV, Stetsenko AV, Kedishvili NY. 2008. Human retinol dehydrogenase 13 (RDH13) is a mitochondrial short-chain dehydrogenase/reductase with a retinaldehyde reductase activity. FEBS J. 275:138-147.

Belyaeva OV, Korkina OV, Stetsenko AV, Kim T, Nelson PS, Kedishvili NY. 2005. Biochemical properties of purified human retinol dehydrogenase 12 (RDH12): catalytic efficiency toward retinoids and C9 aldehydes and effects of cellular retinol-binding protein type I (CRBPI) and cellular retinaldehyde-binding protein (CRALBP) on the oxidation and reduction of retinoids. Biochemistry 44:7035-7047.

Bertrand S, Brunet FG, Escriva H, Parmentier G, Laudet V, Robinson-Rechavi M. 2004. Evolutionary genomics of nuclear receptors: from twenty-five ancestral genes to derived endocrine systems. Mol. Biol. Evol. 21:1923-1937.

Biswas MG, Russell DW. 1997. Expression cloning and characterization of oxidative 17β- and 3α-hydroxysteroid dehydrogenases from rat and human prostate. J. Biol. Chem. 272:15959-15966.

Bray JE, Marsden BD, Oppermann U. 2009. The human short-chain dehydrogenase/reductase (SDR) superfamily: a bioinformatics summary. Chem. Biol. Interact. 178:99-109.

Bridgham JT, Brown JE, Rodríguez-Marí A, Catchen JM, Thornton JW. 2008. Evolution of a new function by degenerative mutation in cephalochordate steroid receptors. PLoS Genet. 4:e1000191.

Callard GV, Pudney JA, Kendall SL, Reinboth R. 1984. *In vitro* conversion of androgen to estrogen in amphioxus gonadal tissues. Gen. Comp. Endocrinol. 56:53-58.

Callard GV, Tarrant AM, Novillo A, et al. (12 co-authors). 2011. Evolutionary origins of the estrogen signaling system: insights from amphioxus. J. Steroid Biochem. Mol. Biol., in press.

Campo-Paysaa F, Marlétaz F, Laudet V, Schubert M. 2008. Retinoic acid signaling in development: tissue-specific functions and evolutionary origins. Genesis 46:640-656.

Cañestro C, Albalat R, Escriva H, Gonzàlez-Duarte R. 2001. Endogenous β-galactosidase activity in amphioxus: a useful histochemical marker for the digestive system. Dev. Genes Evol. 211:154-156.

Cañestro C, Albalat R, Hjelmqvist L, Godoy L, Jörnvall H, Gonzàlez-Duarte R. 2002. Ascidian and amphioxus *Adh* genes correlate functional and molecular features of the ADH family expansion during vertebrate evolution. J. Mol. Evol. 54:81-89.

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. Zoolog. Sci. 27:128-133.

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, Gonzàlez-Duarte R, Albalat R. 2006. Is retinoic acid genetic machinery a chordate innovation? Evol. Dev. 8:394-406.

Chang C, Liu Y, Zhu H. 1985. Steroid hormones and their functional regulation in amphioxus. In Lofts B, Holmes WN, editors. Current Trends in Comparative Endocrinology. Hong Kong University Press, Hong Kong. p. 205-207.

Chetyrkin SV, Belyaeva OV, Gough WH, Kedishvili NY. 2001a. Characterization of a novel type of human microsomal 3α-hydroxysteroid dehydrogenase: Unique tissue distribution and catalytic properties. J. Biol. Chem. 276:22278-22286.

Chetyrkin SV, Hu J, Gough WH, Dumaual N, Kedishvili NY. 2001b. Further characterization of human microsomal 3α-hydroxysteroid dehydrogenase. Arch. Biochem. Biophys. 386:1-10.

Clark BJ, Wells J, King SR, Stocco DM. 1994. The purification, cloning, and expression of a novel luteinizing hormone-induced mitochondrial protein in MA-10 mouse Leydig tumor cells. Characterization of the steroidogenic acute regulatory protein (StAR). J. Biol. Chem. 269:28314-28322.

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.

Duan WR, Linzer DI, Gibori G. 1996. Cloning and characterization of an ovarian-specific protein that associates with the short form of the prolactin receptor. J. Biol. Chem. 271:15602-15607.

Dufort I, Rheault P, Huang XF, Soucy P, Luu-The V. 1999. Characteristics of a highly labile human type 5 17β-hydroxysteroid dehydrogenase. Endocrinology 140:568-574.

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

Eisen JA. 1998. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. Genome Res. 8:163-167.

Escriva H, Bertrand S, Laudet V. 2004. The evolution of the nuclear receptor superfamily. Essays Biochem. 40:11-26.

Escriva H, Delaunay F, Laudet V. 2000. Ligand binding and nuclear receptor evolution. Bioessays 22:717-727.

Escriva H, Holland ND, Gronemeyer H, Laudet V, Holland LZ. 2002. The retinoic acid signaling pathway regulates anterior/posterior patterning in the nerve cord and pharynx of amphioxus, a chordate lacking neural crest. Development 129:2905-2916.

Escriva H, Safi R, Hanni C, Langlois MC, Saumitou-Laprade P, Stehelin D, Capron A, Pierce R, Laudet V. 1997. Ligand binding was acquired during evolution of nuclear receptors. Proc. Natl. Acad. Sci. U. S. A. 94:6803-6808.

Estes KA, Kalamegham R, Hanna-Rose W. 2007. Membrane localization of the NlpC/P60 family protein EGL-26 correlates with regulation of vulval cell morphogenesis in *Caenorhabditis elegans*. Dev. Biol. 308:196-205.

Fang YQ, Weng YZ, Hu XX. 2001. Distribution of sex steroid hormones and their receptors in the gonads and nervous system of amphioxus (*Branchiostoma belcherii)*. Current Zoology 47:398-403.

Flicek P, Aken BL, Ballester B, et al. (57 co-authors). 2010. Ensembl's 10th year. Nucleic Acids Res. 38:D557-D562.

Flower DR. 1996. The lipocalin protein family: structure and function. Biochem. J. 318:1-14.

Folli C, Ramazzina I, Percudani R, Berni R. 2005. Ligand-binding specificity of an invertebrate (*Manduca sexta*) putative cellular retinoic acid binding protein. Biochim. Biophys. Acta 1747:229-237.

Fomitcheva J, Baker ME, Anderson E, Lee GY, Aziz N. 1998. Characterization of Ke 6, a new 17β-hydroxysteroid dehydrogenase, and its expression in gonadal tissues. J. Biol. Chem. 273:22664-22671.

Funder JW, Pearce PT, Smith R, Smith AI. 1988. Mineralocorticoid action: target tissue specificity is enzyme, not receptor, mediated. Science 242:583-585.

Gabaldón T, Huynen MA. 2004. Prediction of protein function and pathways in the genome era. Cell. Mol. Life Sci. 61:930-944.

Galtier N, Gouy M, Gautier C. 1996. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. 12:543-548.

Ganfornina MD, Gutiérrez G, Bastiani M, Sánchez D. 2000. A phylogenetic analysis of the lipocalin protein family. Mol. Biol. Evol. 17:114-126.

Geissler WM, Davis DL, Wu L, Bradshaw KD, Patel S, Mendonca BB, Elliston KO, Wilson JD, Russell DW, Andersson S. 1994. Male pseudohermaphroditism caused by mutations of testicular 17 β-hydroxysteroid dehydrogenase 3. Nat. Genet. 7:34-39.

Godoy L, Gonzàlez-Duarte R, Albalat R. 2006. S-Nitrosogluthathione reductase activity of amphioxus ADH3: insights into the nitric oxide metabolism. Int. J. Biol. Sci. 2:117-124.

Gonzàlez-Duarte R, Albalat R. 2005. Merging protein, gene and genomic data: the evolution of the MDR-ADH family. Heredity 95:184-197.

Gottesman ME, Quadro L, Blaner WS. 2001. Studies of vitamin A metabolism in mouse model systems. Bioessays 23:409-419.

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 3α-hydroxysteroids. J. Biol. Chem. 273:19778-19785.

Gu PL, Gunawardene YI, Chow BC, He JG, Chan SM. 2002. Characterization of a novel cellular retinoic acid/retinol binding protein from shrimp: expression of the recombinant protein for immunohistochemical detection and binding assay. Gene 288:77-84.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52:696-704.

Gutiérrez G, Ganfornina MD, Sánchez D. 2000. Evolution of the lipocalin family as inferred from a protein sequence phylogeny. Biochim. Biophys. Acta 1482:35-45.

Haeseleer F, Jang GF, Imanishi Y, Driessen CA, Matsumura M, Nelson PS, Palczewski K. 2002. Dual-substrate specificity short chain retinol dehydrogenases from the vertebrate retina. J. Biol. Chem. 277:45537-45546.

Haeseleer F, Palczewski K. 2000. Short-chain dehydrogenases/reductases in retina. Methods Enzymol. 316:372-383.

Holland LZ, Albalat R, Azumi K, et al. (64 co-authors). 2008. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. Genome Res. 18:1100-1111.

Holland LZ, Holland ND. 1996. Expression of *AmphiHox-1* and *AmphiPax-1* in amphioxus embryos treated with retinoic acid: insights into evolution and patterning of the chordate nerve cord and pharynx. Development 122:1829-1838.

Huang XF, Luu-The V. 2001. Characterization of the oxidative 3α-hydroxysteroid dehydrogenase activity of human recombinant 11-*cis*-retinol dehydrogenase. Biochim. Biophys. Acta 1547:351-358.

Huang Y, Tang R, Dai J, et al. (11 co-authors). 2001. A novel human hydroxysteroid dehydrogenase like 1 gene (HSDL1) is highly expressed in reproductive tissues. Mol. Biol. Rep. 28:185-191.

Huyghe S, Mannaerts GP, Baes M, Van Veldhoven PP. 2006. Peroxisomal multifunctional protein-2: the enzyme, the patients and the knockout mouse model. Biochim. Biophys. Acta 1761:973-994.

Jahng WJ, Xue L, Rando RR. 2003. Lecithin retinol acyltransferase is a founder member of a novel family of enzymes. Biochemistry 42:12805-12812.

Jez JM, Bennett MJ, Schlegel BP, Lewis M, Penning TM. 1997. Comparative anatomy of the aldo-keto reductase superfamily. Biochem. J. 326:625-636.

Jin M, Li S, Moghrabi WN, Sun H, Travis GH. 2005. Rpe65 is the retinoid isomerase in bovine retinal pigment epithelium. Cell 122:449-459.

Jin XH, Okamoto Y, Morishita J, Tsuboi K, Tonai T, Ueda N. 2007. Discovery and characterization of a $Ca2^{+}$-independent phosphatidylethanolamine N-acyltransferase generating the anandamide precursor and its congeners. J. Biol. Chem. 282:3614-3623.

Kallberg Y, Oppermann U, Persson B. 2010. Classification of the short-chain dehydrogenase/reductase superfamily using hidden Markov models. FEBS J. 277:2375-2386.

Katsu Y, Kubokawa K, Urushitani H, Iguchi T. 2010. Estrogen-dependent transactivation of amphioxus steroid hormone receptor via both estrogen and androgen response elements. Endocrinology 151:639-648.

Kawaguchi R, Yu J, Honda J, Hu J, Whitelegge J, Ping P, Wiita P, Bok D, Sun H. 2007. A membrane receptor for retinol binding protein mediates cellular uptake of vitamin A. Science 315:820-825.

Kedishvili NY, Chumakova OV, Chetyrkin SV, Belyaeva OV, Lapshina EA, Lin DW, Matsumura M, Nelson PS. 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.

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.

Kim TS, Maeda A, Maeda T, Heinlein C, Kedishvili N, Palczewski K, Nelson PS. 2005. Delayed dark adaptation in 11-*cis*-retinol dehydrogenase-deficient mice: a role of RDH11 in visual processes *in vivo*. J. Biol. Chem. 280:8694-8704.

Koop D, Holland ND, Sémon 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.

Kowalik D, Haller F, Adamski J, Moeller G. 2009. In search for function of two human orphan SDR enzymes: hydroxysteroid dehydrogenase like 2 (HSDL2) and short-chain dehydrogenase/reductase-orphan (SDR-O). J. Steroid Biochem. Mol. Biol. 117:117-124.

Kusakabe M, Nakamura I, Young G. 2003. 11β-hydroxysteroid dehydrogenase complementary deoxyribonucleic acid in rainbow trout: cloning, sites of expression, and seasonal changes in gonads. Endocrinology 144:2534-2545.

Kuzniar A, van Ham RC, Pongor S, Leunissen JA. 2008. The quest for orthologs: finding the corresponding gene across genomes. Trends Genet. 24:539-551.
Larkin MA, Blackshields G, Brown NP, et al. (13 co-authors). 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947-2948.
Lee SA, Belyaeva OV, Popov IK, Kedishvili NY. 2007. Overproduction of bioactive retinoic acid in cells expressing disease-associated mutants of retinol dehydrogenase 12. J. Biol. Chem. 282:35621-35628.
Lei Z, Chen W, Zhang M, Napoli JL. 2003. Reduction of all-*trans*-retinal in the mouse liver peroxisome fraction by the short-chain dehydrogenase/reductase RRD: induction by the PPARα ligand clofibrate. Biochemistry 42:4190-4196.
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 B, White JT, Ferguson C, Wang S, Vessella R, Bumgarner R, True LD, Hood L, Nelson PS. 2001. Prostate short-chain dehydrogenase reductase 1 (PSDR1): a new member of the short-chain steroid dehydrogenase/reductase family highly expressed in normal and neoplastic prostate epithelium. Cancer Res. 61:1611-1618.
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.
Lukacik P, Keller B, Bunkoczi G, Kavanagh K, Hwa Lee W, Adamski J, Oppermann U. 2007. Structural and biochemical characterization of human orphan DHRS10 reveals a novel cytosolic enzyme with steroid dehydrogenase activity. Biochem. J. 402:419-427.
Luu-The V, Tremblay P, Labrie F. 2006. Characterization of type 12 17β-hydroxysteroid dehydrogenase, an isoform of type 3 17β-hydroxysteroid dehydrogenase responsible for estradiol formation in women. Mol. Endocrinol. 20:437-443.
Maeda A, Maeda T, Imanishi Y, et al. (13 co-authors). 2005. Role of photoreceptor-specific retinol dehydrogenase in the retinoid cycle *in vivo*. J. Biol. Chem. 280:18822-18832.
Maeda A, Maeda T, Sun W, Zhang H, Baehr W, Palczewski K. 2007. Redundant and unique roles of retinol dehydrogenases in the mouse retina. Proc. Natl. Acad. Sci. U. S. A. 104:19565-19570.
Mangelsdorf DJ, Thummel C, Beato M, et al. (11 co-authors). 1995. The nuclear receptor superfamily: the second decade. Cell 83:835-839.
Mansfield SG, Cammer S, Alexander SC, Muehleisen DP, Gray RS, Tropsha A, Bollenbacher WE. 1998. Molecular cloning and characterization of an invertebrate cellular retinoic acid binding protein. Proc. Natl. Acad. Sci. U. S. A. 95:6825-6830.
Marijanovic Z, Laubner D, Moller G, Gege C, Husen B, Adamski J, Breitling R. 2003. Closing the gap: identification of human 3-ketosteroid reductase, the last unknown enzyme of mammalian cholesterol biosynthesis. Mol. Endocrinol. 17:1715-1725.

Markov GV, Paris M, Bertrand S, Laudet V. 2008. The evolution of the ligand/receptor couple: a long road from comparative endocrinology to comparative genomics. Mol. Cell. Endocrinol. 293:5-16.
Markov GV, Tavares R, Dauphin-Villemant C, Demeneix BA, Baker ME, Laudet V. 2009. Independent elaboration of steroid hormone signaling pathways in metazoans. Proc. Natl. Acad. Sci. U. S. A. 106:11913-11918.
Marks AR, McIntyre JO, Duncan TM, Erdjument-Bromage H, Tempst P, Fleischer S. 1992. Molecular cloning and characterization of (R)-3-hydroxybutyrate dehydrogenase from human heart. J. Biol. Chem. 267:15459-15463.
Marlétaz F, Holland LZ, Laudet V, Schubert M. 2006. Retinoic acid signaling and the evolution of chordates. Int. J. Biol. Sci. 2:38-47.
Matsunaga T, Endo S, Maeda S, Ishikura S, Tajima K, Tanaka N, Nakamura KT, Imamura Y, Hara A. 2008. Characterization of human DHRS4: an inducible short-chain dehydrogenase/reductase enzyme with 3β-hydroxysteroid dehydrogenase activity. Arch. Biochem. Biophys. 477:339-347.
Mindnich R, Adamski J. 2009. Zebrafish 17β-hydroxysteroid dehydrogenases: An evolutionary perspective. Mol. Cell. Endocrinol. 301:20-26.
Mindnich R, Moller G, Adamski J. 2004. The role of 17β-hydroxysteroid dehydrogenases. Mol. Cell. Endocrinol. 218:7-20.
Mizuta T, Asahina K, Suzuki M, Kubokawa K. 2008. *In vitro* conversion of sex steroids and expression of sex steroidogenic enzyme genes in amphioxus ovary. J. Exp. Zool. A Ecol. Genet. Physiol. 309:83-93.
Mizuta T, Kubokawa K. 2007. Presence of sex steroids and cytochrome P450 (CYP) genes in amphioxus. Endocrinology 148:3554-3565.
Mondal MS, Ruiz A, Bok D, Rando RR. 2000. Lecithin retinol acyltransferase contains cysteine residues essential for catalysis. Biochemistry 39:5215-5220.
Moon YA, Horton JD. 2003. Identification of two mammalian reductases involved in the two-carbon fatty acyl elongation cascade. J. Biol. Chem. 278:7335-7343.
Nazarenko I, Schafer R, Sers C. 2007. Mechanisms of the HRSL3 tumor suppressor function in ovarian carcinoma cells. J. Cell Sci. 120:1393-1404.
Nelson DR. 2009. The cytochrome p450 homepage. Hum. Genomics 4:59-65.
Niederreither K, Abu-Abed S, Schuhbaur B, Petkovich M, Chambon P, Dollé P. 2002. Genetic evidence that oxidative derivatives of retinoic acid are not involved in retinoid signaling during mouse development. Nat. Genet. 31:84-88.
Noy N. 2000. Retinoid-binding proteins: mediators of retinoid action. Biochem. J. 348:481-495.
O'Byrne SM, Wongsiriroj N, Libien J, Vogel S, Goldberg IJ, Baehr W, Palczewski K, Blaner WS. 2005. Retinoid absorption and storage is impaired in mice lacking lecithin:retinol acyltransferase (LRAT). J. Biol. Chem. 280:35647-35657.
Oppermann UC, Persson B, Jörnvall H. 1997. Function, gene organization and protein structures of 11β-hydroxysteroid dehydrogenase isoforms. Eur. J. Biochem. 249:355-360.
Orland MD, Anwar K, Cromley D, Chu CH, Chen L, Billheimer JT, Hussain MM, Cheng D. 2005. Acyl coenzyme A dependent retinol esterification by acyl coenzyme A: diacylglycerol acyltransferase 1. Biochim. Biophys. Acta 1737:76-82.
Paris M, Pettersson K, Schubert M, Bertrand S, Pongratz I, Escriva H, Laudet V. 2008. An amphioxus orthologue of the estrogen receptor that does not bind estradiol: insights into estrogen receptor evolution. BMC Evol. Biol. 8:219.

Payne AH, Hales DB. 2004. Overview of steroidogenic enzymes in the pathway from cholesterol to active steroid hormones. Endocr. Rev. 25:947-970.

Peltoketo H, Isomaa V, Maentausta O, Vihko R. 1988. Complete amino acid sequence of human placental 17 β-hydroxysteroid dehydrogenase deduced from cDNA. FEBS Lett. 239:73-77.

Peltoketo H, Luu-The V, Simard J, Adamski J. 1999. 17β-hydroxysteroid dehydrogenase (HSD)/17-ketosteroid reductase (KSR) family; nomenclature and main characteristics of the 17HSD/KSR enzymes. J. Mol. Endocrinol. 23:1-11.

Penning TM. 1997. Molecular endocrinology of hydroxysteroid dehydrogenases. Endocr. Rev. 18:281-305.

Penning TM. 2003. Hydroxysteroid dehydrogenases and pre-receptor regulation of steroid hormone action. Hum. Reprod. Update 9:193-205.

Penning TM, Burczynski ME, Jez JM, Hung CF, Lin HK, Ma H, Moore M, Palackal N, Ratnam K. 2000. Human 3α-hydroxysteroid dehydrogenase isoforms (AKR1C1-AKR1C4) of the aldo-keto reductase superfamily: functional plasticity and tissue distribution reveals roles in the inactivation and formation of male and female sex hormones. Biochem. J. 351:67-77.

Perrière G, Gouy M. 1996. WWW-Query: An on-line retrieval system for biological sequence banks. Biochimie 78:364-369.

Putnam NH, Butts T, Ferrier DE, et al. (37 co-authors). 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature 453:1064-1071.

Rattner A, Smallwood PM, Nathans J. 2000. Identification and characterization of all-*trans*-retinol dehydrogenase from photoreceptor outer segments, the visual cycle enzyme that reduces all-*trans*-retinal to all-*trans*-retinol. J. Biol. Chem. 275:11034-11043.

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.

Ross AC. 1982. Retinol esterification by rat liver microsomes. Evidence for a fatty acyl coenzyme A: retinol acyltransferase. J. Biol. Chem. 257:2453-2459.

Ross SA, McCaffery PJ, Drager UC, De Luca LM. 2000. Retinoids in embryonal development. Physiol. Rev. 80:1021-1054.

Ruiz FX, Gallego O, Ardèvol A, et al. (12 co-authors). 2009. Aldo-keto reductases from the AKR1B subfamily: retinoid specificity and control of cellular retinoic acid levels. Chem. Biol. Interact. 178:171-177.

Russell DW, Wilson JD. 1994. Steroid 5α-reductase: two genes/two enzymes. Annu. Rev. Biochem. 63:25-61.

Sánchez D, Ganfornina MD, Gutiérrez G, Marín A. 2003. Exon-intron structure and evolution of the Lipocalin gene family. Mol. Biol. Evol. 20:775-783.

Schaap FG, van der Vusse GJ, Glatz JF. 2002. Evolution of the family of intracellular lipid binding proteins in vertebrates. Mol. Cell. Biochem. 239:69-77.

Schubert M, Brunet F, Paris M, Bertrand S, Benoit G, Laudet V. 2008. Nuclear hormone receptor signaling in amphioxus. Dev. Genes Evol. 218:651-665.

Schubert M, Escriva H, Xavier-Neto J, Laudet V. 2006a. Amphioxus and tunicates as evolutionary model systems. Trends Ecol. Evol. 21:269-277.

Schubert M, Holland ND, Escriva H, Holland LZ, Laudet V. 2004. Retinoic acid influences anteroposterior positioning of epidermal sensory neurons and their gene expression in a developing chordate (amphioxus). Proc. Natl. Acad. Sci. U. S. A. 101:10320-10325.

Schubert M, Holland ND, Laudet V, Holland LZ. 2006b. A retinoic acid-*Hox* hierarchy controls both anterior/posterior patterning and neuronal specification in the developing central nervous system of the cephalochordate amphioxus. Dev. Biol. 296:190-202.

Schubert M, Yu JK, Holland ND, Escriva H, Laudet V, Holland LZ. 2005. Retinoic acid signaling acts via Hox1 to establish the posterior limit of the pharynx in the chordate amphioxus. Development 132:61-73.

Shafqat N, Marschall HU, Filling C, et al. (11 co-authors). 2003. Expanded substrate screenings of human and *Drosophila* type 10 17β-hydroxysteroid dehydrogenases (HSDs) reveal multiple specificities in bile acid and steroid hormone metabolism: characterization of multifunctional 3α/7α/7β/17β/20β/21-HSD. Biochem. J. 376:49-60.

Shehu A, Mao J, Gibori GB, Halperin J, Le J, Devi YS, Merrill B, Kiyokawa H, Gibori G. 2008. Prolactin receptor-associated protein/17β-hydroxysteroid dehydrogenase type 7 gene (Hsd17β7) plays a crucial role in embryonic development and fetal survival. Mol. Endocrinol. 22:2268-2277.

Simões-Costa MS, Azambuja AP, Xavier-Neto J. 2008. The search for non-chordate retinoic acid signaling: lessons from chordates. J. Exp. Zoolog. B Mol. Dev. Evol. 310:54-72.

Söderhäll I, Tangprasittipap A, Liu H, Sritunyalucksana K, Prasertsan P, Jiravanichpaisal P, Söderhäll K. 2006. Characterization of a hemocyte intracellular fatty acid-binding protein from crayfish (*Pacifastacus leniusculus*) and shrimp (*Penaeus monodon*). FEBS J. 273:2902-2912.

Soref CM, Di YP, Hayden L, Zhao YH, Satre MA, Wu R. 2001. Characterization of a novel airway epithelial cell-specific short chain alcohol dehydrogenase/reductase gene whose expression is up-regulated by retinoids and is involved in the metabolism of retinol. J. Biol. Chem. 276:24194-24202.

Stocco DM. 2001. StAR protein and the regulation of steroid hormone biosynthesis. Annu. Rev. Physiol. 63:193-213.

Storbeck KH, Swart P, Swart AC. 2007. Cytochrome P450 side-chain cleavage: insights gained from homology modeling. Mol. Cell. Endocrinol. 265-266:65-70.

Takeda N, Kubokawa K, Matsumoto G. 2003. Immunoreactivity for progesterone in the giant Rohde cells of the amphioxus, *Branchiostoma belcheri*. Gen. Comp. Endocrinol. 132:379-383.

Takimoto N, Kusakabe T, Horie T, Miyamoto Y, Tsuda M. 2006. Origin of the vertebrate visual cycle: III. Distinct distribution of RPE65 and β-carotene 15,15'-monooxygenase homologues in *Ciona intestinalis*. Photochem. Photobiol. 82:1468-1474.

Tarrant AM, Reitzel AM, Blomquist CH, Haller F, Tokarz J, Adamski J. 2009. Steroid metabolism in cnidarians: insights from *Nematostella vectensis*. Mol. Cell. Endocrinol. 301:27-36.

Tatusov RL, Koonin EV, Lipman D. 1997. A genomic perspective on protein families. Science 278:631-637.

Theodosiou M, Laudet V, Schubert M. 2010. From carrot to clinic: an overview of the retinoic acid signaling pathway. Cell. Mol. Life Sci. 67:1423-1445.

Uemura M, Tamura K, Chung S, Honma S, Okuyama A, Nakamura Y, Nakagawa H. 2008. Novel 5α-steroid reductase (SRD5A3, type-3) is overexpressed in hormone-refractory prostate cancer. Cancer Sci. 99:81-86.

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

Watari H, Arakane F, Moog-Lutz C, Kallen CB, Tomasetto C, Gerton GL, Rio MC, Baker ME, Strauss JF III. 1997. MLN64 contains a domain with homology to the steroidogenic acute regulatory protein (StAR) that stimulates steroidogenesis. Proc. Natl. Acad. Sci. U. S. A. 94:8462-8467.

Wu L, Einstein M, Geissler WM, Chan HK, Elliston KO, Andersson S. 1993. Expression cloning and characterization of human 17β-hydroxysteroid dehydrogenase type 2, a microsomal enzyme possessing 20α-hydroxysteroid dehydrogenase activity. J. Biol. Chem. 268:12964-12969.

Wu X, Lukacik P, Kavanagh KL, Oppermann U. 2007. SDR-type human hydroxysteroid dehydrogenases involved in steroid hormone activation. Mol. Cell. Endocrinol. 265-266:71-76.

Yamamoto H, Simon A, Eriksson U, Harris E, Berson EL, Dryja TP. 1999. Mutations in the gene encoding 11-*cis* retinol dehydrogenase cause delayed dark adaptation and fundus albipunctatus. Nat. Genet. 22:188-191.

Zanotti G, Cendron L, Ramazzina I, Folli C, Percudani R, Berni R. 2006. Structure of zebrafish HIUase: insights into evolution of an enzyme to a hormone transporter. J. Mol. Biol. 363:1-9.

