### letters to nature

pointed along the home vector and the second arm in the opposite direction. A final test (not illustrated) showed that ants were equally quick at learning a fixed path if they had to choose the same side in all compartments.

The ability to learn fixed paths rapidly, without a compatible home vector, indicates that ants may temporarily store a sequence of path segments on every trip home. If path segments can be encoded economically as local vectors<sup>5</sup>, the ant's speculative learning of a path should not make unreasonable demands on its short-term memory. In contrast, ants seem to avoid the speculative learning of views. Instead, they use home vectors to trigger or permit the storage of significant scenes. A reinforcement signal emanating from the home vector would mean that different scenes viewed along a route could all be acquired on the same trip independently of each other. When learning a complex route, an ant may depend initially upon the rapid acquisition of path segments through motor learning. Routes may then be stabilized through visual learning that is guided by path integration. By this means paths can become as economical as the terrain allows, with acquired motor patterns moulded to the improved route.

#### Methods

Maze training. Experiments were done using queenright colonies of Cataglyphis cursor maintained in the laboratory9. At the start of an experiment, some foragers were taken from the sucrose feeder at the end of the tunnel and marked with paint. Marked individuals always returned home through the maze. They were collected on their arrival at the feeder and released in a waiting box with another similar feeder. Ants taken from the waiting box were placed singly at the start of the maze. Each ant performed between five and ten trips a day. During visual training, positive and negative shapes were exchanged after each trial, and the compartments repositioned so that the previously blocked exit was open and the previously open exit blocked. Training thus involved an alternating motor pattern from trial to trial (L(left)R(right)LR or RLRL through the exits in the four compartments). To examine how this pattern influenced the ant's choice, we performed two tests in which ants that had foraged through the linear tunnel, but were naive to the maze, performed a single homeward passage through the linear maze. In test 1, the open exits from compartments 1 to 4 were LLLR or RRRL, and in test 2 the open compartments were LRLR or RLRL. In both tests 1 and 2, ants chose randomly in compartment 1. In compartment 2, they tended to choose the exit on the side that had been open in compartment 1. The behaviour of the two groups diverged in compartments 3 and 4. Ants in test 1 persisted in choosing the exit on the side that had been open in compartments 1 and 2. Thus, in compartment 4, 22 ants chose the negative shape (and the closed exit) and 3 ants chose the positive shape (binomial test, two-tail P < 0.001). In contrast, the ants in test 2 tended to follow the alternating pattern. In compartment 4, 35 ants chose the positive shape and 12 ants chose the negative shape (binomial test, two-tail P < 0.001). This sensitivity to the alternating training sequence is likely to reduce the errors made in the last two compartments during visual acquisition. Although cataglyphid ants do not lay chemical trails12, it is important with fixed-path learning to ensure that ants cannot be guided by chemical cues. Compartments and tunnels were therefore switched to distribute any potential chemical cues evenly over the maze.

**Tests in the arena.** Ants were carried to the circular arena (diameter 20 cm) on a straw which was placed upright in a hole in the centre. The ant climbed off the straw and walked towards the periphery. We recorded in which  $5^{\circ}$  or  $30^{\circ}$  sector the ant was located when it reached the wall of the arena.

**Statistical treatment.** On each trial we scored the choice of exits made in each compartment and the time spent travelling from the first compartment to the end of the maze. The binomial distribution was applied to the group plots of Figs 2 and 4, using a one-tailed test, to show when the error score dropped below chance at P < 0.01. To assess the learning of individual ants contributing to the group plots in Fig. 2, we counted the number of errors made by each ant in its first choice of exit in each compartment. We then used the binomial test to determine whether an individual's score within a compartment was less than that attributable to chance (P < 0.05). Mean directions and confidence intervals in Fig. 3c were obtained as described by Batschelet<sup>13</sup>.

Received 29 December 1998; accepted 23 April 1999.

- Wehner, R., Harkness, R. D. & Schmid-Hempel, P. Foraging strategies in individually searching ants, Cataglyphis bicolor (Hymenoptera: Formicidae). Akad. Wiss. Lit. Mainz, Math.-Naturwiss. Kl. 1, 1–79 (1983)
- Wehner, R. Spatial organization of foraging behaviour in individually searching desert ants, Cataglyphis (Sahara desert) and Ocymyrmex (Namib desert). Experientia 54 (Suppl.), 15–42 (1987).
- Collett, T. S., Dillmann, E., Giger, A. & Wehner, R. Visual landmarks and route following in desert ants. J. Comp. Physiol. A 170, 435