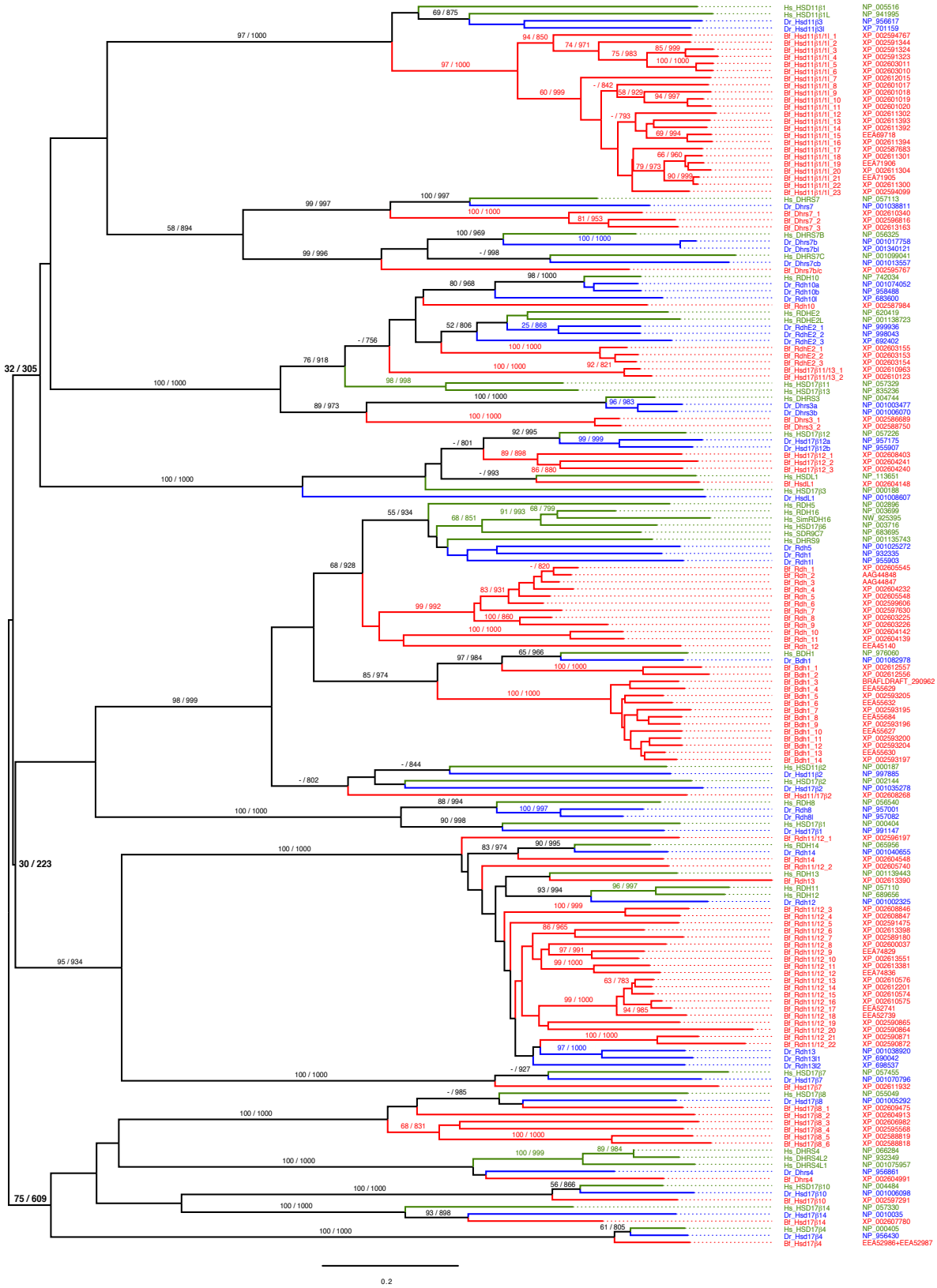


Fig. S1. SDR phylogeny. The topology of the Maximum Likelihood tree is shown. With the exception of the support for the three major SDR clusters (C1, C2 and C3), Bootstrap values (Neighbor Joining/Maximum Likelihood) are only indicated, if the Neighbor Joining value is at least 75 or the Maximum Likelihood value is at least 750. Human sequences are in green, zebrafish sequences are in blue, amphioxus sequences are in red. Accession numbers are given for each sequence.

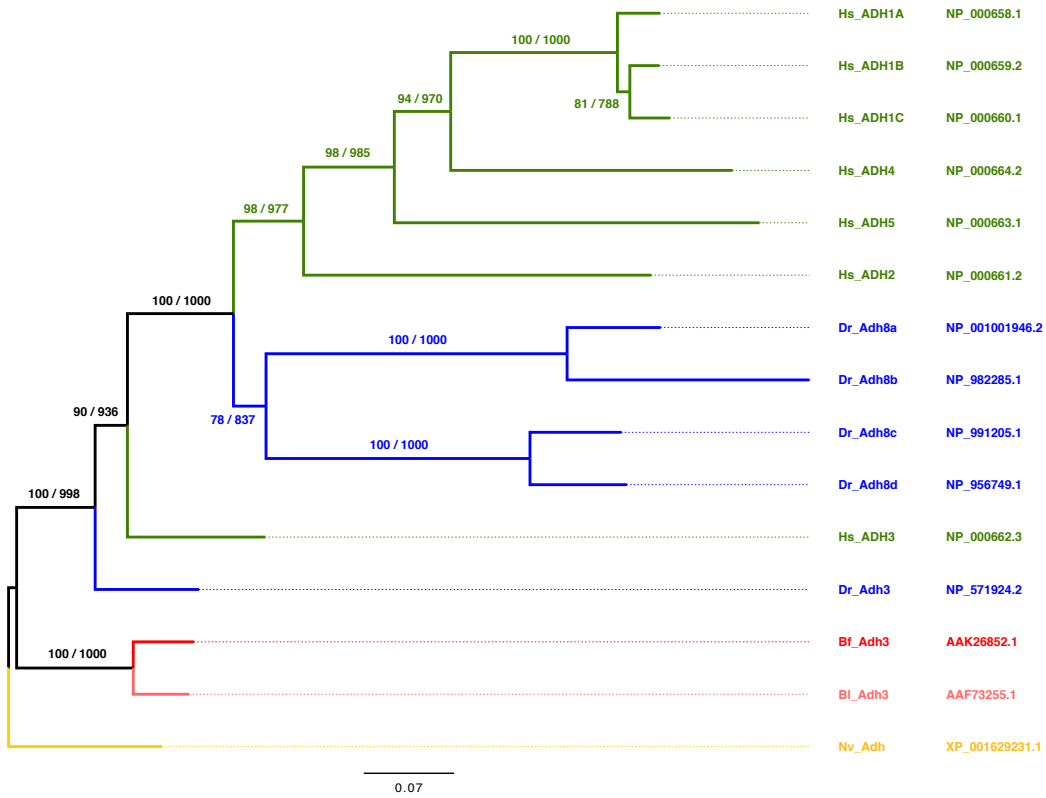


Fig. S2. Adh phylogeny. The topology of the Maximum Likelihood tree is shown. Bootstrap values (Neighbor Joining/Maximum Likelihood) are indicated, if the Neighbor Joining value is at least 75 or the Maximum Likelihood value is at least 750. Human sequences are in green, zebrafish sequences are in blue, amphioxus sequences are in red, sequences from the cnidarian *Nematostella vectensis* are in yellow. Accession numbers are given for each sequence.

Fig. S3. Akr1 phylogeny. The topology of the Maximum Likelihood tree is shown. Bootstrap values (Neighbor Joining/Maximum Likelihood) are indicated, if the Neighbor Joining value is at least 75 or the Maximum Likelihood value is at least 750. Human sequences are in green, zebrafish sequences are in blue, amphioxus sequences are in red, sequences from the cnidarian *Nematostella vectensis* are in yellow. Accession numbers are given for each sequence.

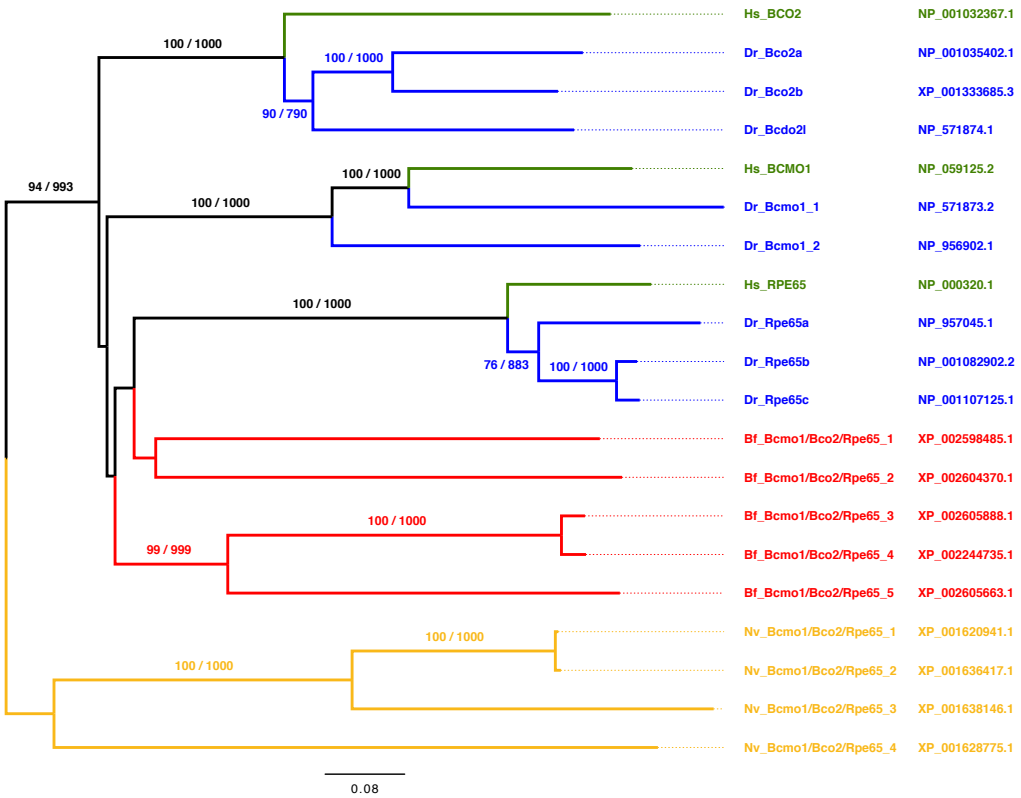


Fig. S4. Bco phylogeny. The topology of the Maximum Likelihood tree is shown. Bootstrap values (Neighbor Joining/Maximum Likelihood) are indicated, if the Neighbor Joining value is at least 75 or the Maximum Likelihood value is at least 750. Human sequences are in green, zebrafish sequences are in blue, amphioxus sequences are in red, sequences from the cnidarian *Nematostella vectensis* are in yellow. Accession numbers are given for each sequence.

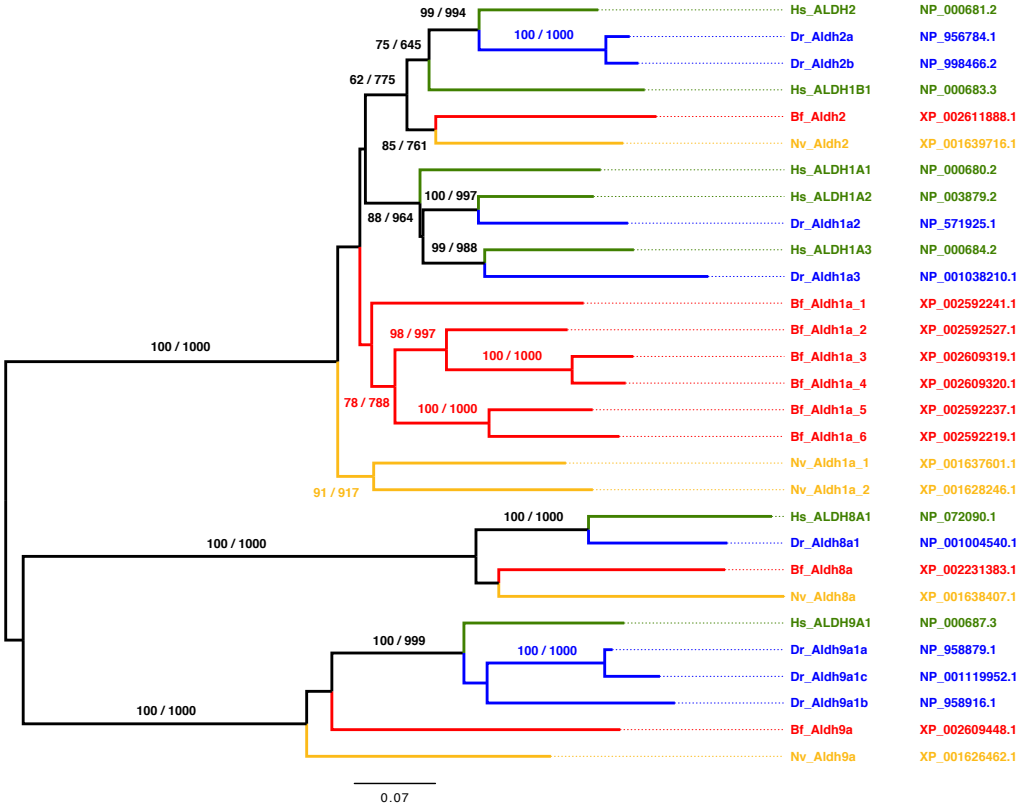


Fig. S5. Aldh phylogeny. The topology of the Maximum Likelihood tree is shown. Bootstrap values (Neighbor Joining/Maximum Likelihood) are indicated, if the Neighbor Joining value is at least 75 or the Maximum Likelihood value is at least 750. Human sequences are in green, zebrafish sequences are in blue, amphioxus sequences are in red, sequences from the cnidarian *Nematostella vectensis* are in yellow. Accession numbers are given for each sequence.

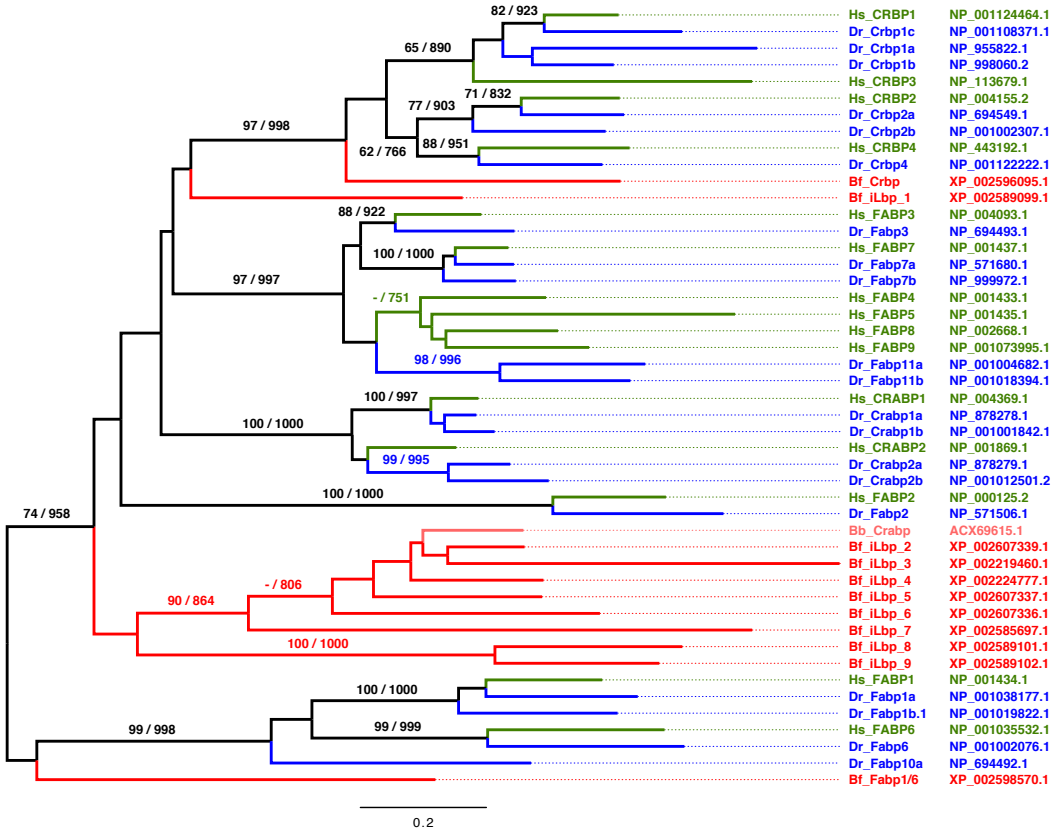


Fig. S6. iLbp phylogeny. The topology of the Maximum Likelihood tree is shown. Bootstrap values (Neighbor Joining/Maximum Likelihood) are indicated, if the Neighbor Joining value is at least 75 or the Maximum Likelihood value is at least 750. Human sequences are in green, zebrafish sequences are in blue, amphioxus sequences are in red. Accession numbers are given for each sequence.

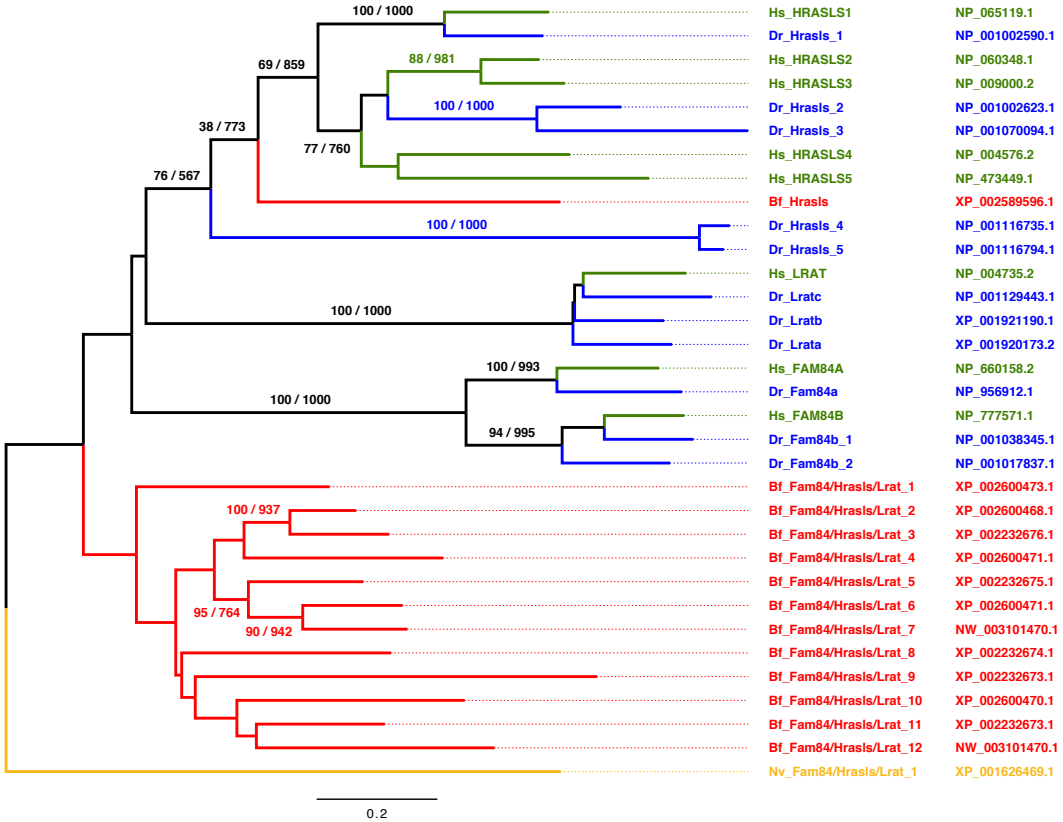


Fig. S7. NlpC/p60 phylogeny. The topology of the Maximum Likelihood tree is shown. Bootstrap values (Neighbor Joining/Maximum Likelihood) are indicated, if the Neighbor Joining value is at least 75 or the Maximum Likelihood value is at least 750. Human sequences are in green, zebrafish sequences are in blue, amphioxus sequences are in red, sequences from the cnidarian *Nematostella vectensis* are in yellow. Accession numbers are given for each sequence.

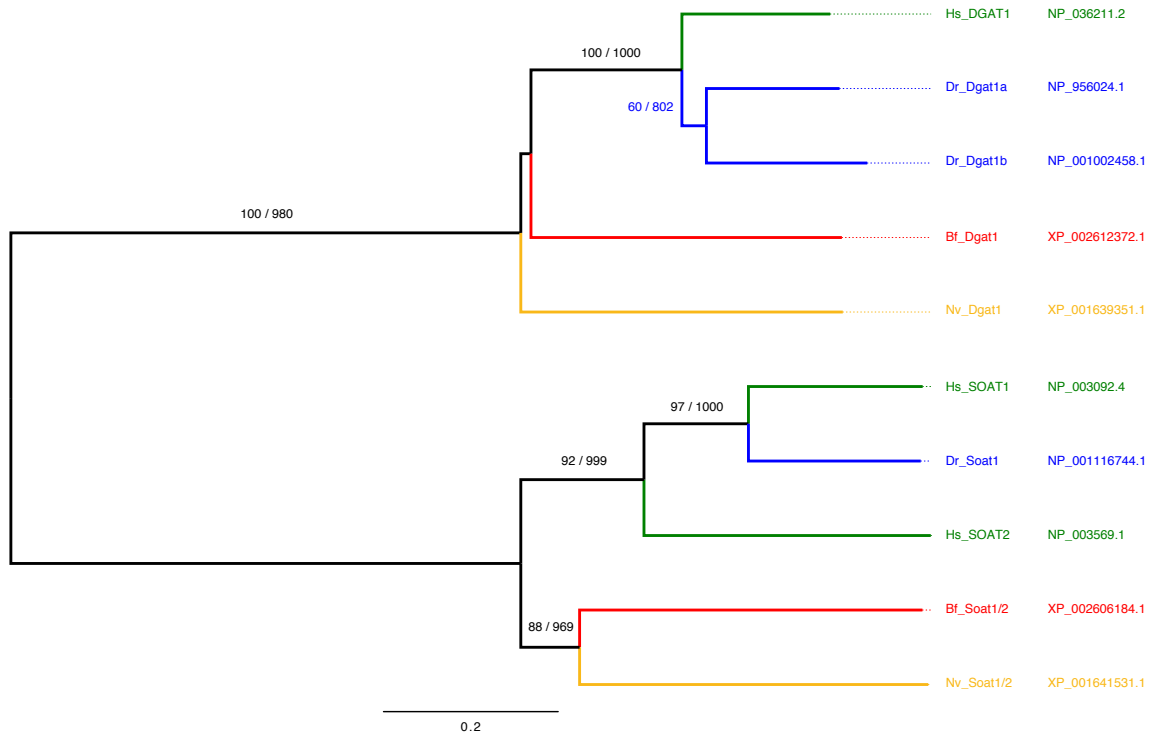


Fig. S8. Dgat phylogeny. The topology of the Maximum Likelihood tree is shown. Bootstrap values (Neighbor Joining/Maximum Likelihood) are indicated, if the Neighbor Joining value is at least 75 or the Maximum Likelihood value is at least 750. Human sequences are in green, zebrafish sequences are in blue, amphioxus sequences are in red, sequences from the cnidarian *Nematostella vectensis* are in yellow. Accession numbers are given for each sequence.

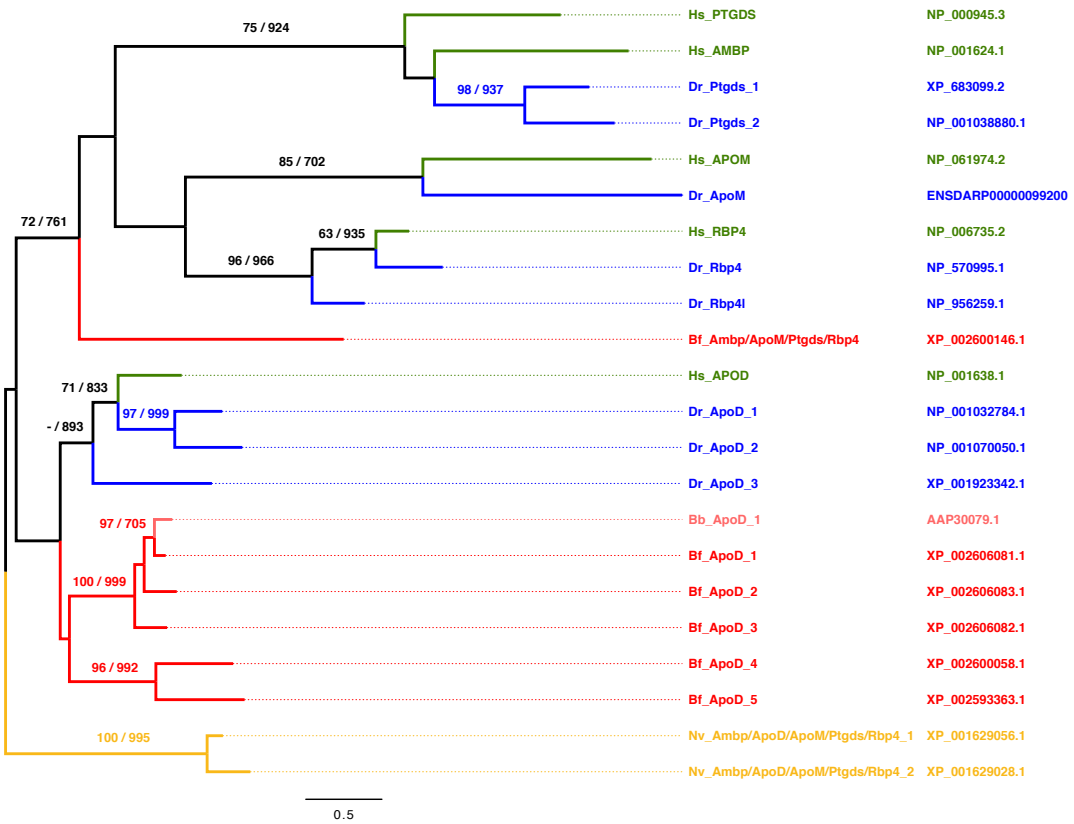


Fig. S9. Lipocalin phylogeny. The topology of the Maximum Likelihood tree is shown. Bootstrap values (Neighbor Joining/Maximum Likelihood) are indicated, if the Neighbor Joining value is at least 75 or the Maximum Likelihood value is at least 750. Human sequences are in green, zebrafish sequences are in blue, amphioxus sequences are in red, sequences from the cnidarian *Nematostella vectensis* are in yellow. Accession numbers are given for each sequence.

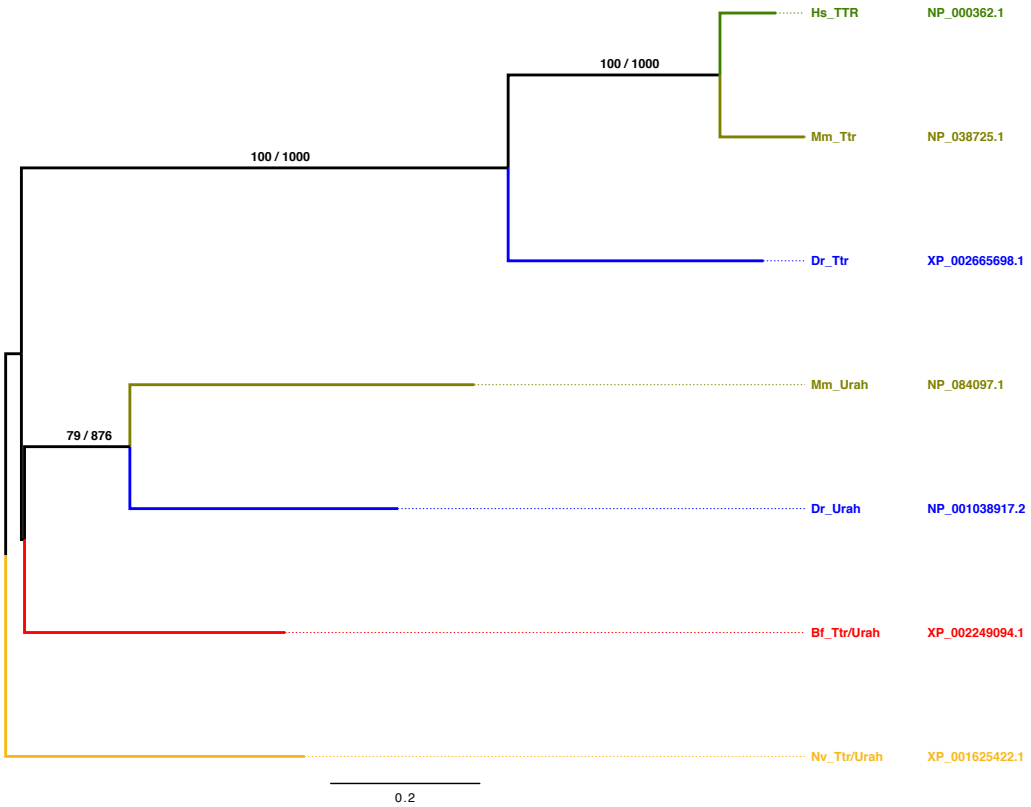


Fig. S10. Ttr/Urah phylogeny. The topology of the Maximum Likelihood tree is shown. Bootstrap values (Neighbor Joining/Maximum Likelihood) are indicated, if the Neighbor Joining value is at least 75 or the Maximum Likelihood value is at least 750. Human sequences are in green, zebrafish sequences are in blue, amphioxus sequences are in red, sequences from the cnidarian *Nematostella vectensis* are in yellow. Accession numbers are given for each sequence.

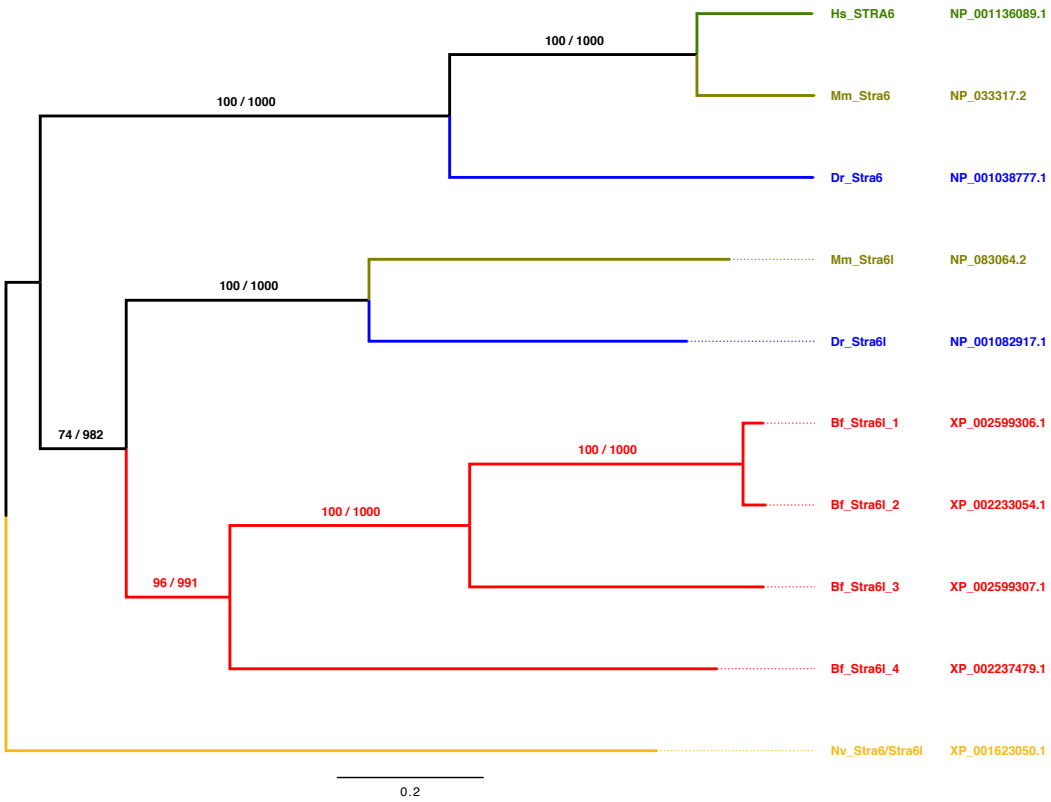


Fig. S11. Stra6 phylogeny. The topology of the Maximum Likelihood tree is shown. Bootstrap values (Neighbor Joining/Maximum Likelihood) are indicated, if the Neighbor Joining value is at least 75 or the Maximum Likelihood value is at least 750. Human sequences are in green, zebrafish sequences are in blue, amphioxus sequences are in red, sequences from the cnidarian *Nematostella vectensis* are in yellow. Accession numbers are given for each sequence.

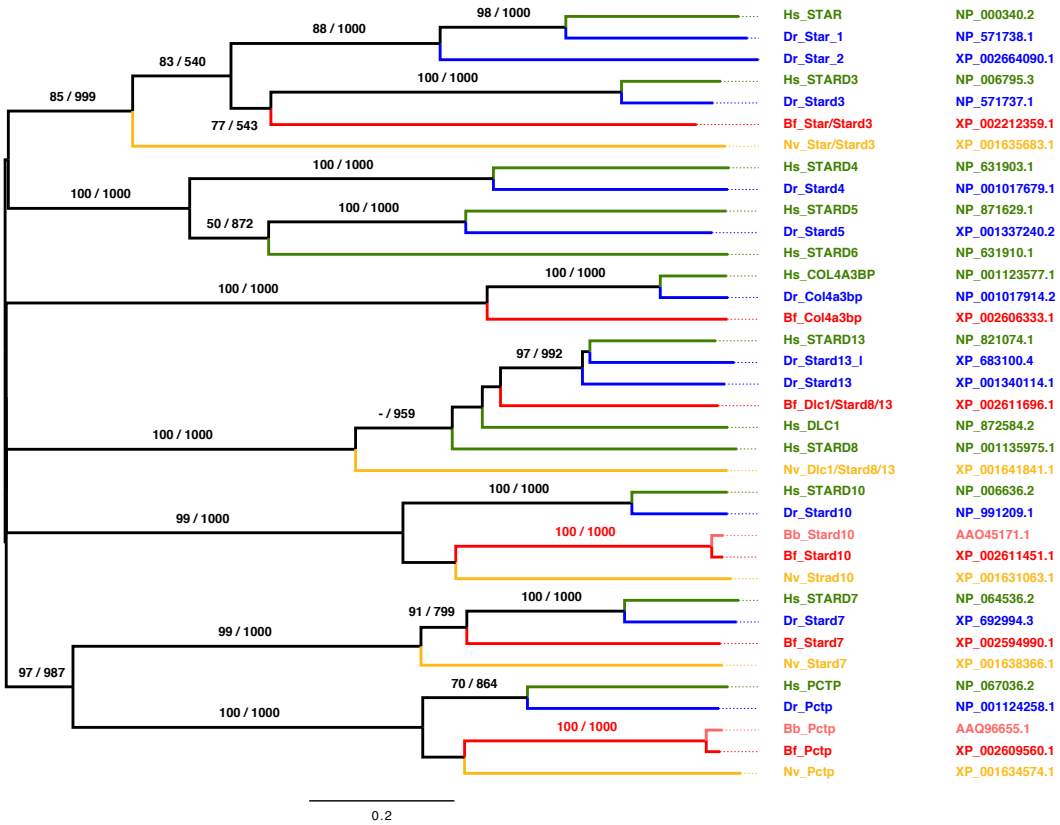


Fig. S12. Star phylogeny. The topology of the Maximum Likelihood tree is shown. Bootstrap values (Neighbor Joining/Maximum Likelihood) are indicated, if the Neighbor Joining value is at least 75 or the Maximum Likelihood value is at least 750. Human sequences are in green, zebrafish sequences are in blue, amphioxus sequences are in red, sequences from the cnidarian *Nematostella vectensis* are in yellow. Accession numbers are given for each sequence.

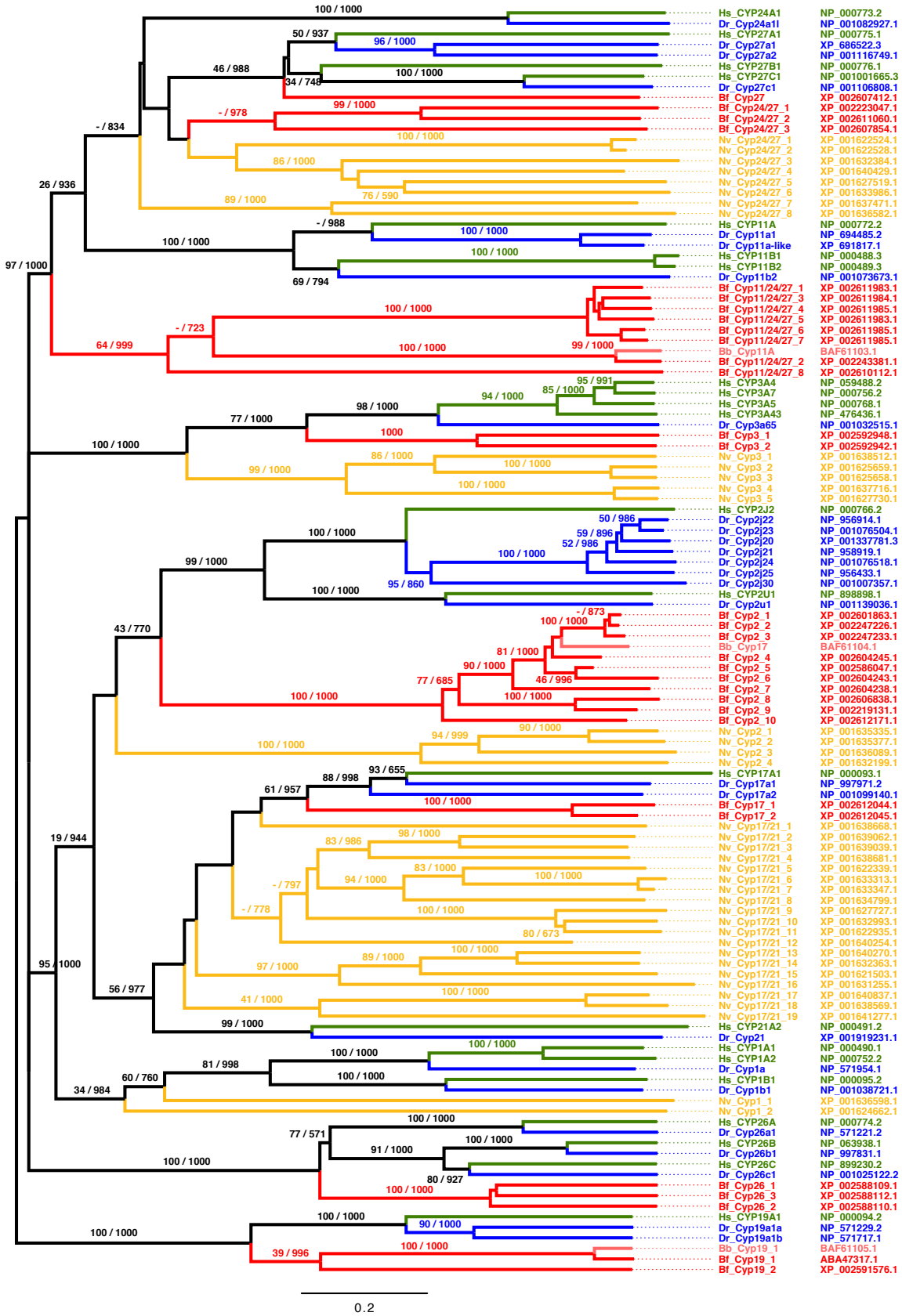


Fig. S13. Cyp phylogeny. The topology of the Maximum Likelihood tree is shown. Bootstrap values (Neighbor Joining/Maximum Likelihood) are indicated, if the Neighbor Joining value is at least 75 or the Maximum Likelihood value is at least 750. Human sequences are in green, zebrafish sequences are in blue, amphioxus sequences are in red, sequences from the cnidarian *Nematostella vectensis* are in yellow. Accession numbers are given for each sequence.

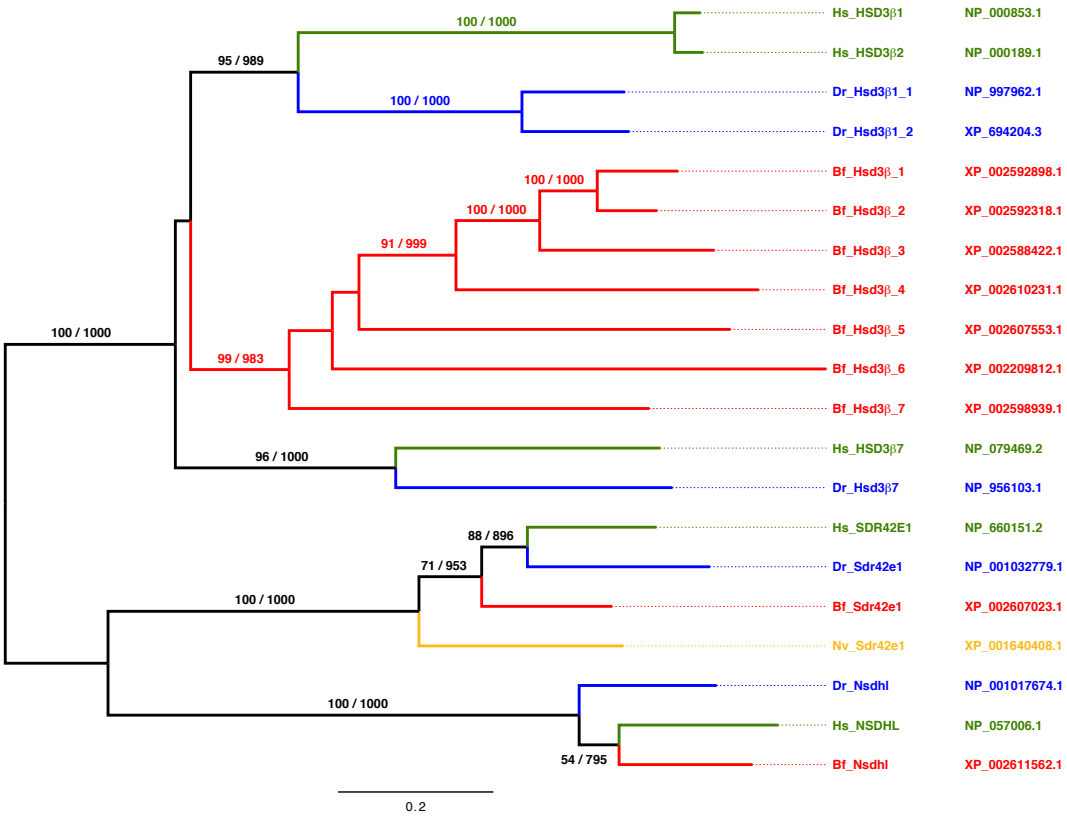


Fig. S14. Hsd3β phylogeny. The topology of the Maximum Likelihood tree is shown. Bootstrap values (Neighbor Joining/Maximum Likelihood) are indicated, if the Neighbor Joining value is at least 75 or the Maximum Likelihood value is at least 750. Human sequences are in green, zebrafish sequences are in blue, amphioxus sequences are in red, sequences from the cnidarian *Nematostella vectensis* are in yellow. Accession numbers are given for each sequence.

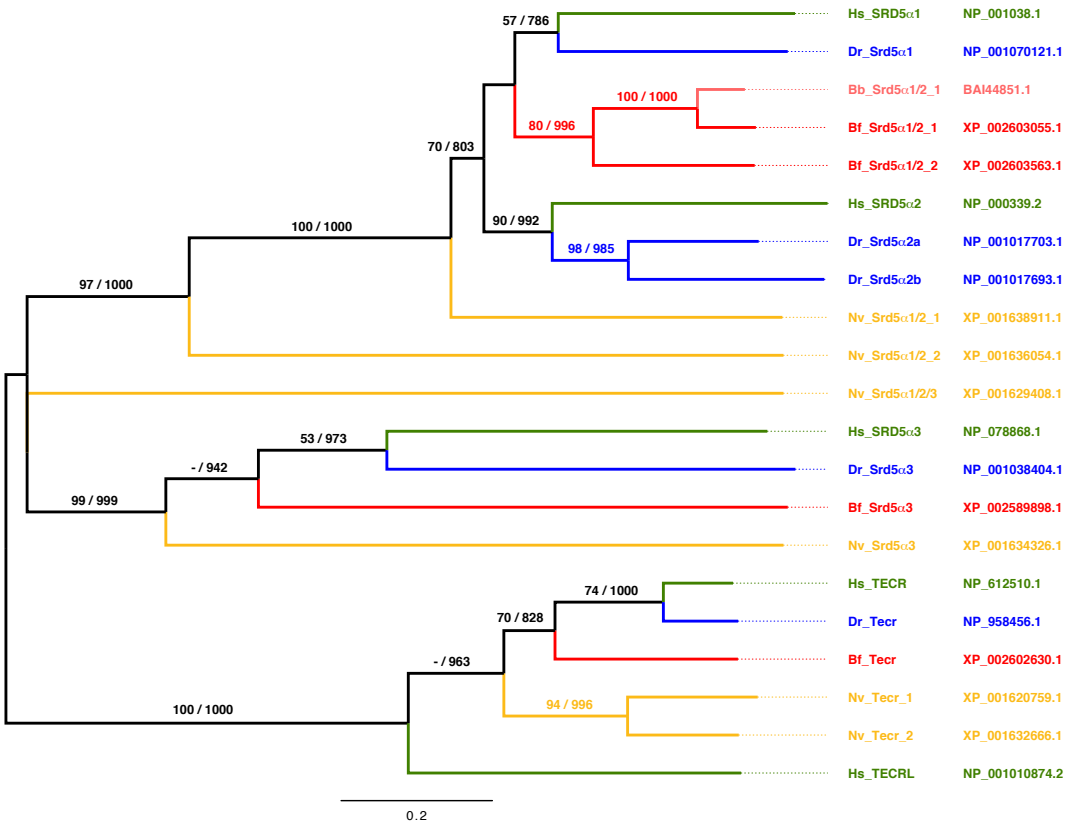


Fig. S15. Srd5α phylogeny. The topology of the Maximum Likelihood tree is shown. Bootstrap values (Neighbor Joining/Maximum Likelihood) are indicated, if the Neighbor Joining value is at least 75 or the Maximum Likelihood value is at least 750. Human sequences are in green, zebrafish sequences are in blue, amphioxus sequences are in red, sequences from the cnidarian *Nematostella vectensis* are in yellow. Accession numbers are given for each sequence.

Structural Shifts of Aldehyde Dehydrogenase Enzymes were Instrumental for the Early Evolution of Retinoid-Dependent Axial Patterning in Metazoans

Tiago J. P. Sobreira[†], Ferdinand Marlétaz[†], Marcos Simões-Costa[†], Deborah Schechtman, Alexandre C. Pereira, Frédéric Brunet, Sarah Sweeney, Ariel Pani, Jochanan Aronowicz, Christopher J. Lowe, Bradley Davidson, Vincent Laudet, Marianne Bronner-Fraser, Paulo S. L. de Oliveira, Michael Schubert[†], José Xavier-Neto[†]

[†]These authors contributed equally to this work.

Abstract

Aldehyde dehydrogenases (ALDHs) catabolize toxic aldehydes and process the vitamin A-derived retinaldehyde into retinoic acid (RA), a small diffusible molecule and a pivotal chordate morphogen. In this study, we combine phylogenetic, structural, genomic, and developmental gene expression analyses to examine the evolutionary origins of ALDH substrate preference. Structural modeling reveals that processing of small aldehydes, such as acetaldehyde, by ALDH2, versus large aldehydes, including retinaldehyde, by ALDH1A is associated with small versus large substrate entry channels (SECs), respectively. Moreover, we show that metazoan ALDH1s and ALDH2s are members of a single ALDH1/2 clade and that during evolution eukaryote ALDH1/2s often switched between large and small SECs after gene duplication, transforming constricted channels into wide opened ones and vice versa. Ancestral sequence reconstructions suggest that during the evolutionary emergence of RA signaling the ancestral, narrow-channeled metazoan ALDH1/2 gave rise to large ALDH1 channels capable of accommodating bulky aldehydes, such as retinaldehyde, supporting the view that retinoid-dependent signaling arose from ancestral cellular detoxification mechanisms. Our analyses also indicate that, on a more restricted evolutionary scale, ALDH1 duplicates from invertebrate chordates (amphioxus and ascidian tunicates) underwent switches to smaller and narrower SECs. When combined with alterations in gene expression, these switches led to neofunctionalization from ALDH1-like roles in embryonic patterning, to systemic, ALDH2-like, roles, suggesting functional shifts from signaling to detoxification.

Introduction

In animal development, major signaling pathways are controlled by morphogens, diffusible molecules whose evolutionary origins are difficult to assess. Aldehyde dehydrogenase (ALDH) enzymes are attractive subjects to study the evolution of morphogen signaling for two main reasons. First, in addition to their acknowledged role in protecting animals by catabolizing reactive biogenic and xenobiotic aldehydes, some ALDHs also synthesize signaling molecules (1-3). Prime examples for these two ALDH enzyme roles are the ALDH2s, which degrade small toxic aldehydes, such as the acetaldehyde derived from ethanol metabolism (1, 2), and the ALDH1s, which process larger aldehydes, including retinaldehyde, a vitamin A-derived precursor of the morphogen retinoic acid (RA). RA plays a critical role during embryonic development of chordates (i.e. amphioxus, tunicates and vertebrates) and has been suggested to have already been involved in patterning the last common ancestor of bilaterian animals (4-8). Second, ALDHs are among the best-characterized proteins and their structure and substrate profiles have been determined with exquisite precision (9-15). Thus, structural modeling of these proteins can be used to study the evolution of substrate specificity without extensive biochemical analyses (16-20).

ALDH1 and ALDH2 enzymes share a high degree of sequence identity, indicating a very close phylogenetic relationship (3). Pioneer observations by Moore et al. (1998) on human ALDH2 and sheep ALDH1 (17) suggested that their respective abilities to detoxify small aldehydes and to process large aldehydes are correlated with the size and shape of their substrate entry channels (SECs), the intramolecular cavities that direct aldehydes to the catalytic sites of ALDH enzymes. Human ALDH2 displays a narrow SEC with a constricted entrance, while sheep ALDH1A1, exhibits a large SEC with a broad opening (17, 18). Thus, SEC topology influences ALDH1/2 substrate preference. For example, while retinaldehyde is a good substrate for vertebrate ALDH1s and acetaldehyde is a natural substrate of ALDH2s, ALDH2s cannot process retinaldehyde and ALDH1s process acetaldehyde only extremely inefficiently (16, 19-22).

To understand the evolutionary origins of the substrate preferences of ALDH1 and ALDH2 enzymes, as well as to illuminate how signaling and protective functions are connected to these different enzyme activities, we used an integrated approach that combined genomic, phylogenetic, and structural analyses. The resulting comprehensive data set was complemented with information on developmental gene expression of ALDH1/2s in the cephalochordate amphioxus (*Branchiostoma floridae*) and the ascidian tunicate *Ciona intestinalis*. These two invertebrate chordate models possess functional RA signaling cascades and are pivotal models for understanding vertebrate origins from both a genomic and a developmental perspective (4, 23-26). Together, this work provides support for the hypothesis that some intercellular signaling mechanisms evolved from cellular detoxification pathways.

Results

Substrate entry channel volumes distinguish vertebrate ALDH1s from ALDH2s

To test whether SEC differences between sheep ALDH1 and human ALDH2 reflect fundamental evolutionary differences between these enzymes, we used the crystal structures of these proteins to create three dimensional (3D) structural models of ALDH1/2s. These models were then analyzed to determine molecular parameters,

such as SEC volume, which are implicated in substrate preference (17). Our dataset shows that ALDH1s generally display larger channel volumes than ALDH2s (589 ± 59 $Å^3$ for ALDH1s versus 403 ± 53 $Å^3$ for ALDH2s, mean ± SD, $p<0.001$) (data not shown). Therefore, channel volume represents a fundamental difference between ALDH1 and ALDH2, reflecting conserved structural requirements associated with processing of large and small aldehydes, respectively.

Since it is likely that the overall geometry of the SEC, rather than its volume alone, determines ALDH1/2 specificities, we looked for further mechanistic clues to the evolution of substrate preference in ALDH1/2 sequences that diverge between the biochemically well-characterized vertebrate ALDH1s and ALDH2s. We hence compiled a list of 34 amino acid (aa) sequence signatures distinguishing the large-channeled ALDH1As (known to synthesize RA) from the narrow-channeled ALDH2s (with well-characterized roles in detoxification of small toxic aldehydes). Six signatures mapped to a subset of the 27 aa that form the SEC (Table S1). Of those, only three signatures are positioned at critical locations at the ALDH1/2 channel, suggesting that they are implicated in determining ALDH1 and ALDH2 channel volumes/functions. The remaining signatures marked residues at oligomerization domains, surface loops or inside the molecule, which, after careful examination, did not suggest immediate functional correlates (Table S1).

The first of the three SEC signatures includes aa124 at the entrance ("mouth") of the ALDH1/2 channel (17). In human ALDH2, a bulky Met124 (124 $Å^3$ Van der Waals volume) protrudes into the channel (Table S1), allowing access of only small aldehydes, while in sheep ALDH1A1, a small, unobtrusive Gly124 (48 $Å^3$) allows entry of large aldehydes (17, 18). Thus, the first aa signature performs a size selection function. The second channel signature includes aa459 at the proximal third ("neck") of the ALDH1/2 SEC (Table S1) (17). In vertebrate ALDH2s, this aa is a large Phe459 (135 $Å^3$). In contrast, vertebrate ALDH1s typically display the smaller Val (93 $Å^3$) or Leu (124 $Å^3$) (Table S1). The third signature corresponds to aa303, close to the catalytic Cys302 (86 $Å^3$) at the end of the channel ("bottom") (17). In vertebrate ALDH2s, this aa is Cys303, while vertebrate ALDH1s typically display Thr303 (93 $Å^3$), Ile303 (124 $Å^3$) or Val303 (Table S1).

To understand the roles of the aa signatures in substrate interaction, we performed docking studies on human ALDH2 and sheep ALDH1 with small aldehydes (data not shown). In the ALDH2 channel, acetaldehyde is kept close to the γ-sulfur of the catalytic Cys302, while the ALDH1 channel does not favor this close association between the aldehyde substrate and Cys302, neither for acetaldehyde, nor for formaldehyde. Analysis of the structural motifs involved in substrate retention close to the ALDH catalytic site indicates that, in ALDH2, Cys302 and Phe465 keep acetaldehyde close to the catalytic γ-sulfur and that Phe465 is latched in position by Phe459, which in turn is fixed by Cys303. In ALDH1 this mechanism is not present, because the interaction surfaces between Cys303 and Phe459 are missing, due to their substitution by Ile303 and Val459, respectively. Thus, neck and bottom signatures keep the aldehyde moiety of small substrates close to the catalytic Cys302 in ALDH2s, consistent with a structural specialization of narrow-channeled ALDHs to process small aldehydes.

Switches from small to large ALDH1/2 substrate entry channels operated at the origin of RA signaling

To understand the evolution of affinities for small and large aldehydes in ALDH2 and ALDH1, we performed large-scale phylogenetic analyses of the ALDH

superfamily (Figs. 1, S1, S2). Contrary to traditional views, we found that metazoan ALDH1s and ALDH2s do not form independent families, but are members of a single, well-supported ALDH1/2 clade, with ALDH2s forming a distinct group nested within this ALDH1/2 clade (Figs. 1, S1, S2). In contrast to ALDH2s, ALDH1s underwent multiple lineage-specific duplications. For example, the genomes of amphioxus (*B. floridae*) and of the ascidian tunicates *C. intestinalis* and *C. savignyi* contain, respectively, six, four, and two ALDH1 duplicates, while in the hemichordate *Saccoglossus kowalevskii* there are five ALDH1 genes (Figs. 1, S1, S2) (27).