  –442 (1992).
- Wehner, R., Michel, B. & Antonsen, P. Visual navigation in insects: coupling of egocentric and geocentric information. J. Exp. Biol. 199, 129–140 (1996).
- Collett, M., Collett, T. S., Bisch, S. & Wehner, R. Local and global vectors in desert ant navigation. Nature 394, 269–272 (1998).
- Mittelstaedt, H. & Mittelstaedt, M. L. in Avian Navigation (eds Papi, F. & Wallraff, H. G.) 290–297 (Springer, New York, 1982).
- 7. Wehner, R. & Wehner, S. Insect navigation: use of maps or Ariadne's thread? *Ethol. Ecol. Evol.* 2, 27–48
- 8. Collett, T. S. Rapid navigational learning in insects with a short lifespan. *Connection Sci.* **10**, 255–270
- Chameron, S., Schatz, B., Pastergue-Ruiz, I., Beugnon, G. & Collett, T. S. The learning of a sequence of visual patterns by the ant Cataglyphis cursor. Proc. R. Soc. Lond. B 265, 2309–2313 (1998).
- Lenoir, A., Nowbahari, E., Querard, L., Pondicq, N. & Delalande, C. Habitat exploitation and intercolonial relationships in the ant Cataglyphis cursor (Hymenoptera: Formicidae). Acta Oecologica 11, 3–18 (1990).
- Schmidt, I., Collett, T. S., Dillier, F.-X. & Wehner, R. How desert ants cope with enforced detours on their way home. J. Comp. Physiol. A 171, 285–288 (1992).
- Harkness, R. D. Quantitative observations on the foraging of nests of an ant (Cataglyphis bicolor F.) in Greece. Acta Entomol. Jugosl. 13, 21–33 (1977).
- 13. Batschelet, E. Circular Statistics in Biology (Academic, London, 1981).

Acknowledgements. We thank V. Fourcassié, P. Graham, M. Jones and J. Lauga for discussion. Financial support came from the European Commission (TMR Program) to B.S. and T.S.C., the French Ministry of Research to S.C., the Science de la Cognition Program to G.B., and the UK BBSRC and the Human Frontier Science Program to T.S.C.

Correspondence and requests for materials should be addressed to T.S.C. (e-mail: T.S.Collett@sussex. ac.uk).

# Hox genes in brachiopods and priapulids and protostome evolution

Renaud de Rosa\*†, Jennifer K. Grenier\*‡, Tatiana Andreeva§, Charles E. Cookl, André Adoutte†, Michael Akaml, Sean B. Carroll‡ & Guillaume Balavoinell¶

† Laboratoire de Biologie Cellulaire 4, CNRS UPRESA Q8080, Université Paris-Sud, 91405 Orsay Cedex, France

‡ Howard Hughes Medical Institute and Laboratory of Molecular Biology, University of Wisconsin, 1525 Linden Drive, Madison, Wisconsin 53706, USA § Department of Embryology, Biological Institute, St Petersburg University, 199034 St Petersburg, Russia

Laboratory for Development and Evolution, University Museum of Zoology, Department of Zoology, Downing St, Cambridge CB2 3EJ, UK

\* These authors contributed equally to this work.

Understanding the early evolution of animal body plans requires knowledge both of metazoan phylogeny and of the genetic and developmental changes involved in the emergence of particular forms. Recent 18S ribosomal RNA phylogenies suggest a three-branched tree for the Bilateria comprising the deuterostomes and two great protostome clades, the lophotrochozoans¹ and ecdysozoans². Here, we show that the complement of Hox genes in critical protostome phyla reflects these phylogenetic relationships and reveals the early evolution of developmental regulatory potential in bilaterians. We have identified Hox genes that are shared by subsets of protostome phyla. These include a diverged pair of posterior (*Abdominal-B-like*) genes in both a brachiopod and a polychaete annelid, which supports the lophotrochozoan assemblage, and a distinct posterior Hox gene shared by a priapulid, a nematode and the arthropods, which supports the

<sup>¶</sup> Present address: Centre de Génétique Moléculaire, CNRS UPR 9061, Ave. de la Terrasse, Bâtiment 26, 91198 Gif-sur-Yvette, France.

ecdysozoan clade. The ancestors of each of these two major protostome lineages had a minimum of eight to ten Hox genes. The major period of Hox gene expansion and diversification thus occurred before the radiation of each of the three great bilaterian clades. Hox genes have key roles in patterning the antero-posterior axis of bilaterians<sup>3</sup>. The reconstruction of the evolutionary history of the Hox gene family is thus central to understanding the evolution of bilaterian body plans and the relationship between genetic and morphological complexity. Hox gene distribution also provides a