Analyses of channel size distribution in eukaryote ALDH1s and ALDH2s indicate that SEC variation can be subdivided into small (< 420 $Å^3$), medium (420-508 $Å^3$), and large channels (> 508 $Å^3$) (Fig. S3). In metazoans, small channels dominate in ALDH2s, while large channels preponderate in ALDH1s (Fig. 1). In plants, there are also two major ALDH1/2 groups, one with small and medium channels and another one with predominantly large channels (345 ± 47 $Å^3$ versus 561 ± 78 $Å^3$, $p<0.05$), suggesting that a single ALDH1/2 ancestor duplicated early in plant evolution giving rise to two distinct functional classes (28). Fungi experienced a distinct diversification pattern with various fungal lineages independently duplicating a single ALDH1/2 ancestor (29). These duplicates subsequently underwent SEC alterations leading to fungal ALDH1/2 enzymes with small, medium or large SECs (Fig. 1).

To understand how small, medium, and large SEC volumes arose during evolution of the ALDH1/2 clade, we reconstructed ancestral sequences at selected nodes of the ALDH1/2 phylogeny using the maximum likelihood method (Table 1). For example, the reconstructed ancestral eukaryote ALDH1/2 enzyme displays a small SEC, in which a bulky Met124 blocks the channel mouth, Phe459 constricts the channel neck, and Cys303 occupies the channel bottom, similar to modern metazoan ALDH2 SECs (Table 1). This indicates that the vertebrate ALDH2 SEC signatures that distinguish them from vertebrate ALDH1s are ancestral, rather than derived, reflecting ancient structural adaptation to process small aldehydes. Table 1 also shows that the emergence of large-channeled ALDH1 enzymes was accompanied by substitution of an ancestral bulky Met124 at the SEC mouth by small aa, such as Gly and Ala. This indicates the wide open channels of ALDH1 enzymes, which process large aldehydes, such as retinaldehyde, evolved from a background of small, constricted channels similar to those displayed by vertebrate ALDH2s, which process small, toxic aldehydes.

ALDH1 duplication and divergence switched large into narrow substrate entry channels in invertebrate chordates

Curiously, the three SEC signatures that efficiently discriminate large-channeled vertebrate ALDH1s from narrow-channeled vertebrate ALDH2s do not distinguish their invertebrate chordate counterparts, because some invertebrate chordate ALDH1s display channel signatures typical for vertebrate ALDH2s (Table S1). To determine, how these SEC variations evolved in the framework of the multiple, lineage-specific, ALDH1 duplications of invertebrate chordates, we focused on the cephalochordate amphioxus, whose genome encodes six ALDH1s and a single ALDH2 (27), and on two closely-related ascidian tunicates: *C. intestinalis* with four ALDH1s and a single ALDH2, and *C. savignyi* with two ALDH1s and a single ALDH2.

Since the first ALDH2 signature at the channel mouth probably selects for smaller substrates, we hypothesized that bulky and small residues may be similarly

present in invertebrate chordate ALDH2s and ALDH1s, respectively. Accordingly, a bulky Leu124 protrudes into the channel mouth of amphioxus, *C. intestinalis*, and *C. savignyi* ALDH2s (Figs. 2, 3, S4). Amphioxus ALDH1s are heterogeneous in that a small Gly124 is embedded into the channel border without constricting the channel in amphioxus ALDH1a and d, while the larger Glu124 (109 $Å^3$) or Ser124 (73 $Å^3$) partially obstruct the channel in the other four amphioxus duplicates, leaving insufficient space to accommodate the β-ionone moiety of retinaldehyde (Fig. 2). The ALDH1s from *C. intestinalis* and *C. savignyi* are also heterogeneous: *C. intestinalis* ALDH1a and d, as well as *C. savignyi* ALDH1a display the small Gly and *C. savignyi* ALDH1b the small Ala (67 $Å^3$) at position 124, which do not obstruct the channel entrance, while *C. intestinalis* ALDH1b and c respectively display the larger Ser124 and Ile124, which interfere with retinaldehyde accommodation in the ALDH1 channel (Figs. 2, 3, S4).

In the ALDH2s from amphioxus, *C. intestinalis*, and *C. savignyi*, the aa at the second channel signature is Phe459, which constricts the channel neck with its large aromatic ring. As in vertebrate ALDH1s, in some amphioxus ALDH1s, the bulky Phe459 is substituted by smaller aa, such as Ile459 in ALDH1a, b, and d, Thr459 (93 $Å^3$) in ALDH1c or Gly459 in ALDH1e and f. In ascidian tunicates, only *C. intestinalis* ALDH1d displays a Leu459, while all other ALDH1s display bulky aa, such as Phe and Met at position 459 (Figs. 3, S4), similar to vertebrate ALDH2s.

As in vertebrates, in amphioxus, *C. intestinalis*, and *C. savignyi*, the aa of the third ALDH2 signature at the channel bottom is Cys303. Amphioxus ALDH1s are heterogeneous in that ALDH1a and d display the vertebrate ALDH1 pattern (Thr303 and Ile303, respectively), while all other amphioxus ALDH1s display the vertebrate ALDH2 pattern (Fig. 2). In ascidian tunicates, only *C. intestinalis* and *C. savignyi* ALDH1a display the vertebrate ALDH1 pattern with Thr303, while other ALDH1s show the vertebrate ALDH2 pattern (Figs. 3, S4). Thus, after lineage-specific duplication some ALDH1 enzymes of invertebrate chordates incorporated aa and/or structural motifs similar to those of the vertebrate ALDH2 channel, shifting from a large, wide open configuration to constricted SEC topologies.

Synteny analyses in amphioxus and *C. intestinalis* suggest that the ALDH1a genes from both species are the sister groups to the other cephalochordate and ascidian tunicate ALDH1s and that the amphioxus duplicates ALDH1b, c, d, e, and f and *C. intestinalis* ALDH1b, c, and d evolved by duplication from an ALDH1a-like ancestor (Fig. S5). Moreover, reconstructions of ancestral SEC signatures and channel structures indicate that amphioxus and *C. intestinalis* ALDH1 ancestors displayed large SECs, structurally consistent with the capacity to process retinaldehyde into RA (Figs. 1-3, S2). The reconstructed amphioxus ALDH1 ancestor already displayed a typical ALDH1 SEC with a 512 $Å^3$ wide opening lined by Gly124, an unobstructed neck flanked by Val459, and a bottom Cys303. The reconstructed *C. intestinalis* ALDH1 ancestor exhibits a large, 540 $Å^3$ SEC, but, curiously, with ancestral Met124, Phe459, and Cys303 SEC signatures, suggesting that large ALDH1 channels emerged independently and by different mechanisms in cephalochordates and ascidian tunicates.

Invertebrate chordate ALDH1 channel switch is associated with transitions between restricted and pleiotropic expression

Our next goal was to understand the specific developmental contexts in which the invertebrate chordate ALDH1/2s are deployed. We hence assessed developmental expression of the large-channeled ALDH1s structurally capable of

accommodating retinaldehyde for RA synthesis, of the narrow-channeled ALDH2 genes adapted for small aldehyde detoxification, and of the divergent, narrow-channeled ALDH1s.

The ALDH1a and d genes of amphioxus and *C. intestinalis* encode enzymes with large and unobstructed SECs. Amphioxus ALDH1a is expressed caudally close to the developing tail bud with a sharp anterior boundary in the neurula (at 12 hours) (Fig. 2). In *C. intestinalis*, ALDH1a is expressed in a sharp, posterior mesodermal domain in the early embryo (at 5 to 8 hours) (Fig. 3) (30). At later embryonic stages, ALDH1a is detectable in a distinct domain in the posterior gut endoderm of the amphioxus late embryo (at 24 hours) and in the posterior trunk of *C. intestinalis* (at 10 to 12 hours). In amphioxus, expression of ALDH1d overlaps that of ALDH1a at the neurula stage, but diverges from that of ALDH1a in the late embryo. At this stage, amphioxus ALDH1d expression is broad and inconspicuous with a moderate accentuation of the signal in the posterior gut endoderm. In *C. intestinalis*, ALDH1d expression is diffuse and weak in the gastrula and becomes diffusely transcribed in the trunk at 10 to 12 hours of development. Thus, in both amphioxus and *C. intestinalis* ALDH1a is consistently expressed in patterns that are entirely consistent with a role in anteroposterior patterning and that are reminiscent of expression of vertebrate ALDH1A2 (RALDH2), which define posterior identity in the vertebrate embryo (31).

In both amphioxus and *C. intestinalis*, there is only a single gene encoding a narrow-channeled ALDH2 enzyme. In amphioxus, ALDH2 expression is restricted to posterior, mesendodermal tissues at 12 hours of development and subsequently spreads throughout the embryo at 24 hours (Fig. 2). In *C. intestinalis*, ALDH2 expression is strong and diffuse at early and late developmental stages (Fig. 3). Thus, ALDH2 genes are expressed in widespread patterns during development of invertebrate chordates.

There are a total of six genes encoding ALDH1s with narrow channels in amphioxus and *C. intestinalis*: amphioxus ALDH1b, c, e, and f, and *C. intestinalis* ALDH1b and c. In amphioxus, these duplicates are weakly expressed in posterior domains overlapping that of ALDH1a in the neurula (at 12 hours) (Fig. 2). However, by 24 hours, they are expressed diffusely and weakly throughout the amphioxus embryo with a weak to moderate concentration of the signal for ALDH1b, c, and e in the posterior gut endoderm. In *C. intestinalis*, ALDH1b is diffusely transcribed in trunk and tail, while ALDH1c is not detectable by in situ hybridization in developing embryos (Fig. 3). Thus, the genes encoding narrow-channeled ALDH1s in amphioxus and *C. intestinalis* generally display either widespread or inconspicuous expression patterns during development, suggesting that these enzymes are not playing major roles in anteroposterior patterning of the embryo.

Discussion

Structural insights into ALDH1 and ALDH2 function

Our modeling studies confirm the notion, first determined by Moore et al. (1998) (17) with only two enzymes, that substrate access channel size is a crucial determinant of ALDH1 and ALDH2 function. Here we extend this concept to the whole metazoan ALDH1/2 clade and provide novel insights on structural adaptations underlying the ancestral ability of narrow-channeled ALDH1/2 enzymes to process small aldehydes and into the evolution of the ability of large-channeled ALDH1s to process large aldehydes.

We determined that position 124 at the channel entrance is a selective gate for aldehyde size. Ancestral reconstructions show that, throughout metazoan ALDH1/2 evolution, increases in channel size associated with substitution of the bulky, ancestral, Met124 by small Ala124 or Gly124 (Table 1). The selective abilities of ALDH1/2 channels are thus regulated by the presence or absence of a steric hindrance to the entry of large aldehydes into the ALDH1/2 channel. Thus, we show that the ALDH2 channel cannot accommodate retinaldehyde, which is consistent with earlier studies (16, 22) showing that retinaldehyde is not an ALDH2 substrate, but a competitive inhibitor of acetaldehyde degradation. This contrasts with the ease with which retinaldehyde is admitted into the ALDH1 channel. Thus, size selection is a fundamental feature of substrate discrimination by the ALDH1/2 channel. We have also shown that, while ALDH2 can keep small aldehydes close to the catalytic Cys302 long enough for catalysis, ALDH1s cannot. Therefore, small aldehyde processing by ALDH2s requires specific structural adaptations to reduce substrate mobility in their channels, an ability that is lacking in large-channeled ALDH1s (16).

Switches between small and large substrate entry channels are common in ALDH1/2 evolution

The changes in the ALDH1 and ALDH2 SEC that we describe reflect a tendency of eukaryote ALDH1/2 genes to alter the structures of their encoded enzymes after gene duplication (Figs. 2, 3, S4, S6). By accumulating mutations at the mouth, neck and/or SEC bottom, ALDH1/2s underwent changes in SEC geometry and/or overall volume, which affected their structural abilities to accommodate their original substrates. This led to switches from small, constricted channels adapted to the handling of small aldehydes to large, broadly opened channels adjusted to receive large aldehyde molecules or vice versa (Fig. 4). Our results provide an important contrast to studies proposing a general non-reversibility of aa changes involved in the functional adaptation of proteins (32).

ALDH1/2 switches and the origins of RA signaling

Although ALDH enzymes can catalyze a range of different substrates, the molecular switches between ALDH1 and ALDH2 SECs reported here very likely represent functional transitions between the ability of ALDH1/2s to process small, toxic, aldehydes for defense against endogenous and xenobiotic aldehyde aggression and its capacity to generate signaling molecules from larger aldehyde precursors. Using ancestral sequence reconstruction, we provide evidence that ALDH1/2 switches were important for the emergence of ALDH1 retinaldehyde dehydrogenases, which probably originated after gene duplication early in metazoan evolution, when a small, narrow-channeled ALDH1/2 ancestor, structurally related to modern ALDH2s, gave rise to a gene encoding a larger SEC capable of accommodating bulkier molecules, including retinaldehyde. This evolutionary scenario supports the view that RA signaling evolved from enzymes implicated in detoxification (3) and, combined with the description of a retinoic acid receptor (RAR) and of other RA cascade members in both protostomes and deuterostomes, pushes the origins of RA signaling to much earlier times than traditionally assumed (4, 7).

ALDH1s underwent independent duplication and extensive diversification in metazoans

Our data substantiate the notion that the metazoan ALDH1 ancestor originated from a eukaryote ALDH1/2 ancestor and underwent duplication before the

origin of bilaterian animals. It is also evident that ALDH1s duplicated independently in various animal groups and underwent extensive diversification, which is supported by two ALDH1 genes (one with a large SEC) in the cnidarian *Nematostella vectensis* and, in amphioxus and *C. intestinalis*, by structurally dissimilar ALDH1 ancestors and by the distinct arrangement of ALDH1 SEC signatures. While duplication and diversification are common in ALDH1s, the metazoan ALDH2s are typically preserved as single copies, and their SECs have kept the same constricted features of the eukaryote ALDH1/2 ancestor. These repeated patterns of ALDH1 duplication and divergence, together with the presence of fast evolving sequences has prevented the resolution of phylogenetic relationships at the base of metazoan ALDH1s, as well as the reconstruction of the metazoan ALDH1 ancestor, which contrasts with our reliable reconstruction of the ALDH2 ancestor as a gene encoding constricted SECs lined by the same aa signatures as the narrow-channeled, eukaryote ALDH1/2 ancestor.

Invertebrate ALDH1 switches suggest shifts from patterning to detoxification

The presence of ALDH1 duplicates in a given animal raises questions about the roles of each duplicate within the RA signaling cascade (4, 7). ALDH1 duplicates in amphioxus and *C. intestinalis* originated by duplication from an ALDH1 ancestor with a large SEC structurally compatible with RA synthesis (Fig. S5). This structure is present in their ALDH1a paralogs, which display sharp posterior domains, consistent with early embryonic anteroposterior patterning. In contrast, genes encoding ALDH1b, c, e, and f in amphioxus and ALDH1b and c in *C. intestinalis* accumulated mutations resulting in constricted ALDH SECs poorly suited to accommodate large molecules, but still capable of admitting small aldehydes. These genes display broad expression patterns suggesting that they have evolved novel functions probably associated with the processing of small, toxic aldehydes for protection against endogenous or xenobiotic aldehydes (1-3).

A plausible scenario leading from patterning to protective ALDH roles can be derived from the expression patterns of the ALDH1d genes of amphioxus and *C. intestinalis*. The molecular structures of the SECs of these two ALDH1ds are consistent with retinaldehyde processing. However, ALDH1d expression is rather broad throughout the amphioxus embryo and the *C. intestinalis* embryonic trunk, suggesting that changes in gene regulation of these two genes occurred after duplication and that these changes were not accompanied by structural remodeling of the SEC.

The transition from restricted signaling functions to generalized roles has not been completed to equivalent degrees in each of the divergent amphioxus duplicates. ALDH1b, c, e, and f developed diffuse patterns in the neurula, while, curiously maintaining weak, but restricted, posterior domains in early gastrulation and, except for Aldh1f, a relative concentration of expression in the posterior gut endoderm, which are likely to represent the ancestral, ALDH1a-like pattern. In *C. intestinalis*, ALDH1b is diffusely and inconspicuously expressed in the trunk, while ALDH1c expression is not detectable during embryogenesis, but seems to be restricted to adult tissues, as indicated by EST database searches.

The fate of these ALDH1 duplicates in amphioxus and *C. intestinalis* is consistent with an evolutionary scenario involving neofunctionalization after duplication, with gene duplicates acquiring more generalized functions during embryogenesis and possibly in the adult. Therefore, in amphioxus and *C. intestinalis*, some duplicated ALDH1s experienced modifications of gene regulation and protein

structure that resulted in neofunctionalization of the duplicates, possibly away from roles in axial patterning, towards generalized, pleiotropic functions similar to those of ALDH2, an enzyme important for protection against small aldehyde toxicity in chordates (33).

The ALDH1/2 case and its implications for anatomic and physiological evolution

Mutations of regulatory regions have been regarded as the major driving force of morphological evolution in development (34), while mutations in coding regions are viewed as major determinants of physiological evolution (35). Here, we describe regulatory alterations affecting duplicated ALDH1/2 genes of amphioxus and ascidian tunicates that are accompanied by fundamental structural shifts of the proteins they encode. Therefore, our data suggest that distinctions between anatomical and physiological evolution may not always be so clear cut and that rapid evolution of novel functions can be achieved when regulatory and protein structure mutations are superimposed after gene duplication, a hypothesis that provides a common ground for these two evolutionary mechanisms that have traditionally been thought to depend on distinct mechanisms. In sum, the ALDH1/2 case probably represents one of many examples that are likely to emerge with the incorporation of protein structure analyses into the collection of approaches used to study the evolution of body plans.

Materials and Methods

Phylogenetic analyses

Whole genomes, EST databases and trace repositories were mined for ALDH sequences using both signature (InterPro IPR002086) and global similarity searches. After sequence alignment, phylogenetic analyses were carried out using both maximum likelihood (RAxML-VI-HPC) and Bayesian inference (MrBayes 3.1) assuming PROTMIXWAG and WAG+I+Γ_4 parameters, respectively (36, 37). Both methods resulted in essentially identical tree topologies. Ancestral sequences at internal nodes of the ALDH1/2 phylogeny were reconstructed using PAML 3.15 based on a data matrix of 447 aa residues using a WAG+Γ_4 model (38).

3D modeling and volume calculations

ALDH1 and ALDH2 structures were modeled by homology using the 3D structure of sheep ALDH1A1 (PDB code 1BXS) or of human ALDH2 in complex with the dipsogenic inhibitor daidzin (PDB code 1OF7). Refined sequence alignment was used as input data for the modeling program Nest in Jackal (39). Graphical analyses were performed with VMD (Visual Molecular Dynamics) (40). For volume calculations, ALDH models were individually placed inside a grid box, in which x, y, and z coordinates were spaced by 0.8 Å and void regions inside this box were determined by moving a probe molecule with a 1.4 Å radius through the grid. ALDH aa residue numbers are based on the classical numbering of the mature human ALDH2 enzyme with the catalytic Cys at position 302 (17).

Gene expression studies

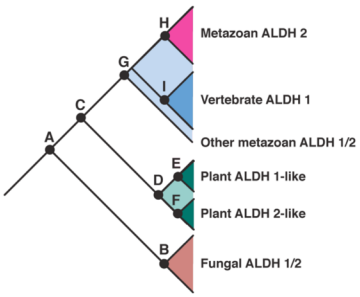
Sexually mature adults of the Florida amphioxus (*B. floridae*) were collected by shovel and sieve in Tampa Bay, Florida. *C. intestinalis* adults were obtained from M-Rep (San Diego, USA). Amphioxus ALDH1/2 genes were retrieved from cDNA libraries (41) and *C. intestinalis* clones were obtained from the Gene Collection

Release 1 (42). Embryo collection and in situ hybridization were carried out as previously described for amphioxus (43) and *C. intestinalis* (44).

Acknowledgements

The authors are grateful to Gérard Benoit, Tiago Pereira as well as to Linda Z. Holland and Nicholas D. Holland for critical reading of the manuscript. This work was supported by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP #06/50843-0) to J.X.N., by funds from ANR (ANR-07-BLAN-0038 and ANR-09-BLAN-0262-02), CNRS, and MENRT to M.S. and by CRESCENDO, a European Union Integrated Project of FP6.

Table 1. Reconstruction of ancestral sequences in the ALDH1/2 family.



Node	Support*	Sequence description	P#	Signature†	Channel size	
					Volume	Category
A	1 / 75	Eukaryote ALDH 1/2 ancestor	0.915	Met,Phe,Cys	378.4 Å^3	Small
B	0.98 / 59	Fungal ALDH 1/2 ancestor	0.918	Met,Phe,Cys	336.5 Å^3	Small
C	1 / 94	Plant/Metazoan ALDH 1/2 ancestor	0.937	Met,Phe,Cys	287.4 Å^3	Small
D	1 / 100	Plant ALDH 1/2 ancestor	0.912	Met,Phe,Cys	286.0 Å^3	Small
E	1 / 100	Plant ALDH 1-like ancestor	0.914	Ala,Phe,Val	559.2 Å^3	Big
F	1 / 100	Plant ALDH 2-like ancestor	0.917	Met,Phe,Cys	333.8 Å^3	Small
G	1 / 65	Metazoan ALDH 1/2 ancestor	0.948	Met,Phe,Cys	368.8 Å^3	Small
H	1 / 69	Metazoan ALDH 2 ancestor	0.966	Met,Phe,Cys	444.9 Å^3	Medium
I	1 / 92	Vertebrate ALDH 1 ancestor	0.956	Gly,Leu,Thr	604.3 Å^3	Large

*Phylogenetic support for the node, where the ancestral sequence was reconstructed (posterior probability / bootstrap value).
#Mean probability of the reconstructed ancestral sequence.
†The order of the amino acids corresponds to the mouth, neck, and bottom signatures, respectively.

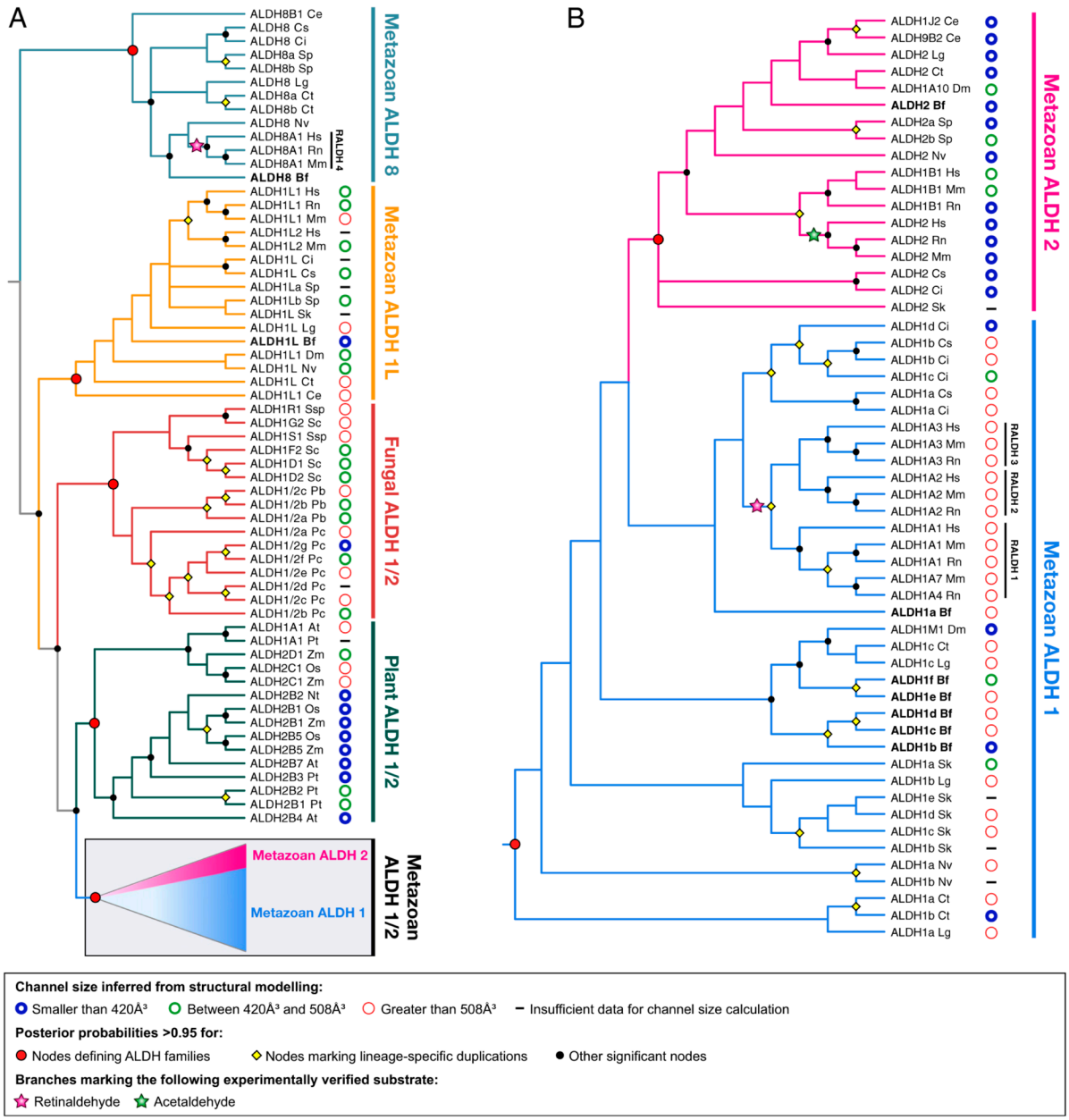


Fig. 1. Phylogeny of ALDH1/2s, ALDH1Ls, and ALDH8s. The overall topology with the ALDH8s used as outgroup is shown in (*A*). The boxed area in (*A*) highlights the metazoan ALDH1/2 clade, which is depicted in (*B*). Nodes with significant posterior probabilities (>0.95) are highlighted with icons. The actual values are given in Fig. S2. Channel size categories are shown diagrammatically for ALDH1/2s and ALDH1Ls with sufficient sequences. ALDH8s are too divergent for these calculations. At, *Arabidopsis thaliana*; Bf, *Branchiostoma floridae*; Ct, *Capitella teleta*; Ce, *Caenorhabditis elegans*; Ci, *Ciona intestinalis*; Cs, *Ciona savignyi*; Dm, *Drosophila melanogaster*; Hs, *Homo sapiens*; Lg, *Lottia gigantea*; Mm, *Mus musculus*; Nv, *Nematostella vectensis*; Nt, *Nicotiana tabacum*; Os, *Oryza sativa*; Pc, *Phanerochaete chrysosporium*; Pb, *Phycomyces blakesleeanus*; Pt, *Populus trichocarpa*; Rn, *Rattus norvegicus*; Sc, *Saccharomyces cerevisiae*; Sk, *Saccoglossus kowalevskii*; Sp, *Strongylocentrotus purpuratus*; Ssp, *Schizosaccharomyces pombe*; Zm, *Zea mays*.

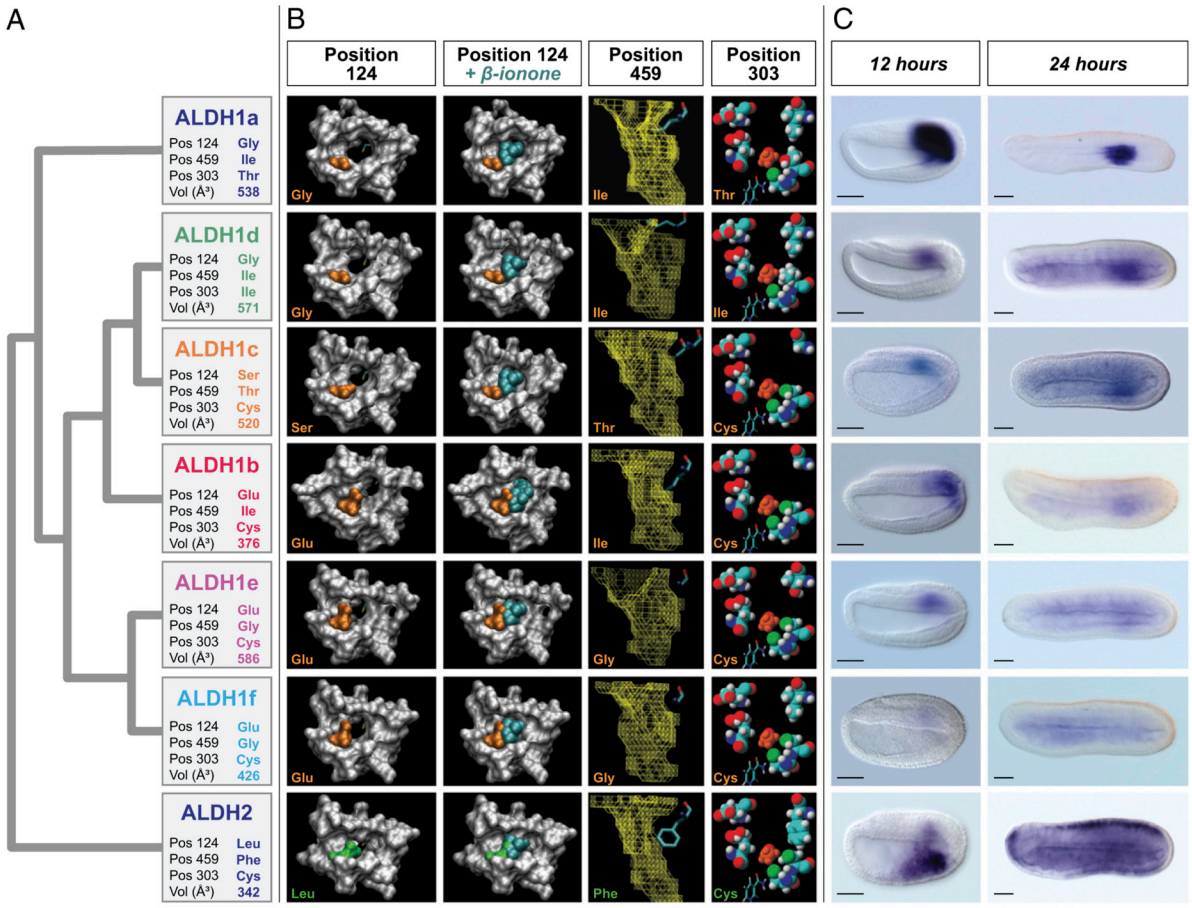


Fig. 2. Amphioxus ALDH1/2 duplicates. Phylogeny (*A*), channel structure (*B*), and developmental expression (*C*). Amino acid signatures of the substrate entry channel at positions 124 (the mouth), 459 (the neck), and 303 (the bottom) are indicated. For the expression analyses, neurulae (12 hours) and late embryos (24 hours) are shown.

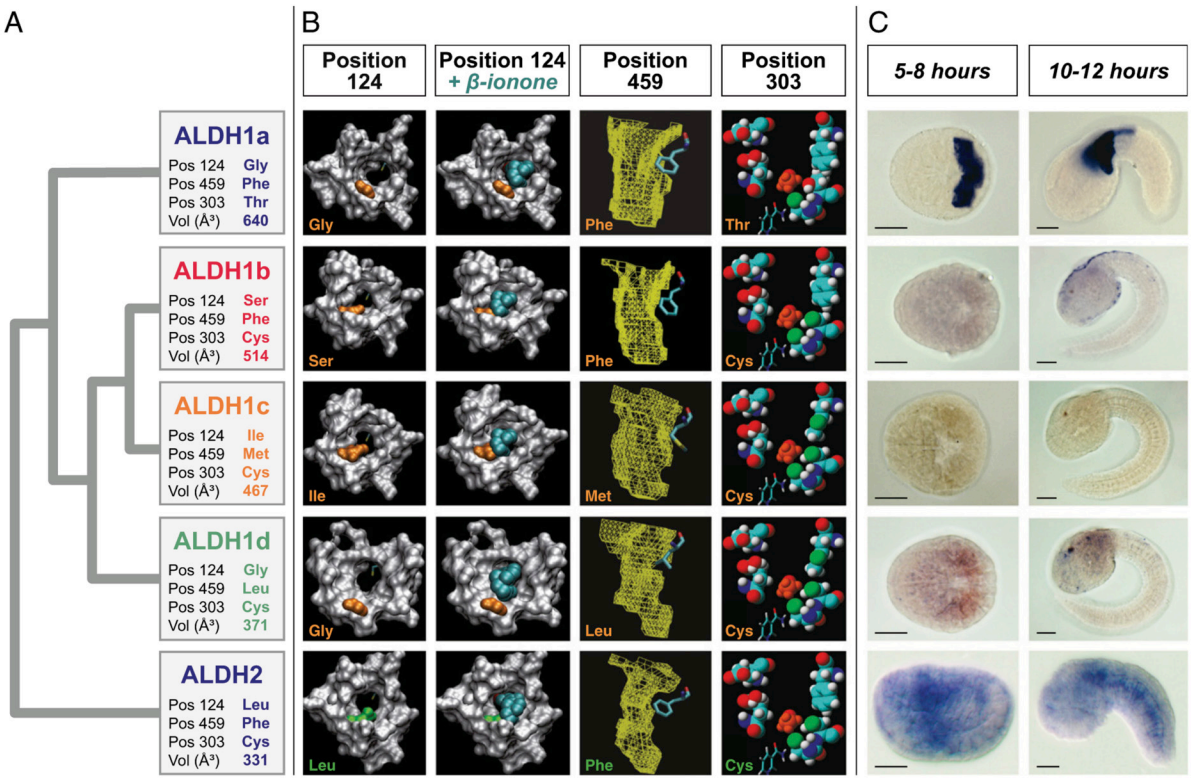


Fig. 3. *Ciona intestinalis* ALDH1/2 duplicates. Phylogeny (*A*), channel structure (*B*), and developmental expression (*C*). Amino acid signatures of the substrate entry channel at positions 124 (the mouth), 459 (the neck), and 303 (the bottom) are indicated. For the expression analyses, early embryos (5-8 hours) and tailbud stage embryos (10-12 hours) are shown.

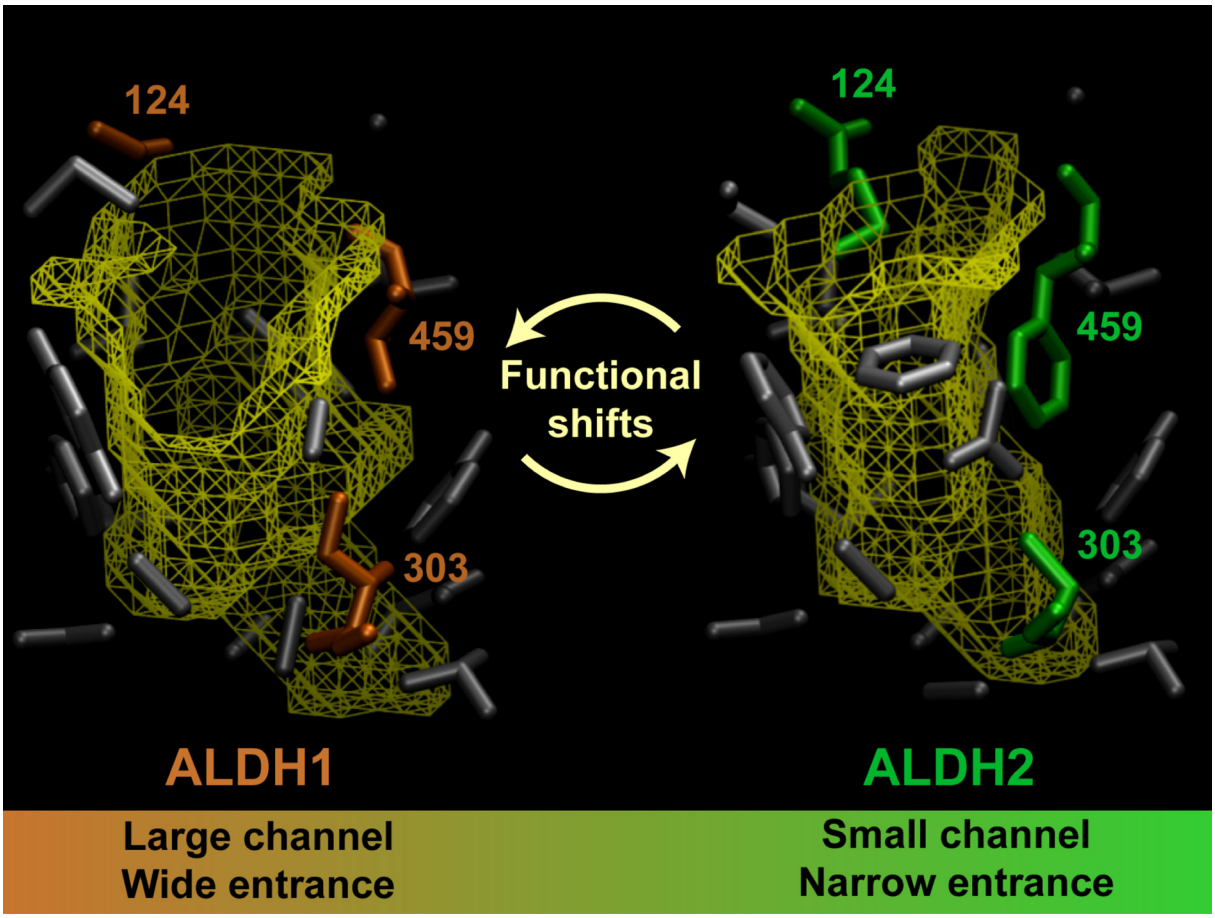


Fig. 4. Gene duplication and functional evolution in the ALDH1/2 family. Metazoan ALDH1s are generally characterized by a large substrate entry channel (SEC) with a wide channel entrance, while metazoan ALDH2s have a small SEC with a narrow channel entrance (the three channel signatures important for defining the ALDH1 and ALDH2 SECs are indicated: position 124 located at the channel mouth, position 459 at the neck of the channel, and position 303 at the channel bottom). After duplication, ALDH1/2 duplicates accumulated mutations in their SECs that led to smaller channel sizes in ALDH1 duplicates (e.g. amphioxus and ascidian tunicate ALDH1s) and to bigger channel volumes in ALDH2 duplicates (e.g. vertebrate ALDH1B1 and sea urchin ALDH2). Hence, duplicated ALDH1/2s experienced functional shifts of their SECs, which are indicative of changes in specificity towards aldehydes of different sizes.

References

1. Seiler N, Eichentopf B (1975) 4-aminobutyrate in mammalian putrescine catabolism. *Biochem J* 152:201-210.
2. Kikonyogo A, Pietruszko R (1996) Aldehyde dehydrogenase from adult human brain that dehydrogenates gamma-aminobutyraldehyde: purification, characterization, cloning and distribution. *Biochem J* 316:317-324.
3. Yoshida A, Rzhetsky A, Hsu LC, Chang C (1998) Human aldehyde dehydrogenase gene family. *Eur J Biochem* 251:549-557.
4. Campo-Paysaa F, Marlétaz F, Laudet V, Schubert M (2008) Retinoic acid signaling in development: tissue-specific functions and evolutionary origins. *Genesis* 46:640-656.
5. Marlétaz F, Holland LZ, Laudet V, Schubert M (2006) Retinoic acid signaling and the evolution of chordates. *Int J Biol Sci* 2:38-47.
6. Duester G (2008) Retinoic acid synthesis and signaling during early organogenesis. *Cell* 134:921-931.
7. Simões-Costa MS, Azambuja AP, Xavier-Neto J (2008) The search for non-chordate retinoic acid signaling: lessons from chordates. *J Exp Zool B Mol Dev Evol* 310:54-72.
8. Theodosiou M, Laudet V, Schubert M (2010) From carrot to clinic: an overview of the retinoic acid signaling pathway. *Cell Mol Life Sci* 67:1423-1445.
9. Lamb AL, Newcomer ME (1999) The structure of retinal dehydrogenase type II at 2.7 Å resolution: implications for retinal specificity. *Biochemistry* 38:6003-6011.
10. Dräger UC, Wagner E, McCaffery P (1998) Aldehyde dehydrogenases in the generation of retinoic acid in the developing vertebrate: a central role of the eye. *J Nutr* 128:463S-466S.
11. Graham CE, Brocklehurst K, Pickersgill RW, Warren MJ (2006) Characterization of retinaldehyde dehydrogenase 3. *Biochem J* 394:67-75.
12. 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.
13. Niederreither K, Subbarayan V, Dollé P, Chambon P (1999) Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. *Nat Genet* 21:444-448.
14. Wang X, Penzes P, Napoli JL (1996) Cloning of a cDNA encoding an aldehyde dehydrogenase and its expression in *Escherichia coli*. Recognition of retinal as substrate. *J Biol Chem* 271:16288-16293.
15. Zhao D*, et al.* (1996) Molecular identification of a major retinoic-acid-synthesizing enzyme, a retinaldehyde-specific dehydrogenase. *Eur J Biochem* 240:15-22.
16. Klyosov AA (1996) Kinetics and specificity of human liver aldehyde dehydrogenases toward aliphatic, aromatic, and fused polycyclic aldehydes. *Biochemistry* 35:4457-4467.
17. Moore SA*, et al.* (1998) Sheep liver cytosolic aldehyde dehydrogenase: the structure reveals the basis for the retinal specificity of class 1 aldehyde dehydrogenases. *Structure* 6:1541-1551.

18. Steinmetz CG, Xie P, Weiner H, Hurley TD (1997) Structure of mitochondrial aldehyde dehydrogenase: the genetic component of ethanol aversion. *Structure* 5:701-711.
19. Yoshida A, Hsu LC, Davé V (1992) Retinal oxidation activity and biological role of human cytosolic aldehyde dehydrogenase. *Enzyme* 46:239-244.
20. Yoshida A, Hsu LC, Yanagawa Y (1993) Biological role of human cytosolic aldehyde dehydrogenase 1: hormonal response, retinal oxidation and implication in testicular feminization. *Adv Exp Med Biol* 328:37-44.
21. Duester G (2001) Genetic dissection of retinoid dehydrogenases. *Chem Biol Interact* 130-132:469-480.
22. Klyosov AA, Rashkovetsky LG, Tahir MK, Keung WM (1996) Possible role of liver cytosolic and mitochondrial aldehyde dehydrogenases in acetaldehyde metabolism. *Biochemistry* 35:4445-4456.
23. Holland LZ*, et al.* (2008) The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Res* 18:1100-1111.
24. Cañestro C, Yokoi H, Postlethwait JH (2007) Evolutionary developmental biology and genomics. *Nat Rev Genet* 8:932-942.
25. Schubert M, Escriva H, Xavier-Neto J, Laudet V (2006) Amphioxus and tunicates as evolutionary model systems. *Trends Ecol Evol* 21:269-277.
26. Putnam NH*, et al.* (2008) The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453:1064-1071.
27. Cañestro C, Postlethwait JH, Gonzàlez-Duarte R, Albalat R (2006) Is retinoic acid genetic machinery a chordate innovation? *Evol Dev* 8:394-406.
28. Skibbe DS*, et al.* (2002) Characterization of the aldehyde dehydrogenase gene families of *Zea mays* and *Arabidopsis*. *Plant Mol Biol* 48:751-764.
29. Perozich J, Nicholas H, Wang BC, Lindahl R, Hempel J (1999) Relationships within the aldehyde dehydrogenase extended family. *Protein Sci* 8:137-146.
30. Fujiwara S, Kawamura K (2003) Acquisition of retinoic acid signaling pathway and innovation of the chordate body plan. *Zool Sci* 20:809-818.
31. Niederreither K, McCaffery P, Dräger UC, Chambon P, Dollé P (1997) Restricted expression and retinoic acid-induced downregulation of the retinaldehyde dehydrogenase type 2 (RALDH-2) gene during mouse development. *Mech Dev* 62:67-78.
32. Bridgham JT, Ortlund EA, Thornton JW (2009) An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* 461:515-519.
33. Chen C-H*, et al.* (2008) Activation of aldehyde dehydrogenase-2 reduces ischemic damage to the heart. *Science* 321:1493-1495.
34. Carroll SB (2008) Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134:25-36.
35. Hoekstra HE, Coyne JA (2007) The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 61:995-1016.
36. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574.
37. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688-2690.
38. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555-556.

39. Petrey D*, et al.* (2003) Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling. *Proteins* 53 Suppl 6:430-435.
40. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14:33-38, 27-38.
41. Yu J-K*, et al.* (2007) Axial patterning in cephalochordates and the evolution of the organizer. *Nature* 445:613-617.
42. Satou Y*, et al.* (2002) A cDNA resource from the basal chordate *Ciona intestinalis*. *Genesis* 33:153-154.
43. Holland LZ, Holland PWH, Holland ND (1996) Revealing homologies between body parts of distantly related animals by in situ hybridization to developmental stages: amphioxus versus vertebrates. *Molecular zoology: advances, strategies, and protocols*, eds Ferraris JD & Palumbi SR (Wiley-Liss, New York), pp 267-282; 473-483.
44. Christiaen L*, et al.* (2008) The transcription/migration interface in heart precursors of *Ciona intestinalis*. *Science* 320:1349-1352.

Supplementary Materials and Methods

ALDH sequences

The following genomes were examined for ALDH sequences: *Aedes aegypti*, *Anopheles gambiae*, *Bos taurus*, *Branchiostoma floridae*, *Caenorhabditis elegans*, *Canis familiaris*, *Ciona intestinalis*, *Ciona savignyi*, *Drosophila melanogaster*, *Gallus gallus*, *Gasterosteus aculeatus*, *Homo sapiens*, *Macaca mulatta*, *Monodelphis domestica*, *Mus musculus*, *Nematostella vectensis*, *Ornithorhyncus anatinus*, *Oryzias latipes*, *Pan troglodytes*, *Phanerochaete chrysosporium*, *Phycomyces blakesleeanus*, *Populus trichocarpa*, *Rattus norvegicus*, *Saccharomyces cerevisiae*, *Strongylocentrotus purpuratus*, *Takifugu rubripes*, *Tetraodon nigroviridis*, and *Xenopus tropicalis*. To these we added selected sequences from *Saccoglossus kowalevskii* EST and trace archives as well as ALDH sequences from various other animals (*Apis mellifera*, *Bombyx mori*, *Drosophila pseudoobscura*, *Macaca fasciculata*, *Mesocricetus auratus*, *Oryctolagus cuniculus*, *Ovis aries*, *Pongo pygmaeus*, *Taeniopygia guttata*, *Tribolium castaneum*, *Xenopus laevis*), plants (*Arabidopsis thaliana*, *Nicotiana tabacum*, *Oryza sativa*, *Populus trichocarpa*, *Secale cereale*, *Sorghum bicolor*, *Zea mays*), and fungi (*Chaetomium globosum*, *Cordyceps bassiana*, *Emericella nidulans*, *Magnaporthe grisea*, *Phaeosphaeria nodorum*, *Phanerochaete chrysosporium*, *Phycomyces blakesleeanus*). To facilitate structural comparisons with published work, numbering of ALDH amino acid (aa) residues was based on the classical numbering of residues in the mature human ALDH2 enzyme, which places the catalytic Cys at aa position 302 (1).

Large-scale phylogenetic analyses

ALDH sequences were obtained from www.aldh.org, genomic data, and trace archives using both signature (InterPro IPR002086) (2) and BLAST searches (3). ALDH aa sequences were aligned using MUSCLE (4) followed by manual refinement. Bayesian inference (BI) was carried out using MrBayes 3.1 (5) with a WAG+I+Γ_4 model predicted by ProtTest (6). Two runs of 5,000,000 generations were computed for each tree. Convergence was verified and the burn-in period determined by plotting log likelihood versus time. Consensus trees and posterior probabilities were calculated using the 50% majority rule. A maximum likelihood (ML) analysis was performed using RAxML-VI-HPC with the PROTMIXWAG parameter (7) and with 100 bootstrap pseudo-replicates to assess node support. Phylogenetic analyses carried out with BI and ML resulted in essentially identical tree topologies.

ALDH signatures

ALDH1 and ALDH2 sequence signatures were obtained by aligning human ALDH1A2 and human ALDH2 and confirmed in an exhaustive alignment containing vertebrate ALDH1 and ALDH2 sequences using MultAlin (bioinfo.genopole-toulouse.prd.fr/multalin/multalin.html). The individual aa frequencies were obtained from weblogo.berkeley.edu.

Ancestral sequence reconstruction

Ancestral protein sequences at internal nodes of the ALDH1/2 phylogeny were reconstructed using the program PAML 3.15 (8) assuming a

WAG+Γ_4 model for a data matrix of 447 aa residues. To evaluate and limit the influence of fast evolving sequences on the ancestral sequence reconstruction, five separate data sets were analyzed (one including all available ALDH1/2 sequences, one excluding all sequences with long branches, one excluding all protostome sequences, one excluding all lophotrochozoan sequences, and one excluding all tunicate sequences). For each analysis, a corresponding ML tree was calculated using RAxML-VI-HPC with the PROTMIXWAG (7) with 100 bootstrap replicates to assess node support, which served as input tree for the ancestral sequence calculation. These control calculations, based on varying taxonomic sampling, successfully tested the robustness of the ancestral sequence reconstruction approach as well as of the subsequent ancestral channel modeling and ancestral channel volume calculation.

3D modeling

ALDH1 and ALDH2 structures were modeled by homology using the 3D structure of sheep ALDH1A1 (PDB code 1BXS) or of human ALDH2 in complex with the dipsogenic inhibitor daidzin (PDB code 1OF7). Each target sequence was globally aligned with the template using ClustalW. This alignment was used as input data for the modeling program Nest in Jackal (9). Parameters were set for refinement in all loops and secondary structure regions. The stereochemical quality of the 3D models was evaluated using Procheck (10). Retinal binding to the substrate access channel was carried out by anchoring the aldehyde structure in a position equivalent to that of daidzin. The 3D structure of retinal was obtained from MSD Ligand Chemistry (www.ebi.ac.uk/msd-srv/msdchem). Graphical analyses were performed with VMD (Visual Molecular Dynamics) (11).

Cluster analysis

We have compiled data on 202 ALDH1/2 volumes derived from 3D structure studies. TwoStep Cluster algorithm as implemented in SPSS 13.0 (SPSS for Windows, Rel. 13.0.2004, SPSS Inc., Chicago, USA) was used both to define the best number of clusters that accommodate the empirical data and to classify individual ALDH1/2s into their respective volume clusters. The best number of clusters was chosen as a function of the smallest Schwartz Bayesian Criterion (BIC).

Retinal binding

Retinal binding to the substrate access channel was carried out anchoring the aldehyde structure in a position equivalent to daidzin. The retinal 3D structure was obtained from MSD Ligand Chemistry (www.ebi.ac.uk/msd-srv/msdchem). The retinal atoms O1, C15, C14, C13, C12, C11, and C10 were superposed on the daidzin atoms O34, C31, C30, C29, C28, C26, and C25, respectively.

Substrate access channel volume calculations

Aldehyde dehydrogenase (ALDH) models were individually placed inside a grid box, in which x, y, and z coordinates were spaced by 0.8 Å. Void regions inside this box were determined by sequentially moving a probe molecule with a 1.4 Å radius through all grid points. At each point, the volume

occupied by the probe plus an average atomic radius of 1.6 Å was scanned. If no protein atoms were detected, the point was considered to be in a void region. Multiple cavities were detected and to isolate the substrate access channel from other spaces, the structure of human ALDH2 in complex with the dipsogenic inhibitor daidzin (PDB code 1OF7) was superposed on the boxed model. The void region in the boxed ALDH model matching the localization of daidzin inside the ALDH2 channel was isolated for volume calculation. The total cavity volume was determined as a sum of all distinct volume elements within the isolated void region. Each volume element was a cube defined by eight points enclosing a volume of 0.512 $Å^3$. Modeled cavities were manually inspected to remove subsidiary spaces not associated with the substrate access channel. The Van der Waals volumes for all aa have previously been described (12).

Statistical analyses

Substrate access channel volumes were analyzed in unpaired sets of eukaryote ALDH1/2 and ALDH1L enzymes and in paired sets of vertebrate ALDH1/2s (ALDH1As, ALDH1B1s, and ALDH2s) by non-parametric ANOVA using the Kruskal-Wallis test and Dunn's Multiple Comparisons test.

Gene expression studies

Sexually mature adults of the Florida amphioxus (*B. floridae*) were collected by shovel and sieve in Tampa Bay, Florida. Gametes were obtained by electric stimulation of adults, and after fertilization embryos and larvae were raised in the laboratory and fixed at different developmental stages. Five cDNA libraries were used for cloning ALDH1 and ALDH2 genes from amphioxus. The cDNA libraries were made from RNA of unfertilized eggs, gastrulae, neurulae, early larvae, and mature adults (13). In situ hybridization experiments were carried out as previously described (14) using increased hybridization temperatures (65 °C to 70 °C) and non-conserved regions of the different genes as templates for antisense riboprobe synthesis to ensure probe specificity. Control experiments were carried out with sense riboprobes to verify the specificity of the expression patterns. After in situ hybridization, embryos were mounted on glass slides, analyzed under the microscope, and photographed as whole mounts.

C. intestinalis adults were obtained from M-Rep (San Diego, USA). Embryos and sperm were surgically removed and kept separate until in vitro fertilization. Fertilized eggs were dechorionated as described (15). Embryos were raised at 18 °C on gelatin/formaldehyde-coated dishes in artificial seawater. Embryos were collected and fixed in 4% paraformaldehyde at various developmental stages (5 to 8 hours, 10 to 12 hours). Template DNA plasmids were retrieved from the *C. intestinalis* Gene Collection Release 1 library (16). Digoxigenin-labeled antisense RNA probes for ALDH1a, b, and d, and ALDH2 were synthesized from clones GC38i10, GC38i10, GC30b02, and GC07d10, respectively. The ALDH1c probe was generated using PCR amplification of the last exon from *C. intestinalis* genomic DNA. Whole mount in situ hybridization was essentially performed as described (17) using an increased hybridization temperature of 65 °C. Embryos were mounted in Permount on glass slides and expression was analyzed using a Leica DMR microscope.

Supplementary References

1. Moore SA*, et al.* (1998) Sheep liver cytosolic aldehyde dehydrogenase: the structure reveals the basis for the retinal specificity of class 1 aldehyde dehydrogenases. *Structure* 6:1541-1551.
2. Mulder NJ*, et al.* (2003) The InterPro Database, 2003 brings increased coverage and new features. *Nucleic Acids Res* 31:315-318.
3. Altschul SF*, et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389-3402.
4. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792-1797.
5. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574.
6. Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104-2105.
7. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688-2690.
8. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555-556.
9. Petrey D*, et al.* (2003) Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling. *Proteins* 53 Suppl 6:430-435.
10. Laskowski RA, Moss DS, Thornton JM (1993) Main-chain bond lengths and bond angles in protein structures. *J Mol Biol* 231:1049-1067.
11. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14:33-38, 27-38.
12. Richards FM (1974) The interpretation of protein structures: total volume, group volume distributions and packing density. *J Mol Biol* 82:1-14.
13. Yu J-K*, et al.* (2007) Axial patterning in cephalochordates and the evolution of the organizer. *Nature* 445:613-617.
14. Holland LZ, Holland PWH, Holland ND (1996) Revealing homologies between body parts of distantly related animals by in situ hybridization to developmental stages: amphioxus versus vertebrates. *Molecular zoology: advances, strategies, and protocols*, eds Ferraris JD & Palumbi SR (Wiley-Liss, New York), pp 267-282; 473-483.
15. Mita-Miyazawa I, Ikegami S, Satoh N (1985) Histospecific acetylcholinesterase development in the presumptive muscle cells isolated from 16-cell-stage ascidian embryos with respect to the number of DNA replications. *J Embryol Exp Morphol* 87:1-12.
16. Satou Y*, et al.* (2002) A cDNA resource from the basal chordate *Ciona intestinalis*. *Genesis* 33:153-154.
17. Christiaen L*, et al.* (2008) The transcription/migration interface in heart precursors of *Ciona intestinalis*. *Science* 320:1349-1352.

Table S1. Sequence signatures of vertebrate ALDH1 and ALDH2 enzymes. Signatures are displayed as motifs with four to ten amino acids (aa). Motifs are shown as pictograms with relative frequencies symbolized by aa height as determined in an alignment of 24 vertebrate ALDH1s and 12 vertebrate ALDH2s. Represented here are only aa that effectively distinguish vertebrate ALDH1s from ALDH2s. Species-specific and vertebrate ALDH1A paralog-specific signatures are not included. The aa that are not conserved between ALDH1s and ALDH2s are surrounded by a red box and their localization along with the nature of the substitutions between ALDH1s and ALDH2s is displayed. The ability of each vertebrate signature to distinguish ALDH1s from ALDH2s in invertebrate chordates (amphioxus and ascidian tunicates) is also shown.

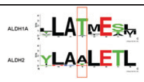
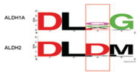
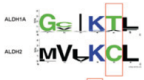
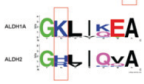
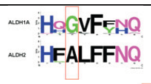
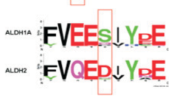
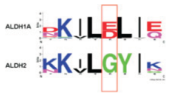
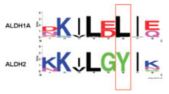
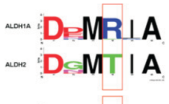
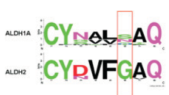
Amino acid substitution	Vertebrate signature	Cephalochordate signature	Tunicate signature	Type of substitution	Structural location of the signature
Thr-104-Ala		No	No	Polar to small aliphatic apolar	Exposed surface (α helix-B)
Gly-110-Asn		No	Yes	Small aliphatic apolar to polar	Buried residue (α helix-B)
Asn-116-Ile		No	Yes	Polar to aliphatic apolar	Buried residue (α helix-C)
Gly-123-Asp		No	No	Polar to negatively charged	Channel (α helix-C)
Gly-124-Met		No	No	Small aliphatic apolar to large aliphatic	Channel mouth (α helix-C)
Thr-128-Cys		Yes	Yes	Polar to polar thiol-containing	Buried residue (α helix-C)
Phe-175-Gln		Yes	Yes	Bulky aromatic to polar	Buried residue (α helix-D)
Ser-184-Ala		No	No	Polar to small aliphatic apolar	Buried residue (α helix-D)
Cys-185-Thr		Yes	Yes	Thiol-containing to polar	Buried residue (α helix-D)
Thr-188-Val		Yes	Yes	Polar to aliphatic apolar	Buried residue (β sheet-8)
Val-191-Met		Yes	Yes	Aliphatic apolar to large aliphatic apolar	Buried residue (β sheet-8)
Pro-193-Val		Yes	Yes	Apolar cyclic to aliphatic apolar	Buried residue (β sheet-8)
Lys-251-Arg/His		No	No	Positively charged to larger positively charged imidazole	Channel (α helix-G)
Glu-255-Val		No	No	Negatively charged to aliphatic apolar	Oligomerization surface (α helix-G)
Ala-279-Ser		No	Yes	Small aliphatic apolar to polar	Exposed surface (connecting loop between β sheet-12 and α helix-H)
Asn-285-Trp		No	No	Polar to large aromatic	Exposed surface (α helix-H)
Gln-292-Phe		Yes	Yes	Polar to aromatic	Exposed surface (α helix-H)

Table S1. Continued.

Amino acid substitution	Vertebrate signature	Cephalochordate signature	Tunicate signature	Type of substitution	Structural location of the signature
Gly-293-Ala		No	No	Small aliphatic apolar to smaller aliphatic apolar	Buried residue (α helix-H)
Ile-303-Cys		No	No	Aliphatic apolar to thiol-containing	Channel bottom (connecting loop between α helix-H and β sheet-13)
Glu-311-Gln		No	No	Negatively charged to polar	Exposed surface (β sheet-13)
Ser-313-Asp		No	No	Polar to negatively charged	Exposed surface (connecting loop between β sheet-13 and α helix-I)
Arg-320-Glu		No	No	Positively charged to negatively charged	Exposed surface (α helix-I)
Asp-355-Gly		No	No	Negatively charged to small aliphatic apolar	Exposed surface (α helix-J)
Leu-356-Tyr		No	No	Aliphatic apolar to large aromatic	Exposed surface (α helix-J)
Glu-368-Leu		No	Yes	Negatively charged to large aliphatic apolar	Buried residue (β sheet-16)
Ser-387-Gly		No	Yes	Polar to small aliphatic apolar	Exposed residue (β sheet-14)
Arg-394-Thr		No	No	Positively charged to polar	Exposed residue (α helix-K)
Gln-405-Met		No	No	Polar to large aliphatic apolar	Buried residue (β sheet-16)
Ile/Val/Leu-440-Asp		No	No	Aliphatic apolar to negatively charged	Oligomerization surface (α helix-M)
Thr-441-Tyr		Yes	No	Polar to large aromatic	Exposed surface (α helix-M)
Ser-444-Gln		No	No	Polar to larger polar	Exposed surface (α helix-M)
Ser-457-Asp		No	No	Polar to negatively charged	Channel (connecting loop between β sheet-18 and α helix-N)
Val-459-Phe		Yes	No	Aliphatic apolar to aromatic	Channel neck (connecting loop between β sheet-18 and α helix-N)
Ser-460-Gly		No	No	Polar to small aliphatic apolar	Exposed surface (connecting loop between β sheet-18 and α helix-N)

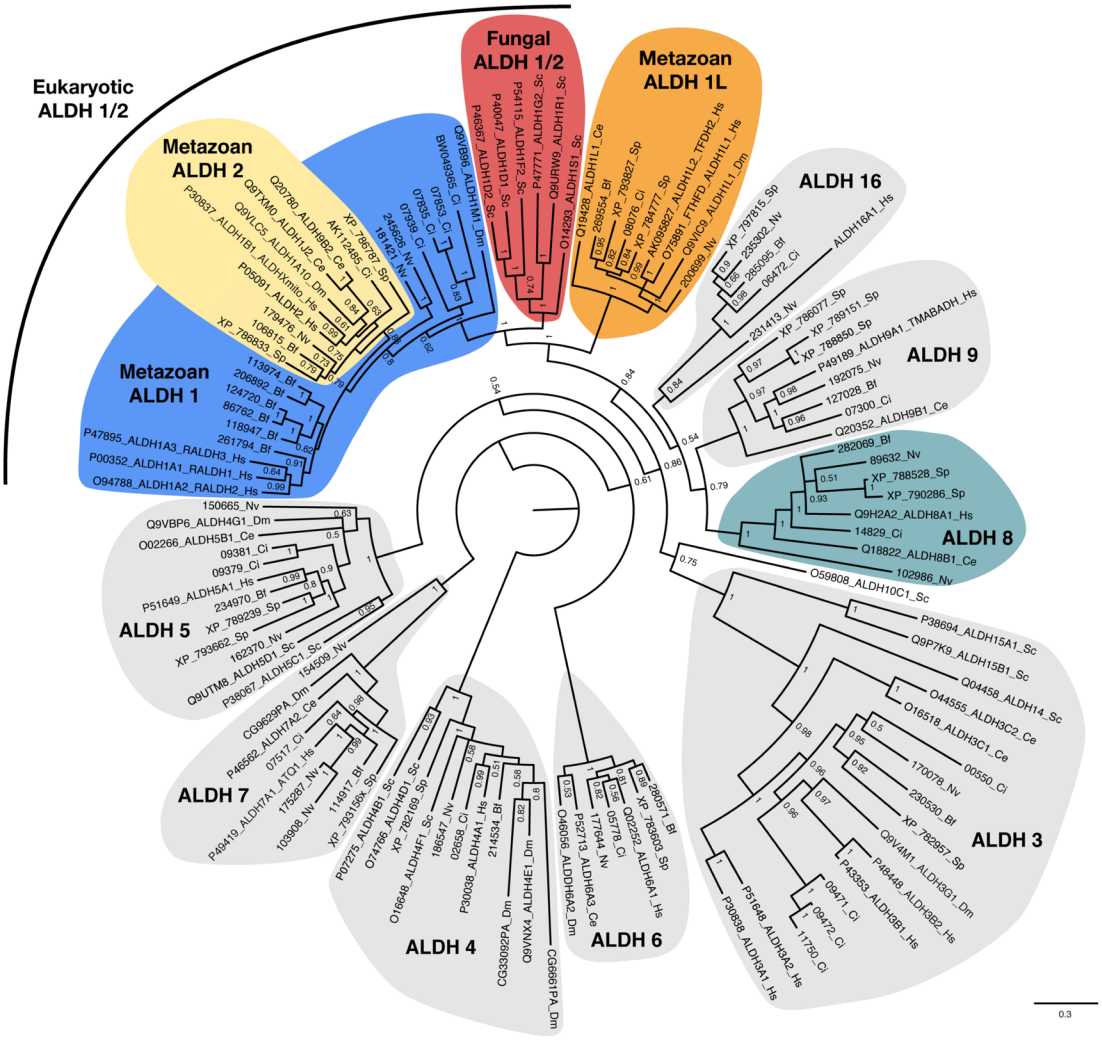


Fig. S1. Phylogram of a comprehensive phylogenetic analysis of the aldehyde dehydrogenase (ALDH) superfamily. Posterior probabilities are indicated at each node. Each individual ALDH family is supported by significant posterior probabilities (>0.95). Abbreviations: Bf, *Branchiostoma floridae*; Ce, *Caenorhabditis elegans*; Ci, *Ciona intestinalis*; Dm, *Drosophila melanogaster*; Hs, *Homo sapiens*; Nv, *Nematostella vectensis*; Sc, *Saccharomyces cerevisiae*; Sp, *Strongylocentrotus purpuratus*.

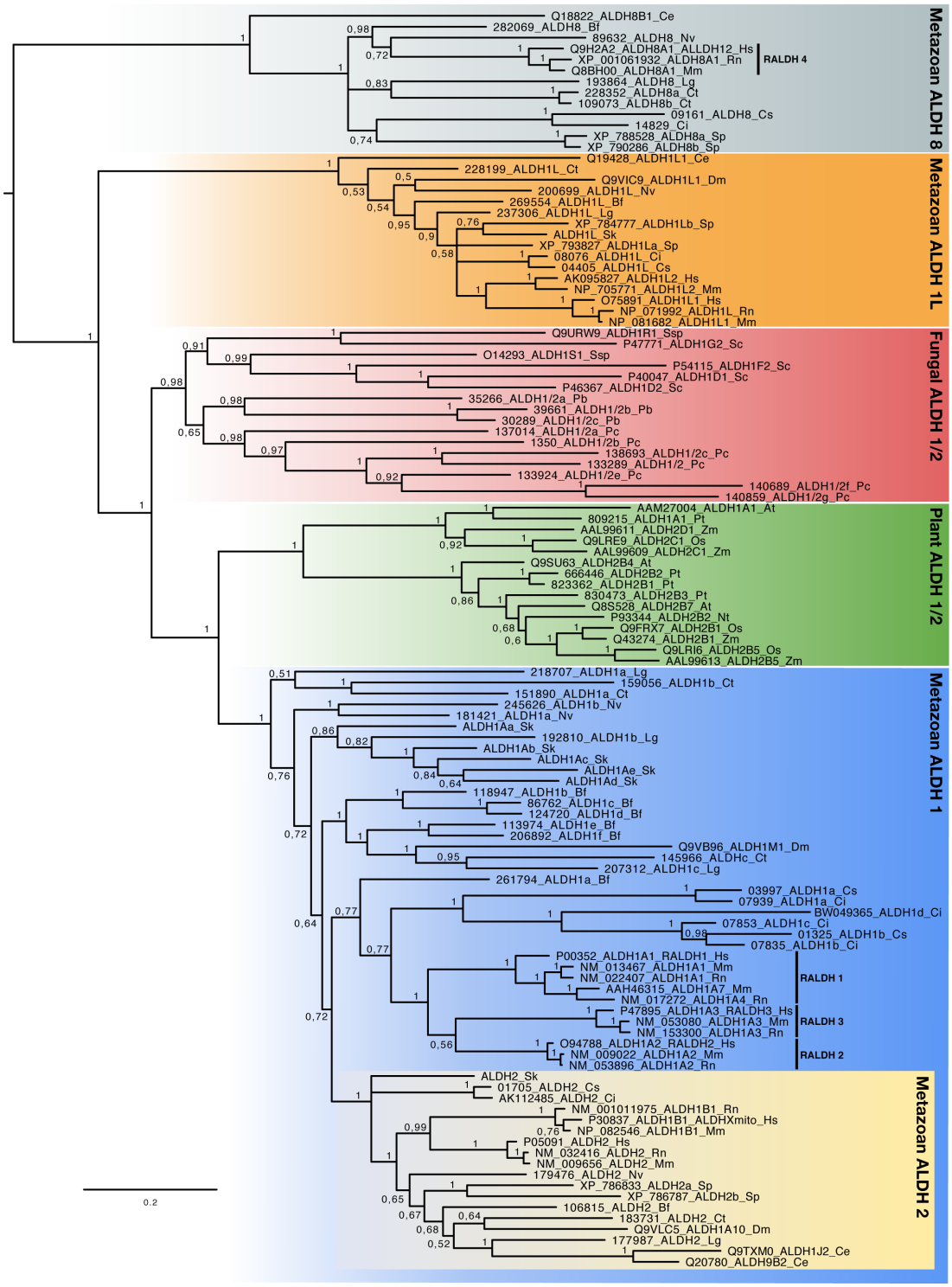


Fig. S2. Phylogram of a phylogenetic analysis of the ALDH1/2, ALDH1L, and ALDH8 families with the latter as outgroup. Posterior probabilities are indicated at each node. Each individual ALDH family is supported by significant posterior probabilities (>0.95). Abbreviations: At, *Arabidopsis thaliana*; Bf, *Branchiostoma floridae*; Ct, *Capitella teleta*; Ce, *Caenorhabditis elegans*; Ci, *Ciona intestinalis*; Cs, *Ciona savignyi*; Dm, *Drosophila melanogaster*; Hs, *Homo sapiens*; Lg, *Lottia gigantea*; Mm, *Mus musculus*; Nv, *Nematostella vectensis*; Nt, *Nicotiana tabacum*; Os, *Oryza sativa*; Pc, *Phanerochaete chrysosporium*; Pb, *Phycomyces blakesleeanus*; Pt, *Populus trichocarpa*; Rn, *Rattus norvegicus*; Sc, *Saccharomyces cerevisiae*; Sk, *Saccoglossus kowalevskii*; Sp, *Strongylocentrotus purpuratus*; Ssp, *Schizosaccharomyces pombe*; Zm, *Zea mays*.

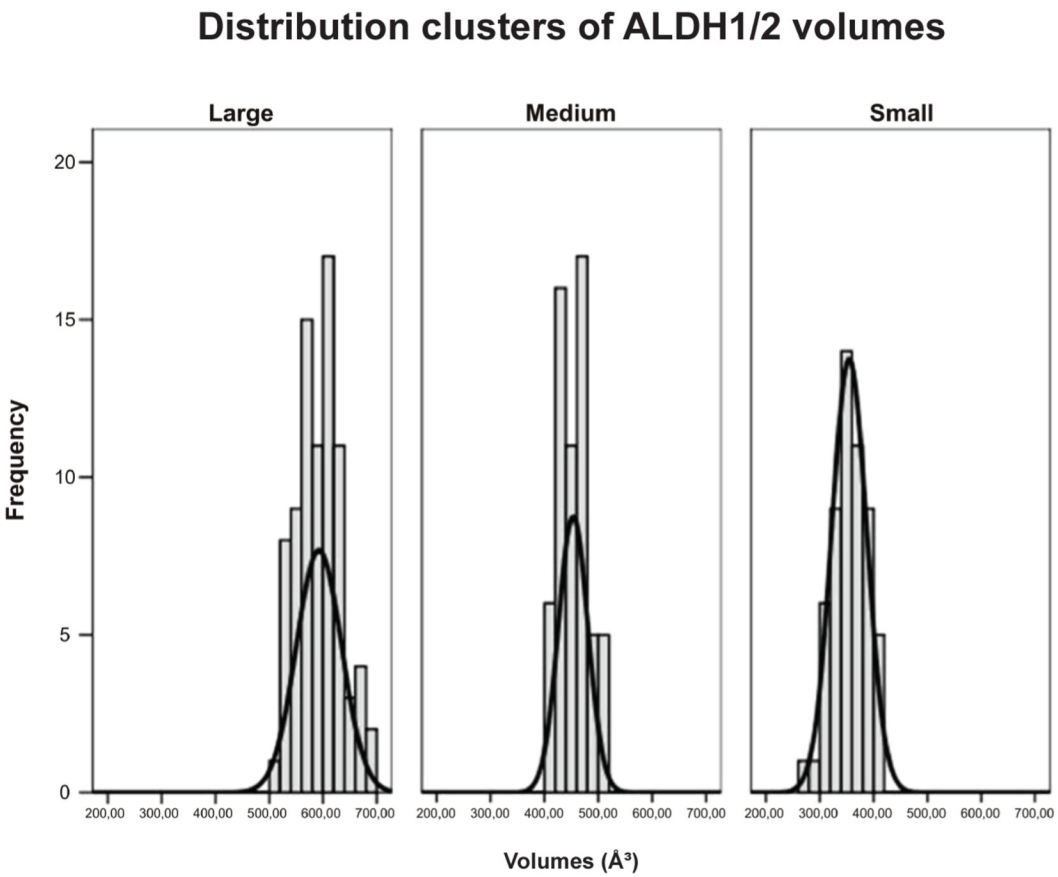


Fig. S3. ALDH1/2 clusters according to substrate access channel volume distribution. Average and standard deviation values for the three clusters are: 592.29 ± 42.07 for large channels, 453.33 ± 27.39 for medium channels, and 354.79 ± 32.50 for small channels.

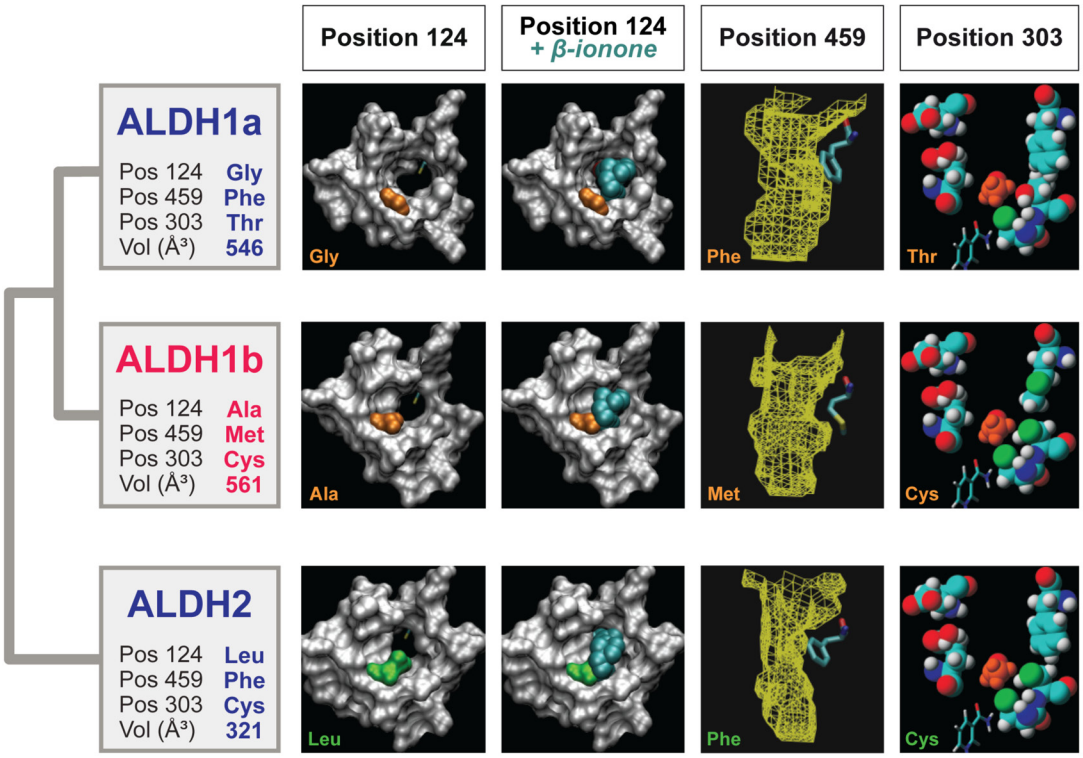


Fig. S4. Phylogenetic and structural analysis of ALDH1/2s from the ascidian tunicate *Ciona savignyi*. The two ALDH1 duplicates from *C. savignyi* display heterogeneous properties at their substrate entry channels. Although ALDH1a and b both display bulky amino acids (aa) at position 459, corresponding to the channel neck, ALDH1a displays the smallest aa (Gly) at position 124, which corresponds to the channel mouth. In contrast, in ALDH1b this residue is substituted by the slightly larger Ala, which creates an obstacle to accommodate the large β-ionone moiety of retinaldehyde. Moreover, in ALDH1a the aa at position 303 is Thr, which is typical for ALDH1s, while in ALDH1b this aa is Cys, which is typical for ALDH2s. This suggests that in *C. savignyi* ALDH1a is best adapted for retinoic acid synthesis and that ALDH1b incorporated structural features reminiscent of ALDH2s, which process smaller toxic aldehydes.

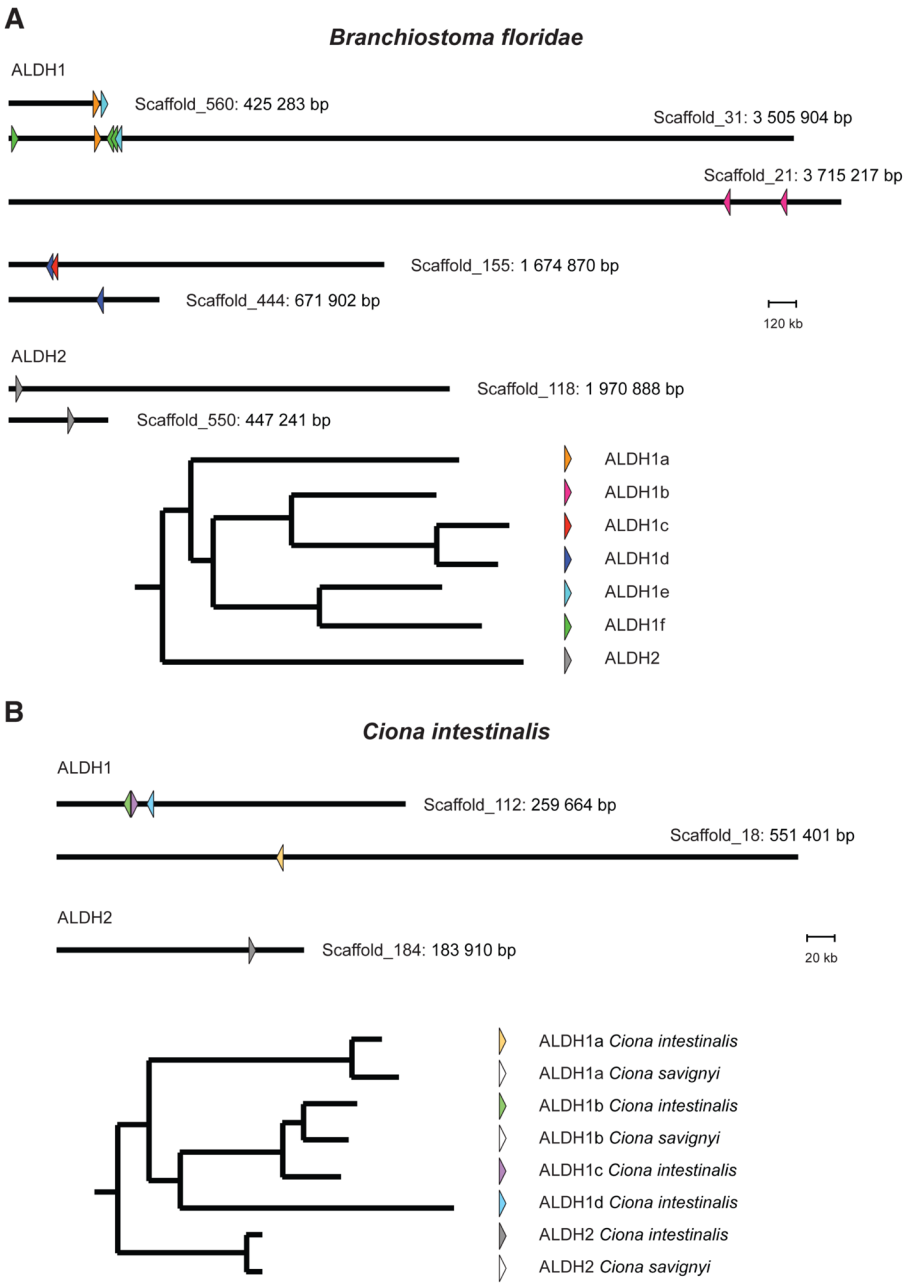


Fig. S5. Genomic linkage of ALDH1 and ALDH2 genes in the cephalochordate amphioxus and in the ascidian tunicate *Ciona intestinalis*. The position of predicted ALDH1 and ALDH2 sequences on assembled scaffolds of the amphioxus (*A*) and the *C. intestinalis* (*B*) genome are shown. (*A*) The phylogenetic relationship between the six lineage-specific amphioxus ALDH1 duplicates and the single amphioxus ALDH2 is indicated. For ALDH2, the analysis identified two copies in the amphioxus genome, one on scaffold 550 and one on scaffold 118, representing a single copy of amphioxus ALDH2 on each of the two homologous chromosomes. For the amphioxus ALDH1 duplicates, the analysis suggests that at least some genes, like ALDH1c and d (on scaffold 155), ALDH1a and e (on scaffold 560) or ALDH1a, e, and f (on scaffold 31) are located on the same scaffold and might hence be linked on the chromosome. In sum, the phylogenetic and synteny data suggest that the six amphioxus-specific ALDH1 duplicates evolved by tandem duplications from an ALDH1a-like ancestor. (*B*) The phylogenetic relationship between *C. intestinalis* and *C. savignyi* ALDH1s and ALDH2s is indicated. In the *C. intestinalis* genome, ALDH2 and ALDH1a are found on individual scaffolds (on scaffold 184 and 18, respectively), while the other three *C. intestinalis* ALDH1 duplicates, ALDH1b, c, and d, are clustered on a single scaffold (on scaffold 112) suggesting that these three genes are linked on the same chromosome and might have originated by tandem duplications.

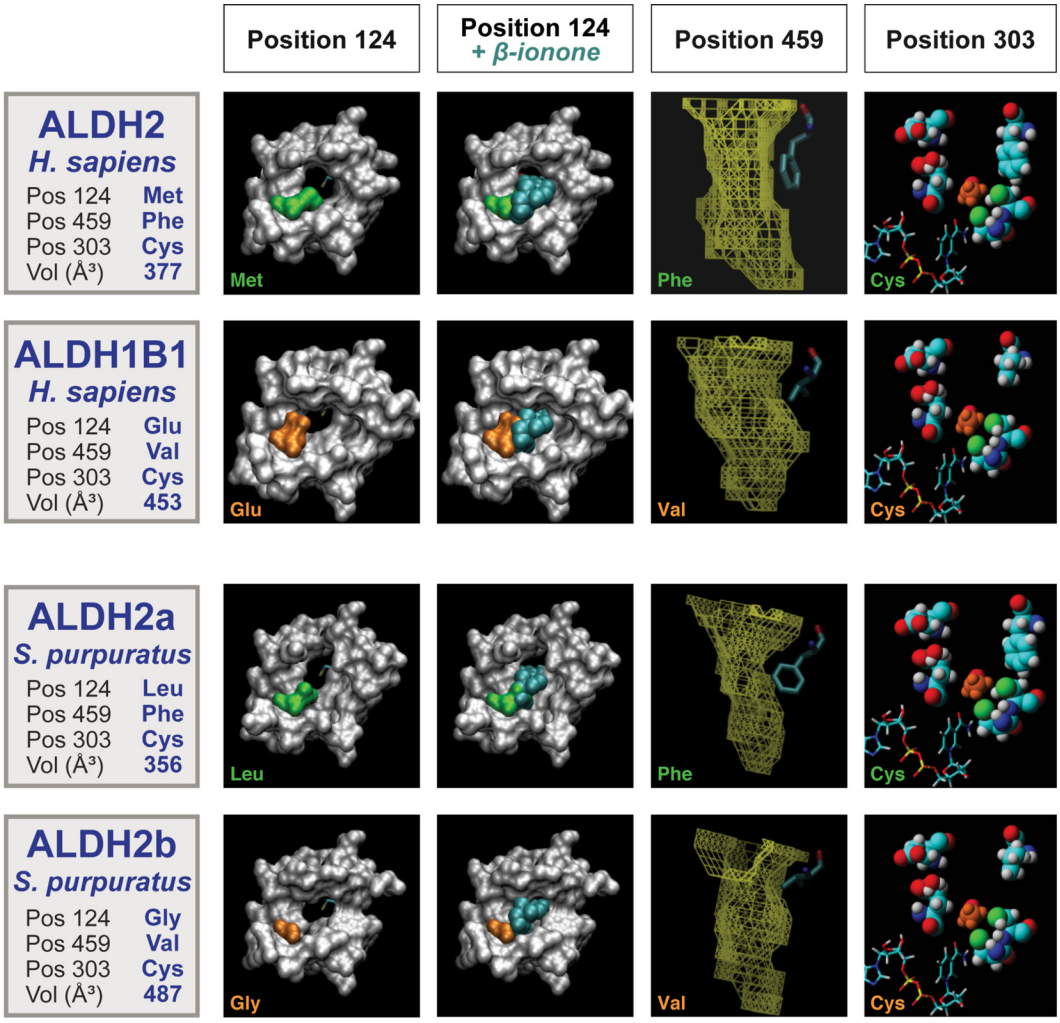


Fig. S6. Plasticity of the substrate access channel in deuterostomes. Vertebrate and echinoderm ALDH2s are examples for the evolutionary plasticity of ALDH1/2 channels. In vertebrates, the ALDH2s have duplicated into the ALDH2s and the ALDH1B1s. Vertebrate ALDH2s have conserved the amino acid (aa) signatures (Met124, Phe459) for shaping a narrow entrance (the mouth) and a tight-fitting proximal third (the neck) of the substrate access channel. The ALDH1B1s incorporated smaller Glu124 and Val459, which offer less resistance for accommodating the β-ionone moiety of retinaldehyde and alleviate constriction at the channel neck hence increasing the overall channel volumes (362 ± 30 $Å^3$ versus 451 ± 30 $Å^3$ for ALDH2 and ALDH1B1, respectively, $p<0,05$). A similar trend is observed for sea urchin (*Strongylocentrotus purpuratus*) ALDH2s. *S. purpuratus* ALDH2a displays the typical ALDH2 pattern with bulky Leu124 and Phe459 at the channel mouth and neck, respectively. In contrast, *S. purpuratus* ALDH2b incorporated small aa (Gly124 and Val459), which leave the channel mouth wide open to accommodate the β-ionone moiety of retinaldehyde and alleviate constriction at the channel neck thus increasing the volume from 356 $Å^3$ in ALDH2a to 487 $Å^3$ in ALDH2b.

Future Directions

RAvolution: the (R)evolutionary History of Morphogen-Dependent Signaling

Michael Schubert

Proposal Abstract

Extensive research carried out over the last 100 years has established that retinoids, which constitute a group of fat-soluble morphogens related to retinol (vitamin A), play crucial roles in early development, organogenesis, tissue homeostasis, proliferation, differentiation, and apoptosis. In vertebrates, most, but not all, retinoid functions are mediated by retinoic acid (RA) binding to heterodimers of two nuclear receptors: retinoic acid receptor (RAR) and retinoid X receptor (RXR). Retinoid signaling was long thought to be vertebrate-specific, but developmental studies in invertebrate chordates have revealed roles for retinoids that are conserved in all chordates. Outside chordates, however, evidence for functional roles of retinoids and of the RAR/RXR heterodimer remains scarce, although recent analyses have implicated retinoid-dependent signaling in embryogenesis of echinoderms and mollusks, in tissue regeneration of insects and crustaceans and in cell proliferation and neuronal differentiation of cnidarians. Taking into consideration that the genomes of insects, crustaceans, and cnidarians do not encode a RAR gene, these data suggest that roles for retinoids as morphogens originated before a functional RAR/RXR heterodimer and, given the importance of the RAR/RXR heterodimer for mediating retinoid signaling in chordates, evoke the question of the evolutionary origins of the ligand binding capacities of RAR and RXR. This project thus proposes to address these points by studying retinoid signaling in five distinct invertebrate taxa. The experiments are designed to independently characterize the roles of retinoids and of RAR and RXR, which will allow a detailed assessment of the functions of retinoids versus the functions of RAR and RXR. Altogether, these data will reveal the evolutionary diversification of the retinoid signaling cascade, hence creating an important source of information illuminating the complexity of the vertebrate (and human) retinoid system.

State of the Art

The retinoids constitute a group of fat-soluble morphogens related to retinol, which is more commonly known as vitamin A. In the course of the last century, basic and clinical research has established that vitamin A is essential for the maintenance of various tissues and organs including skin, bone and vasculature and also promotes vision and immune functions (1,2). Vitamin A derivatives (i.e. retinoids) are equally important for human reproduction and embryonic development (2-5). In the body, vitamin A is converted to more active retinoids, mainly to the potent morphogen retinoic acid (RA), through which this vitamin exerts its multiple effects during development, organogenesis, tissue homeostasis, cell proliferation, differentiation, and apoptosis (5-8).

Due to the anti-proliferation and pro-differentiation properties of retinoids, their therapeutic value has equally been the subject of intensive study (9). In 1982, the retinoid 13-*cis* RA (also called isotretinoin) was licensed under the name Accutane® for treatment of severe, cystic acne. Taken in early pregnancy, Accutane® also has strong teratogenic effects on the developing human fetus resulting in high percentages of spontaneous abortions and embryonic malformations (known as RA embryopathy). This retinoid-induced phenotype includes craniofacial malformations as well as abnormal central nervous system, heart and thymus development (10).

The roles of vitamin A derivatives as morphogens during development have mainly been investigated in vertebrates (chiefly in rodents, chicken and quail, frogs and fish) (3-5,11). While vitamin A deficiency models and pharmacological approaches, such as treatment with vitamin A or RA, were initially used to study retinoid signaling, the development of genetic tools, such as targeted gene inactivation or gene overexpression, has allowed a rather detailed assessment of the functions of retinoid morphogens during vertebrate embryogenesis. The vertebrate embryo thus appears to be subdivided into several modules each with a different responsiveness to retinoids, which depends on the presence of local sources and sinks of RA and on the sensitivity of a given tissue to RA, which in turn varies both temporally and spatially during development (7).

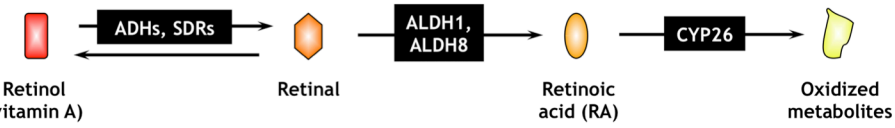


Fig. 1. Retinoic acid synthesis and degradation. ADH, alcohol dehydrogenase; ALDH, aldehyde dehydrogenase, CYP26, cytochrome P450, subfamily CYP26; RA, retinoic acid; SDR, short-chain dehydrogenase/reductase.

During vertebrate development, RA, as the main mediator of retinoid signaling, is synthesized from vitamin A in two steps of NAD-dependent oxidation (Fig. 1) (5,7,8,12). The first step is the reversible oxidation of vitamin A to retinal by either alcohol dehydrogenases (ADHs) or short-chain dehydrogenases/reductases (SDRs) (13-15). The second step is the irreversible oxidation of retinal to RA by enzymes of the aldehyde dehydrogenase family (ALDH) (16). This conversion is mainly carried out by ALDH1A1, 2, 3 and ALDH8 enzymes, also known as, respectively, RALDH1,

2, 3 and 4 (for retinaldehyde dehydrogenases) (17-19). RALDH-dependent RA synthesis plays crucial roles during embryonic development. For example, loss of *raldh2* in mice results in severe developmental defects that are almost completely rescued by maternal RA administration (20). The availability of endogenous RA is also controlled through its degradation by proteins of the cytochrome P450 family, chiefly CYP26 (21). These enzymes catalyze the oxidation of RA into a wide variety of metabolites, such as 4-oxo-RA, 4-OH-RA or 18-OH-RA. Whether these metabolites retain a biological activity as morphogens is still controversial (22,23).

Until very recently, the main dogma of retinoid research stated that RA functions are mediated by heterodimers of two nuclear receptors: the retinoic acid receptor (RAR) and the retinoid X receptor (RXR) (Fig. 2) (24). RAR/RXR heterodimers bind to specific DNA elements in the regulatory regions of target genes, called RA response elements (RAREs). Most RAREs consist of two direct repeats (DRs) with the canonical nucleotide sequence (A/G)G(G/T)TCA separated by a variable number of nucleotide spacers (usually either 1, 2 or 5 nucleotides) (25-27). Fixation of RA activates the ligand-dependent transcription factor function of the RAR/RXR heterodimer and leads to the binding of a coactivator complex to the heterodimer (24). *In vitro,* RAR can bind both all-*trans* and 9-*cis* RA, whereas RXR only binds 9-*cis* RA. However, it is still unclear, whether 9-*cis* RA is actually present in vertebrate embryos and adults and, thus, whether this compound and its binding to RXR have a biological function in vertebrates (28,29).

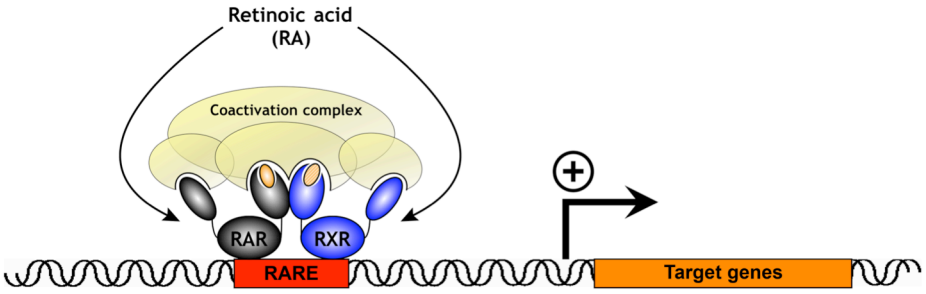


Fig. 2. Classical retinoid signaling. RA, retinoic acid; RAR, retinoic acid receptor; RARE, retinoic acid response element; RXR, retinoid X receptor.

In addition to this classical mode of retinoid action, mediated by RAR/RXR heterodimers, several alternative mechanisms for triggering retinoid-dependent signaling cascades in vertebrates have recently been proposed. Some of these non-canonical signaling cascades are apparently controlled by other nuclear receptors, such as PPARβ/δ, RORβ or COUP-TF, all three of which are bound and activated by retinoids (30-32). Moreover, members of completely different protein families have also been suggested to convey retinoid signals. For example, RA can directly activate CREB, the cAMP response element-binding protein, in a RAR/RXR-independent manner (33). Moreover, RA has been described as an important modulator of PKCα (a protein kinase C) activity: PKCα apparently contains a RA binding site and the RA-dependent modulation of PKCα activity is controlled by competitive binding of PKCα to all-*trans* RA and to acidic phospholipids (34). Albeit intriguing, the biological roles during vertebrate development of these non-canonical retinoid signaling mechanisms remain to be determined.

Over the last few decades, more than 500 vertebrate genes have been put forth as *in vivo* regulatory targets of the morphogen RA. In some cases, direct regulation, driven for example by liganded RAR/RXR heterodimers bound to RAREs, was demonstrated, but in most cases, regulation of the proposed gene target was indirect thus occurring through intermediate transcription factors. Of the target genes suggested, 27 were found to be direct targets of the classical RAR/RXR-dependent RARE pathway and another 100 are good candidates for being direct targets (2). More recently, genomic RARα and RARγ targets were identified using a chromatin immunoprecipitation on chip (ChIP on chip) approach in MCF-7 breast cancer cells (35). Although the relevance of this *in cellulo* study for the overall composition of the retinoid signaling cascade *in vivo* remains to be demonstrated, this ChIP on chip analysis identified a total of 1413 genes significantly regulated by RA. About half of these RA-regulated genes probably represent direct effects of RAR/RXR heterodimers rather than secondary effects (35). Taken together, although a great effort has been undertaken to decipher the vertebrate RA signaling network *in vivo*, our current understanding of retinoid targets and their roles in development and disease unfortunately still needs to be described as being in its infancy.

Retinoid signaling was long thought to be vertebrate-specific, but studies in invertebrate chordates (e.g. the cephalochordate amphioxus and different tunicate species) have revealed major roles for retinoids that are conserved amongst chordates (i.e. in vertebrates, cephalochordates and tunicates) (Fig. 3) (7,36,37). This work has also established that, of all invertebrates, the cephalochordate amphioxus has the most vertebrate-like retinoid signaling system, both in terms of molecular composition and biological functions, with the important added value of a lack of significant genetic redundancy (7,8,38,39).

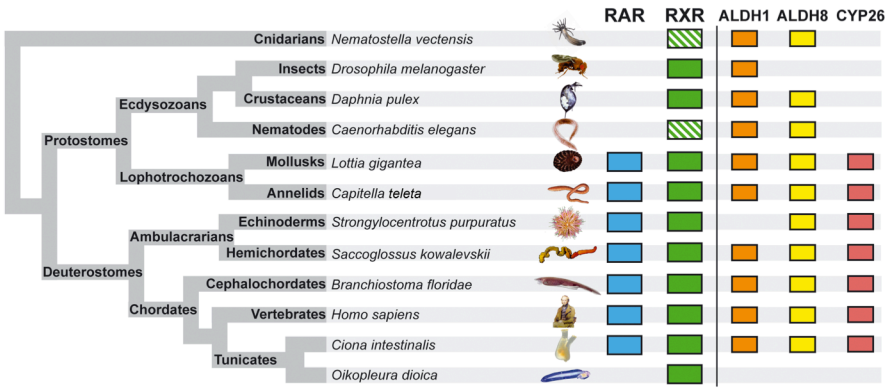


Fig. 3. Retinoid signaling in metazoans. ALDH, aldehyde dehydrogenase, CYP26, cytochrome P450, subfamily CYP26; RAR, retinoic acid receptor, RXR, retinoid X receptor.

Outside the chordate lineage, evidence for functional roles of retinoids and of the RAR/RXR heterodimer becomes much scarcer (Fig. 3). Although genes encoding orthologs of the basic vertebrate components for synthesis (RALDH) and degradation (CYP26) of endogenous RA as well as orthologs of the retinoid receptors RAR and RXR are present in ambulacrarians (such as hemichordates and echinoderms) and lophotrochozoans (such as annelids

and mollusks) (7,40), extensive experimental evidence for retinoid and/or RAR/RXR heterodimer functions in these lineages is unfortunately still lacking. Moreover, ecdysozoans (which include insects, crustaceans, and nematodes) have probably secondarily lost both the receptor RAR and the RA-degrading machinery (CYP26) (7,40).

Nonetheless, it has been suggested that RA treatment results in delay of embryonic development of the echinoderm *Paracentrotus lividus* (41) and, in the mollusk *Lymnea stagnalis*, retinoids (in particular 9-*cis* RA) disrupt embryogenesis and are involved in neuronal differentiation, outgrowth, and growth cone guidance (42). This role of retinoids in mollusk neurons is apparently independent of a functional RAR/RXR heterodimer, but instead requires a cytoplasmic localization of RXR (43). In addition, retinoid functions have also been described in tissue regeneration of insects and crustaceans (44,45) and in cell proliferation and neuronal differentiation of cnidarians (46), i.e. in animals, whose genomes do a priori not encode RAR.

Collectively, these data suggest (i) that in the course of metazoan evolution functions for retinoids as morphogens originated well before a functional RAR/RXR heterodimer, (ii) that presence of both RAR and RXR in a genome does not automatically imply a functional RAR/RXR heterodimer capable of binding and being activated by retinoids, and (iii) that at the base of the chordates novel functions for retinoid signaling have evolved, most of which are mediated by the RAR/RXR heterodimer.

Objectives

This project proposes to test the above hypotheses using a comparative work approach involving different invertebrate animal systems marking key positions of retinoid diversification (annelids, mollusks, sea urchins, hemichordates, and cephalochordates).

The proposed work can chiefly be subdivided into two tasks, which will be described in detail below: (Task 1) retinoid-dependent morphogen signaling in invertebrates and (Task 2) RAR/RXR functions in invertebrates. Whereas the first task will focus on the presence of endogenous retinoids in the different invertebrates studied, on their capacity to synthesize and degrade retinoids and on pharmacological studies of developing embryos, the second task will address the functions of the RAR/RXR heterodimer in these species (Fig. 4).

The results of the proposed work will yield information about the differential roles of retinoids and of RAR/RXR during development of five invertebrate species located at key positions of the bilaterian tree: the annelid worm *Platynereis dumerilii* and the mollusk snail *Nucella lapillus* are both lophotrochozoans belonging to the protostome clade of the bilaterian tree, the echinoderm *P. lividus* and the hemichordate *Saccoglossus kowalevskii* are ambulacrarians located at the base of the deuterostomes, and *Branchiostoma lanceolatum* is a cephalochordate and as such constitutes the base of the chordate phylum. The data obtained with this project will hence allow, for the first time, a reconstruction of the evolutionary diversification of retinoid signaling and will, in parallel, address the question when retinoid receptors (RAR and RXR) were incorporated into the retinoid cascade as the main mediators of this morphogen.

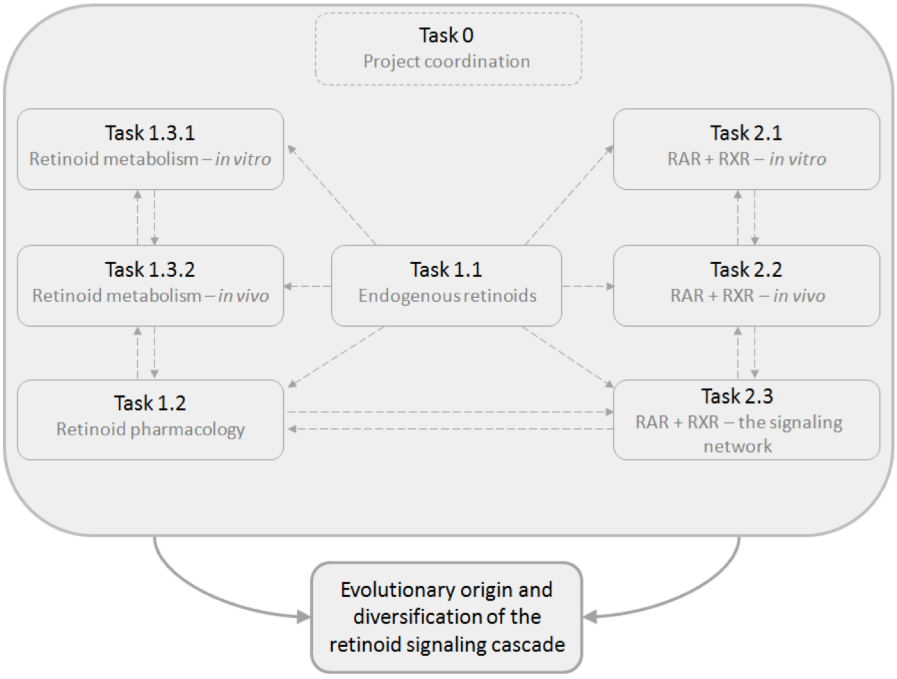


Fig. 4. Overall organization of the proposed work. The interconnection of the different tasks as well as the final outcome of the research project are indicated.

The proposal relies heavily on its comparative approach based on five different invertebrate taxa, all of which are characterized by the presence, in their genomes, of orthologs of vertebrate RA synthesis (RALDH) and degradation (CYP26) enzymes as well as orthologs of vertebrate RAR and RXR. The choice of the model organisms is thus not, as in most cases, a prerequisite for, but rather a result of, the hypotheses put forth in this proposal, which are focusing on the differential functions of retinoid morphogens during development.

Moreover, the experiments proposed in this project, designed to test the emitted hypotheses, represent an elaborate network of different technical approaches from various scientific domains and are based on the latest cutting-edge technologies. Thus, analytic chemistry methods are being used in parallel with pharmacological tools, biochemical experiments, cell culture assays, molecular biology techniques, high throughput sequencing, and genomic approaches, all of which is embedded in a very thorough developmental biology study. The combination of these methods will reveal the overall structure of the retinoid signaling network and its evolution in an unparalleled level of detail, which will without any doubt also have major repercussions on our understanding of retinoid signaling in vertebrates and, most importantly, in humans.

In sum, although the roles for retinoids in vertebrates have extensively been studied (1-8,11), we are still very far away from a comprehensive understanding of this morphogen-dependent signaling cascade, which has crucial roles both in developing and adult vertebrates. Moreover, retinoid signaling in invertebrates has received very little attention and most biological functions of this morphogen pathway in invertebrates still remain elusive (7,8,12,40-46). The work detailed herein proposes a remedy for these shortcomings by studying retinoid signaling and RAR/RXR functions in a number of different invertebrate phyla. The results will not only reveal the evolution of retinoid signaling and the origin and diversification of RAR/RXR

functions within this morphogen-controlled pathway, but also provide important new insights into the structure of an ancestral retinoid signaling network, an elaborated version of which can be found in extant vertebrates.

Scientific Program

The proposed project is taking advantage of previous *in silico* analyses (7,40) that identified molecular components of retinoid signaling in various invertebrates and will focus on taxa, whose genomes contain candidates for RA synthesis (RALDH) and degradation (CYP26) enzymes as well as orthologs of vertebrate RAR and RXR. In these taxa (annelids, mollusks, echinoderms, hemichordates, and cephalochordates), a detailed analysis of the retinoid signaling machinery will be carried out focusing on the roles played on one hand by retinoids and on the other hand by RAR and RXR during development (Fig. 4):

Task 1 Characterization of retinoid functions
- Task 1.1 Identification of endogenous retinoids
- Task 1.2 Retinoid pharmacology and phenotype analyses
- Task 1.3 Endogenous RA synthesis/degradation
  - Task 1.3.1 Cell culture-based validation
  - Task 1.3.2 Developmental roles of RA synthesis/degradation

Task 2 Characterization of RAR and RXR functions
- Task 2.1 Molecular characterization of RAR and RXR
- Task 2.2 Developmental roles of RAR and RXR
- Task 2.3 Reconstruction of an ancestral RAR/RXR signaling network

Task 0.1 Project coordination

The overall organization of this project, including scientific, administrative and financial aspects, will be the responsibility of the principal investigator, Michael Schubert. The comparative and technically diverse nature of this project necessitates a very careful coordination of the work. This is made possible, on one hand, by state-of-the-art facilities at the host institution and, on the other hand, by collaborations with some of the best research laboratories in their respective fields.

On the one hand, thus, in addition to suitable animal facilities, the host institution offers access to cutting-edge research platforms including imaging, microinjection, HPLC, mass spectrometry, genome sequencing, and bioinformatic facilities. On the other hand, the collaborative partners of this project, together with the principal investigator, in turn ensure the feasibility of all proposed experiments. Thus, while embryonic and adult material from *P. dumerilii*, *N. lapillus* and *S. kowalevskii* will respectively be provided by the laboratories of Detlev Arendt (EMBL, Heidelberg, Germany), Eduardo Jorge Sousa Rocha (CIIMAR, Porto, Portugal) and Christopher Lowe (Hopkins Marine Station of Stanford University, Pacific Grove, USA), the corresponding material from *P. lividus* and *B. lanceolatum* will be obtained from, respectively, the Station Biologique de Roscoff (Roscoff, France) and the Observatoire Océanologique de Banyuls-sur-Mer (Banyuls-sur-Mer, France). Ongoing scientific collaborations with research groups at both marine institutes (with the laboratory of Sylvie Mazan at the Station Biologique de Roscoff and with

the laboratory of Hector Escriva at the Observatoire Océanologique de Banyuls-sur-Mer) ensure both a reliable supply of animal material for the project and, at the same time, represent a significant intellectual asset for the successful initiation and termination of the proposed research.

Gene expression analyses as well as manipulative experiments of developing embryos, including pharmacological treatments and microinjections, on *P. dumerilii*, *N. lapillus* and *S. kowalevskii* will be carried out in collaboration with the laboratories of, respectively, Detlev Arendt, Eduardo Jorge Sousa Rocha and Christopher Lowe, all of which are international references for their respective model organisms (47-52). There is thus no doubt that these laboratories have the technical know-how to carry out the proposed experiments. For closely coordinating these collaborations, the principal investigator is already engaged in regular interactions with the heads of the respective laboratories. Moreover, the principal investigator is also in direct contact with the research scientists in charge of carrying out the proposed work. He is thus working, for example, with Mette Handberg-Thorsager in the laboratory of Detlev Arendt, with Filipe Castro in the laboratory of Eduardo Jorge Sousa Rocha and with Ariel Pani in the laboratory of Christopher Lowe.

Manipulative experiments and gene expression analyses in *P. lividus* and *B. lanceolatum* will be performed in the laboratory of the principal investigator with the support, in the case of *P. lividus*, of the laboratory of Thierry Lepage from the Observatoire Océanologique de Villefranche-sur-Mer (Villefranche-sur-Mer, France). The excellent track record of the principal investigator for pharmacological and manipulative work with amphioxus (38,53-62) and that of the laboratory of Thierry Lepage on developmental studies with sea urchins (63-69) combined with the technical assistance for the RNAseq and ChIPseq experiments provided by the laboratory of Gérard Benoit (CGΦMC, Lyon, France), who is world-renowned for his work on the molecular biology of nuclear receptors (70,71), leave no doubt that the proposed work on *P. lividus* and *B. lanceolatum* is very much feasible.

Task 0.2 Choice of the animal models

The animal models selected for this study are the annelid worm *P. dumerilii*, the mollusk snail *N. lapillus*, the echinoderm *P. lividus*, the hemichordate *S. kowalevskii*, and the cephalochordate *B. lanceolatum*, which together cover all the invertebrate phyla, discovered so far, with a complete set of a vertebrate-like RA signaling system encoded in the genome (RALDH, CYP26, RAR, and RXR) (7,40). While annelids and mollusks are both lophotrochozoan taxa hence belonging to the protostome clade of the bilaterian tree, echinoderms and hemichordates together constitute the so-called ambulacrarian clade located at the base of the deuterostomes. The invertebrate chordate amphioxus is a representative of the cephalochordates, which occupy a basal position within the chordates (Fig. 3). Together, this sampling of different animal taxa provides a unique opportunity to obtain novel insights into the evolution of retinoid-dependent morphogen signaling.

The annelid worm *P. dumerilii* is the best-studied lophotrochozoan model organism (47,48) and access to this biological system is going to be possible due to an ongoing collaboration with the laboratory of Detlev Arendt, an international leader in the field of evolution and development (47,48). In

addition, the genome of *P. dumerilii* is currently being sequenced with some sequence data already having been made available to the community. Another annelid genome (that of the worm *Capitella teleta*) is already publicly available (http://genome.jgi-psf.org/Capca1/Capca1.home.html).

The gastropod snail *N. lapillus* has been chosen, because it represents one of the very few accessible mollusk model species and this species has already been used extensively for pharmacological studies, also with retinoids (49,50). Access to this animal is going to be provided through a collaboration with the laboratory of Eduardo Jorge Sousa Rocha, which has a long track record working on various marine organisms, including *N. lapillus*. Concerning existing genomic resources for mollusks, the genome of *Lottia gigantea* is already publicly available (http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html).

The sea urchin *P. lividus* is a very well developed developmental model organism (65-69). Work on *P. lividus* will be carried out as part of an ongoing collaboration with the laboratory of Thierry Lepage. Sequencing of the *P. lividus* genome is currently underway and the laboratory of Thierry Lepage is actively involved in the analysis and annotation of the genomic sequence. Moreover, with the participation of the laboratory of Thierry Lepage, the genome sequence of another sea urchin species (*Strongylocentrotus purpuratus*) has already been determined and published (63,64).

S. kowalevskii is the most advanced hemichordate model system (51,52) and access to this animal is ensured through an ongoing collaboration with the laboratory of Christopher Lowe. Sequencing of the *S. kowalevskii* genome is currently underway and some genomic resources are already publicly available.

With strong involvement of the principal investigator, the cephalochordate amphioxus has been at the center of research into the evolution of retinoid signaling for well over a decade. Using extensive pharmacological and manipulative approaches, it has been established that amphioxus has a vertebrate-like RA signaling cascade that lacks the extensive gene duplications characteristic of vertebrates (53-59,72,73). For example, there is only one RAR and one RXR in amphioxus, while vertebrates have 3 (or more) RAR and 3 (or more) RXR proteins (73). We will use the European amphioxus (*B. lanceolatum*) for our studies, because it is very well suited for laboratory use with on-demand spawning inducible by simple thermal shock (60-62). A proposal for sequencing the *B. lanceolatum* genome has just been accepted by the Genoscope (Evry, France) and the genome of the Florida amphioxus (*B. floridae*) has already been published (74,75). The principal investigator significantly contributed to analysis and annotation of the *B. floridae* genome and is also an active member of the *B. lanceolatum* genome sequencing consortium, which will ensure full access to the *B. lanceolatum* genome sequence.

Task 1 Characterization of retinoid functions

The experiments proposed in this task will assess the roles of retinoids during development in different invertebrate taxa (Fig. 4).

Task 1.1 Identification of endogenous retinoids

The very first step for analyzing retinoid signaling in invertebrates is to assay for the retinoids that are present in the adult and, more importantly, in

the developing embryo. The retinoid content will be established by preparing extracts of animal tissues followed by HPLC analyses coupled with mass spectrometry detection. HPLC and mass spectrometry facilities are available at the host institution. Very detailed methods for the quantification of endogenous retinoids have recently been published (76) and the retinoid content, for example, of adult *B. lanceolatum* (77) has already successfully been established, which will serve as a positive control for our analyses. Given that an overall tissue volume of 100μl and 300μl is sufficient for successful retinoid extraction and subsequent HPLC separation (76), both embryonic and adult material from *P. dumerilii*, *N. lapillus*, *P. lividus*, *S. kowalevskii*, and *B. lanceolatum* will be analyzed. For all species, developmental stage(s) characterized by expression of putative RA synthesizing enzymes (RALDH, see Task 1.3) will be selected preferentially for retinoid extraction and analysis. These assays will yield a list of retinoids present in embryos and adults of the different invertebrates under study, which, in turn, will allow comparisons of the endogenous retinoid contents of these invertebrate taxa and will, more importantly, lead to the identification of specific retinoids with putative biological activity.

Task 1.2 Retinoid pharmacology and phenotype analyses

The next aim is to study the roles, during development, of retinoids in the five invertebrate taxa (*P. dumerilii*, *N. lapillus*, *P. lividus*, *S. kowalevskii*, and *B. lanceolatum*). Developing embryos will thus be treated with different concentrations of retinoids known to be biologically active in vertebrates (e.g. all-*trans* RA, 9-*cis* RA, 13-*cis* RA) at different time points of development. Moreover, corresponding treatments will be carried out with retinoids identified in the retinoid content analyses (see Task 1.1). The phenotypes obtained from the treatments will be analyzed morphologically and by *in situ* hybridization experiments using region- and cell population-specific marker genes of the developing embryos, which are readily available for annelids, mollusks, echinoderms, hemichordates, and cephalochordates. These analyses will reveal time- and concentration-dependent effects of the different retinoids during development of these invertebrates. Comparisons of these effects will show, which retinoid is active in which animal at which developmental stage and the combination of all these data will reveal retinoids that are active in all or most of the taxa studied, which is indicative of a conserved, putatively ancestral, function.

Once retinoids with conserved functions in the studied taxa have been identified, the treatments of developing embryos with these compounds will be repeated and the embryos will subsequently be fixed for RNA extraction. This RNA (and corresponding control RNA of untreated embryos) will then be subjected to high throughput RNA sequencing (RNAseq) (using the paired-end Illumina technique) to obtain expression profiles of the retinoid-dependent developmental transcriptome. These RNAseq experiments will be carried out in collaboration with the laboratory of Gérard Benoit, who is currently finalizing an RNAseq analysis of the effects of RA in the embryonal carcinoma cell line F9. Since these experiments require a well-annotated genome, they will only become feasible, once assembled genome sequences of *P. dumerilii*, *N. lapillus*, *P. lividus*, *S. kowalevskii*, and *B. lanceolatum* become available, which, at least for *P. dumerilii*, *P. lividus*, *S. kowalevskii*, and *B. lanceolatum*

will be before the end of 2011. Comparison of the RNAseq transcriptomic data from different invertebrates will highlight the basic, conserved building blocks of the ancestral retinoid-dependent signaling network.

The pharmacological and developmental aspects of this task in *P. dumerilii*, *N. lapillus*, and *S. kowalevskii* will be carried out in the laboratories of, respectively, Detlev Arendt, Eduardo Jorge Sousa Rocha and Christopher Lowe.

Task 1.3 Endogenous RA synthesis/degradation

Although it is very important to study the effects of exogenous retinoids during development, the biochemical potential to endogenously synthesize and degrade active retinoids is also crucial to assess. To this end, orthologs of vertebrate RA synthesis (RALDH) and degradation (CYP26) enzymes will be cloned from *P. dumerilii*, *N. lapillus*, *P. lividus*, *S. kowalevskii*, and *B. lanceolatum* for functional analyses. Our preliminary *in silico* analyses (7,59) indicate that there are multiple RALDH and CYP26 orthologs in the five lineages and preliminary cloning work has already yielded at least pieces of most RALDH and CYP26 orthologs of the five species.

Task 1.3.1 Cell culture-based validation

Just because a given invertebrate species possesses an ortholog of an enzyme that, in vertebrates, is responsible for the synthesis or the degradation of RA, the enzymatic function of this protein in the invertebrate might still be very different (59). Thus, after full-length cloning, the biochemical properties of the putative invertebrate RALDH and CYP26 enzymes need to be tested. These assays will be carried out using a cell culture-based system with readouts based on HPLC analyses coupled with mass spectrometry detection. The cell lines used for these experiments will vary depending on the experimental set up: while for RA synthesis assays cells devoid of endogenous RA synthesis (such as MCF-7, T47D or OVCAR3) will be used, the RA degradation assays will be based on cells lacking RA-dependent induction of endogenous CYP26 (such as HEK293 or MDA-MB-231) (78,79). The RALDH enzymes will be assayed in the presence of retinol and the CYP26 proteins in the presence of both retinoids known to be biologically active in vertebrates (e.g. all-*trans* RA, 9-*cis* RA, 13-*cis* RA) and endogenous retinoids identified in the retinoid extracts (see Task 1.1). These experiments will yield the capacities of the different invertebrate RALDH orthologs to synthesize and of the different invertebrate CYP26 orthologs to degrade retinoids, which in turn will allow the identification of the major components for RA synthesis and degradation in the five taxa.

Task 1.3.2 Developmental roles of RA synthesis/degradation

The biochemical analysis of the different RALDH and CYP26 orthologs will be followed by a developmental characterization of the functions of RALDH and CYP26 in invertebrates, focusing on the RALDH and CYP26 orthologs with evident RA synthesis and degradation capacities, respectively (see Task 1.3.1). Developmental expression of the RALDH and CYP26 orthologs will be assessed using *in situ* hybridization. A survey of developmental RALDH expression has recently been published for the Florida amphioxus (*B. floridae*) (59), and the data reported in this article will serve as

a basis for a much more detailed expression analysis proposed as part of this project. The functions of the RALDH and CYP26 orthologs will be analyzed by gene knockdown and overexpression experiments in *P. dumerilii*, *P. lividus*, *S. kowalevskii*, and *B. lanceolatum* (manipulative experiments are currently not possible for *N. lapillus*). The phenotypes obtained from these experiments will be analyzed morphologically and by *in situ* hybridization using the available region- and cell population-specific marker genes of developing embryos. Collectively, these analyses will yield a very detailed picture of the localization of sources and sinks of endogenous RA in developing invertebrates and of the functions of RA synthesizing and degrading enzymes during the development of these animals.

The developmental aspects of this task involving *P. dumerilii*, *N. lapillus*, and *S. kowalevskii* will be carried out in the laboratories of, respectively, Detlev Arendt, Eduardo Jorge Sousa Rocha and Christopher Lowe.

Task 2 Characterization of RAR and RXR functions

As much as the first task is meant to elucidate functions of retinoids during development of five different invertebrate taxa, this second task aims at assessing the functions of the classical vertebrate retinoid receptors, RAR and RXR, in the context of the overall retinoid signaling cascade in these animals (Fig. 4). Our preliminary *in silico* analyses (7) indicate that there is at least one RAR and one RXR ortholog in each of the five invertebrate lineages. Although suggestive of a function in retinoid signaling, the presence of RAR and RXR in the genome of a given animal alone is not sufficient to assign a specific function to these receptors. It is thus very important to carry out a very detailed functional characterization of the invertebrate RAR and RXR proteins.

Task 2.1 Molecular characterization of RAR and RXR

After full-length cloning of RAR and RXR from *P. dumerilii*, *N. lapillus*, *P. lividus*, *S. kowalevskii*, and *B. lanceolatum*, which is currently underway, the proteins will be subjected to a molecular characterization to assess their DNA binding, dimerization, ligand binding, and transactivation capacities. While the DNA binding and dimerization properties of the different RAR and RXR proteins will be studied using electrophoretic mobility shift assays (EMSAs), ligand binding capacities of the receptors will initially be assessed with limited proteolysis assays (LPAs). In these LPA experiments, the endogenous retinoids identified in the retinoid extracts (see Task 1.1) will be tested alongside retinoids known to be biologically active in vertebrates (e.g. all-*trans* RA, 9-*cis* RA, 13-*cis* RA). Once a retinoid has been shown to interact with a RAR or RXR protein in a LPA assay, the affinity and kinetics of this interaction will be studied using a Biacore system, which is accessible through a collaboration with the laboratory of Vincent Laudet (IGFL, Lyon, France). Finally, the potential of the RAR/RXR heterodimer to activate transcription in the presence of a ligand will be studied using cell culture-based transactivation experiments. One amphioxus RAR/RXR heterodimer (that of the Florida amphioxus, *B. floridae*) has already been characterized using EMSA, LPA, and transactivation experiments and these data can thus be used as a positive control (73,80). Together, the data obtained as part of this task will yield crucial information about the molecular properties of the

invertebrate RAR and RXR proteins, both in terms of DNA and dimerization properties and in terms of ligand binding and transactivation capacity. This work will thus show, whether RAR/RXR heterodimers from lophotrochozoans and ambulacrarians, just like their chordate orthologs, can bind retinoids and activate transcription upon ligand binding.

Task 2.2 Developmental roles of RAR and RXR

In addition to this molecular characterization, the developmental roles of RAR and RXR will also be studied in the five invertebrate taxa. It is conceivable that the RAR/RXR heterodimer has functional roles during development even if both RAR and RXR are incapable of binding retinoids. Developmental expression of RAR and RXR will hence be assessed using *in situ* hybridization experiments. Expression of RAR and RXR has already been described in the Florida amphioxus (*B. floridae*), which will serve as a positive control for our analyses (73). Moreover, the functions of the RAR and RXR orthologs will be analyzed by gene knockdown and overexpression experiments in *P. dumerilii*, *P. lividus*, *S. kowalevskii*, and *B. lanceolatum* (manipulative experiments are currently not possible for *N. lapillus*). The phenotypes obtained from these experiments will be analyzed morphologically and by *in situ* hybridization using region- and cell population-specific marker genes of developing embryos. The results obtained from these experiments will allow a very fine-grained analysis of the functional roles played by RAR and RXR during the development of *P. dumerilii*, *N. lapillus*, *P. lividus*, *S. kowalevskii*, and *B. lanceolatum*. Comparing and contrasting these functions with the roles observed for retinoids during development represents the key to testing the three working hypotheses put forth in this proposal.

The developmental aspects of this task involving *P. dumerilii*, *N. lapillus*, and *S. kowalevskii* will be carried out in the laboratories of, respectively, Detlev Arendt, Eduardo Jorge Sousa Rocha and Christopher Lowe.

Task 2.3 Reconstruction of an ancestral RAR/RXR signaling network

The final step proposed here for the characterization of RAR/RXR-dependent functions in invertebrates is the attempt to reconstruct the RAR/RXR signaling network of the studied species. In order to achieve this, ChIP experiments on RAR and RXR will be performed on whole embryos at developmental stages showing significant expression of the two receptors. The ChIP result will subsequently be subjected to high throughput DNA sequencing (ChIPseq) (using the paired-end Illumina technique) in order to obtain a genome-wide analysis of RAR and RXR binding sites. These ChIPseq experiments will be carried out in collaboration with the laboratory of Gérard Benoit, who is currently finalizing a ChIPseq analysis targeting RAR and RXR in the embryonal carcinoma cell line F9. Moreover, for setting up the ChIP experiments in annelids, echinoderms, and amphioxus, in collaboration with the laboratory of Detlev Arendt, the development of antibodies directed against RAR and RXR of *P. dumerilii*, *P. lividus*, and *B. lanceolatum* has already been initiated. Since these experiments require a well-annotated genome, they will only become feasible, once assembled genome sequences of *P. dumerilii*, *N. lapillus*, *P. lividus*, *S. kowalevskii*, and *B. lanceolatum* become available, which, at least for *P. dumerilii*, *P. lividus*, *S. kowalevskii*,

and *B. lanceolatum* will be before the end of 2011. Comparison of the ChIPseq data from the different species will allow the definition of a minimal, ancestral RAR/RXR signaling network, before taxon-specific diversification.

Exploitation of the Anticipated Results

The biological functions of retinoids have extensively been studied in vertebrates and invertebrates chordates (1-8, 11). In contrast, much less is known about retinoid functions in other animal groups and, more generally, about the evolutionary origins and diversification of this morphogen-dependent signaling cascade. Previous analyses have suggested that functions for retinoids as morphogens originated well before a functional RAR/RXR heterodimer, the main mediator of retinoid signaling in chordates, and that the presence of RAR and RXR in a given invertebrate taxon is no automatic proof for a functional RAR/RXR heterodimer capable to bind to and be activated by retinoids (7,8,12,40-46). This latter fact evokes the question of the biological functions of RAR and RXR proteins putatively incapable of binding retinoids.

The project proposes work to study retinoid signaling in invertebrate taxa, whose genomes encode orthologs of both vertebrate RA synthesis and degradation enzymes and of RAR and RXR. The experiments are designed to independently assess the roles of retinoids and of RAR and RXR in five invertebrate models (two lophotrochozoan protostomes, two ambulacrarian deuterostomes and one invertebrate chordate), which will allow a detailed assessment of the functions of retinoids versus the functions of RAR and RXR in each of these lineages. Comparisons of these respective functions will reveal roles for retinoids that are dependent on RAR and RXR and those that are independent of these receptors. Moreover, cross-comparisons of these results between the five species will identify conserved and divergent elements of retinoid signaling, which in turn will reveal the evolutionary diversification of this morphogen-dependent signaling cascade in bilaterian animals.

On a more applicable scale, the use of latest high throughput sequencing approaches will create a wealth of data about the molecular composition of the retinoid signaling network in non-standard model systems, including animals located at key positions of the bilaterian tree of life. For example, the RNAseq and ChIPseq data from lophotrochozoans and ambulacrarians will reveal the retinoid signaling network (dependent on RAR and RXR or not) of Urbilateria, the last common ancestor of protostomes and deuterostomes.

Likewise, the comprehensive RNAseq and ChIPseq data set from the invertebrate chordate amphioxus will allow the reconstitution of the ancestral chordate retinoid signaling cascade, which is equivalent to a simple form of the vertebrate retinoid signaling network devoid of duplications. Given the bifurcated work approach proposed here, both the RAR/RXR-dependent signaling network and the functions of retinoids that are independent of RAR and RXR will be revealed. In this respect, by studying animals that, at first glance, do not exhibit obvious ties to humans, the work proposed within this project offers a unique opportunity to instruct the complexity of the human

retinoid system, with its crucial functions during embryology and in tissue homeostasis.

It goes without saying that these incredibly important data have to be made available in full to the scientific community and beyond. All significant advances will thus be published in the form of research and review articles or book chapters and presented at scientific meetings. Moreover, all sequences generated in the course of this project will be deposited in the appropriate online databases and all protocols developed and material produced during the project will be made freely available. In order to reach a broader audience including non-specialists, efforts will be undertaken to publicize the results in popular and even non-science magazines. The principal investigator will also actively participate in the organization of public visits to the host institution and in scientific exchange initiatives between the host institution and local schools to interest the youngest public in the miracles of evolution and animal development.

Task Schedule, Deliverables and Milestones

Task schedule

	Year 1		**Year 2**		**Year 3**	
	1st semester	**2nd semester**	**1st semester**	**2nd semester**	**1st semester**	**2nd semester**
Task 0 Project coordination	Kick-off meeting + report	Informal meetings	Informal meetings	Intermediary meeting + report	Informal meetings	Project conclusion meeting + report
Task 1.1 Endogenous retinoids	Sample collection + extraction - adult/embryo	Sample collection + extraction - adult/embryo	Sample analyses	Sample analyses (final retinoid list)		Publication of results
Task 1.2 Retinoid pharmacology	Retinoid treatments + sample preparation	Phenotype analyses + RNAseq experiments	Phenotype + RNAseq data analyses	Retinoid treatments (with retinoid list) + sample preparation	Phenotype analyses + RNAseq experiments	Phenotype + RNAseq data analyses + publication of results
Task 1.3.1 Retinoid metabolism – *in vitro*	Gene cloning + cell culture assays	Cell culture assays + sample analyses	Cell culture assays + sample analyses	Cell culture assays + sample analyses (final gene list)		Publication of results
Task 1.3.2 Retinoid metabolism – *in vivo*	Gene cloning + expression analyses	Expression analyses + manipulative experiments	Manipulative experiments + phenotype analyses	Manipulative experiments (with gene list) + phenotype analyses	Phenotype analyses	Phenotype analyses + publication of results
Task 2.1 RAR + RXR – *in vitro*	Gene cloning + molecular analyses	Molecular analyses	Molecular analyses + cell culture assays	Molecular analyses + cell culture assays	Cell culture assays	Publication of results
Task 2.2 RAR + RXR – *in vivo*	Gene cloning + expression analyses	Expression analyses + manipulative experiments	Manipulative experiments + phenotype analyses	Manipulative experiments + phenotype analyses	Phenotype analyses	Phenotype analyses + publication of results
Task 2.3 RAR + RXR – the signaling network	Gene cloning + antibody development + validation	Antibody validation + initial ChIP assays	ChIP assays	ChIPseq experiments	ChIPseq experiments + ChIPseq data analysis	ChIPseq data analysis + publication of results

Deliverables and milestones

Number	Title	Delivery time (in months after the beginning of the project)
Task 0 Project coordination		
0a	Evolutionary origin and diversification of retinoid signaling	36
Task 1 Characterization of retinoid functions		
Task 1.1 Identification of endogenous retinoids		
1a	Biological sample collection and extraction	12
1b	List of endogenous retinoids	24
Task 1.2 Retinoid pharmacology and phenotype analyses		
1c	Developmental pharmacology of exogenous retinoids	36
1d	Retinoid-dependent developmental transcriptome	36
Task 1.3 Endogenous RA synthesis/degradation		
1e	Cloning of full-length RA synthesis/degradation genes	6
Task 1.3.1 Cell culture-based validation		
1f	Identification of enzyme capacities	24
Task 1.3.2 Developmental roles of RA synthesis/degradation		
1g	Survey of developmental expression	12
1h	Detailed characterization of developmental functions	36
Task 2 Characterization of RAR and RXR functions		
2a	Cloning of full-length RAR and RXR genes	6
Task 2.1 Molecular characterization of RAR and RXR		
2b	Molecular properties of RAR and RXR	30
Task 2.2 Developmental roles of RAR and RXR		
2c	Developmental expression of RAR and RXR	12
2d	Functions of RAR and RXR in development	36
Task 2.3 Reconstruction of an ancestral RAR/RXR signaling network		
2e	Antibody development and validation	12
2f	Chromatin immunoprecipitation of RAR and RXR	24
2g	RAR and RXR signaling network	36

References

1. Collins, M. D., Mao, G. E. (1999) Annu. Rev. Pharmacol. Toxicol. 39:399-430.
2. Blomhoff, R., Blomhoff, H. K. (2006) J. Neurobiol. 66:606-630.
3. Glover, J. C., et al. (2006) J. Neurobiol. 66:705-725.
4. Mark, M., et al. (2006) Annu. Rev. Pharmacol. Toxicol. 46:451-480.
5. Duester, G. (2008) Cell 134:921-931.
6. Morriss-Kay, G. M., Ward, S. J. (1999) Int. Rev. Cytol. 188:73-131.
7. Campo-Paysaa, F., et al. (2008) Genesis 46:640-656.
8. Theodosiou, M., et al. (2010) Cell. Mol. Life Sci. 67:1423-1445.
9. Fields, A. L., et al. (2007) J. Cell. Biochem. 102:886-898.
10. Lammer, E. J., et al. (1985) N. Engl. J. Med. 313:837-841.
11. Maden, M. (2002) Nat. Rev. Neurosci. 3:843-853.
12. Simões-Costa, M. S., et al. (2008) J. Exp. Zool. B 310:54-72.
13. Kim, C. I., et al. (1992) Arch. Biochem. Biophys. 294:388-393.
14. Boleda, M. D., et al. (1993) Arch. Biochem. Biophys. 307:85-90.
15. Yang, Z. N., et al. (1994) Alcohol. Clin. Exp. Res. 18:587-591.
16. Duester, G. (1996) Biochemistry 35:12221-12227.
17. Lee, M. O., et al. (1991) Biochem. Pharmacol. 42:1279-1285.
18. Niederreither, K., et al. (1997) Mech. Dev. 62:67-78.
19. Lin, M., et al. (2003) J. Biol. Chem. 278:9856-9861.
20. Niederreither, K., et al. (1999) Nat. Genet. 21:444-448.
21. White, R. J., et al. (2007) PloS Biol. 5:e304.
22. Niederreither, K., et al. (2002) Nat. Genet. 31:84-88.
23. Reijntjes, S., et al. (2005) Dev. Biol. 285:224-237.
24. Gronemeyer, H., et al. (2004) Nat. Rev. Drug. Discov. 3:950-964.
25. Balmer, J. E., Blomhoff, R. (2005) J. Steroid Biochem. Mol. Biol. 96:347-354.
26. Chambon, P. (1996) FASEB J. 10:940-954.
27. Ross, S. A., et al. (2000) Physiol. Rev. 80:1021-1054.
28. Mic, F. A., et al. (2003) Proc. Natl. Acad. Sci. U.S.A. 100:7135-7140.
29. Calléja, C., et al. (2006) Genes Dev. 20:1525-1538.
30. Stehlin-Gaon, C., et al. (2003) Nat. Struct. Biol. 10:820-825.
31. Schug, T. T., et al. (2007) Cell 129:723-733.
32. Kruse, S. W., et al. (2008) PLoS Biol. 6:e227.
33. Aggarwal, S., et al. (2006) Mol. Biol. Cell 17:566-575.
34. Ochoa, W. F., et al. (2003) Biochemistry 42:8774-8779.
35. Hua, S., et al. (2009) Cell 137:1259-1271.
36. Cañestro, C., et al. (2006) Evol. Dev. 8:394-406.
37. Marlétaz, F., et al. (2006) Int. J. Biol. Sci. 2:38-47.
38. Schubert, M., et al. (2006) Trends Ecol. Evol. 21:269-277.
39. Koop, D., Holland, L. Z. (2008) Birth Defects Res. C Embryo Today 84:175-187.
40. Albalat, R., Cañestro, C. (2009) Chem. Biol. Interact. 178:188-196.
41. Sciarrino, S., Matranga, V. (1995) Cell Biol. Int. 19:675-680.
42. Farrar, N. R., et al. (2009) J. Neurosci. 29:14136-14142.
43. Carter, C. J., et al. (2010) Dev. Biol. 343:124-137.
44. Hopkins, P. (2001) Am. Zool. 41:389-398.
45. Halme, A., et al. (2010) Curr. Biol. 20:458-463.

46. Estephane, D., Anctil, M. (2010) Dev. Neurobiol. 70:842-852.
47. Christodoulou, F., et al. (2010) Nature 463:1084-1088.
48. Tomer, R., et al. (2010) Cell 142:800-809.
49. Castro, L. F., et al. (2007) Aquat. Toxicol. 85:57-66.
50. Lima, D., et al. (2011) Aquat. Toxicol. 101:221-227.
51. Lowe, C. J., et al. (2003) Cell 113:853-865.
52. Lowe, C. J., et al. (2006) PLoS Biol. 4:e291.
53. Schubert, M., et al. (2004) Proc. Natl. Acad. Sci. U.S.A. 101:10320-10325.
54. Schubert, M., et al. (2005) Development 132:61-73.
55. Schubert, M., et al. (2006) Dev. Biol. 296:190-202.
56. Onai, T., et al. (2009) Dev. Biol. 332:223-233.
57. Osborne, P. W., et al. (2009) Dev. Biol. 327:252-262.
58. Koop, D., et al. (2010) Dev. Biol. 338:98-106.
59. Sobreira, T. J. P., et al. (2011) Proc. Natl. Acad. Sci. U.S.A. 108:226-231.
60. Fuentes, M., et al. (2004) J. Exp. Zool. B 302:384-391.
61. Fuentes, M., et al. (2007) J. Exp. Zool. B 308:484-493.
62. Theodosiou, M., et al. (2011) J. Exp. Zool. B 316:263-275.
63. Lapraz, F., et al. (2006) Dev. Biol. 300:132-152.
64. Sodergren, E., et al. (2006) Science 314:941-952.
65. Range, R., et al. (2007) Development 134:3649-3664.
66. Röttinger, E., et al. (2008) Development 135:353-365.
67. Lapraz, F., et al. (2009) PLoS Biol. 7:e1000248.
68. Duboc, V., et al. (2010) Development 137:223-235.
69. Saudemont, A., et al. (2010) PLoS Genet. 6:e1001259.
70. Benoit, G., et al. (2006) Pharmacol. Rev. 58:798-836.
71. Wang, Z., et al. (2003) Nature 423:555-560.
72. Holland, L. Z., Holland, N. D. (1996) Development 122:1829-1838.
73. Escriva, H., et al. (2002) Development 129:2905-2916.
74. Holland, L. Z., et al. (2008) Genome Res. 18:1100-1111.
75. Putnam, N. H., et al. (2008) Nature 453:1064-1071.
76. Kane, M. A., Napoli, J. L. (2010) Methods Mol. Biol. 652:1-54.
77. Dalfó, D., et al. (2002) Dev. Genes Evol. 212:388-393.
78. Mira-Y-Lopez, R., et al. (2000) J. Cell. Physiol. 185:302-309.
79. Williams, S.J., et al. (2009) Gynecol. Oncol. 112:637-645.
80. Escriva, H., et al. (2006) PLoS Genet. 2:e102.
81. Holland, L. Z., et al. (2004) Cell. Mol. Life Sci. 61:2290-2308.
82. Kozmik, Z., et al. (1999) Development 126:1295-1304.
83. Schubert, M., et al. (2000) Mol. Biol. Evol. 17:1896-1903.
84. Schubert, M., et al. (2001) Dev. Biol. 240:262-273.
85. Schubert, M., et al. (2008) Dev. Genes Evol. 218:651-665.
86. Paris, M., et al. (2008) Dev. Genes Evol. 218:667-680.
87. Campo-Paysaa, F., et al. (2011) Evol. Dev. 13:15-27.
88. Paris, M., et al. (2008) Curr. Biol. 18:825-830.