Dme Antp		RKRGRQTYTRYQTL	ELEKEFHFNRY	LIKKKKIEIAHA	LCLTERQIKIV	VFQNRRMKWKKEN	
Dme lab	I.SSCSI.SSNT	NNSTNF-NK-LT		ANT	-0-N-T-V	BV	KEGLIPADILI
Pca lab		1110 1111 1111 111		<mark>A</mark> AS		2 2. 0	
Lan lab	KNYQDYSGNI	PNMTNFSDK-LT				<mark>Q</mark> <mark>R</mark> M	KESQTLPTAA
Nvi lab		PNMTNF-NK-LT					
Bfl Hox1		PNNTNF-TK-LT					
Dme pb	/EFVPENGL	PR-L-TANT-L-	K-	-C-PAS	-DV-V-	H- <mark>R</mark> QT	LSKTDDEDNKI
Pca pb							
Nvi pb	DGGSDGGSTGTGSN						
Bfl Hox2	/NADFLTSPTDQVNS						
Sgr Hox3	SSSPTRPSTGSGPTI	SA-TASQ-LI				K	KAKEAMEAAEA
Pca Hox3	/ ====================================	A-LV		-C-PM-AM	-N		
Lan Hox3		TA-TASA-LV					
Nvi Hox3	SANGGDSTNGEGEKP	SA-TA-NSA-LV				<u>Y</u> DQ	KMKPNSEKEL:
Pvu Hox3 Bfl Hox3	/ CTTEDCE CDCI	CA-TASA-IV		-C-PM-AI			KIIK CCCSCCC
Dme Dfd	/GTTEPGESPGLGGAA						
Pca Dfd	/ NGSIQFGME	PQ-TAH-I-					VPLNIVNAVV
Nvi Dfd	/ A NC A V CTD			·S			VI DNTVN DIO
Bfl Hox4	/ANGAIGID	ST- <mark>TA</mark> <u>H</u> -V- TS- <u>TA</u> Q-V-			-C	D-	BIDNEKE BC
Dme Scr	/TVNANG E	TQ-TS					KMASMNIVPYE
Pca HB1	/ I VIIIII . L	1 0 10				<del>H</del>	
Lan Scr	/GM <mark>NG</mark> TE	ST-TSH	G		-N		KLAHLTKTEAN
Nvi Scr	/ANGVE	ST-TSH			- N	<del></del>	KLGHLAKSOG
Bfl Hox5	/AGTGD	NT-TA					KLKSLSOCOO
Dme ftz	/CKD	S		IDN-	-S-S	SDR	TLDSSPEHCG
Dme Antp	/						KTKGEPGSGGI
Pca HB2	/ D			NM		C	KARLIPNGDF
Pca HB3	/	-RF		NM		K	QAYGATCSTE'
Lan HB1	/PD	HS-H					KGIELSRDSP:
Nvi HB1				S			
Pvu HB1			- X - K -		-N		
Pvu HB2							
Pvu HB3				VM			
Lsa Hox7	/ PD						
Hro Lox5		QT	YS	S	-A-S		NVQKLTGPGG\
Lan Lox5	/IGYE	QT					
Nvi Lox5		QT					
Lsa Hox6 Bfl Hox6	/NANRE				<u>-G</u>		NTOCINATORI
Bfl Hox7	/ MGEE / DE	K		KI	-G		KIESTNAIIII
Bfl Hox8	/ PE	-RS	K-			T7	VMICDDKVELI
Dme Ubx	/TNGL.	- <u>R</u>	T_H_	M		T	OAIKELNEOE
Aka Ubx	/ANGLQ		T-H-			M	OTIKDINEOE
Pca Ubx		-R					
Dme abdA							
Aka abdA							
		-R	H-	V	·	L	RAVKEINEQAI
Hme Lox2	/TNGCP	-R	H-	V	'	L L	RAVKEINEQA: RAVKEINEQA:
Hme Lox2 Lan Lox2	/TNGCP	-R	<u>K</u>	V <mark>LS</mark> -T	' '-Y	L L	RAVKEINEQA: RAVKEINEQA:
	/TNGCP	-R	<u>K</u>	V	 -Y [	L L	RAVKE INEQAI RAVKE INEQAI
Lan Lox2	/T <mark>NGCP</mark> /PNSNQ	-R	K  -K K				RAVKEINEQA RAVKEINEQA QAIRELNEIE:
Lan Lox2 Nvi Lox2	/TNGCP /PNSNQ PNSSQ	-RS			-Y    -C	L EL	RAVKEINEQA RAVKEINEQA QAIRELNEIE: QAIKELNAQT QQIKELNEVG
Lan Lox2 Nvi Lox2 Pvu Lox2 Hme Lox4 Lan Lox4	/TNGCP /PNSNQ PNSSQ	-RS -RS	H- K -K K Q		-Y    -C	L L E K K	RAVKEINEQA RAVKEINEQA QAIRELNEIE: QAIKELNAQT QQIKELNEVG QQIKEINDTL
Lan Lox2 Nvi Lox2 Pvu Lox2 Hme Lox4 Lan Lox4 Nvi Lox4	/TNGCP /PNSNQ PNSSQ	-RS -RS	H- K -K K Q R-H-		-Y -C	LLELEL	RAVKEINEQA RAVKEINEQA QAIRELNEIE QAIKELNAQT QQIKELNEVG QQIKEINDTL LQIKELNRDI
Lan Lox2 Nvi Lox2 Pvu Lox2 Hme Lox4 Lan Lox4	/TNGCP /PNSNQ PNSSQ	-RS -RS	H- K -K K Q R-H-		-Y -C 	L L E K K	RAVKEINEQA RAVKEINEQA QAIRELNEIE QAIKELNAQT QQIKELNEVG QQIKEINDTL LQIKELNRDI
Lan Lox2 Nvi Lox2 Pvu Lox2 Hme Lox4 Lan Lox4 Nvi Lox4 Pvu Lox4	/TNGCP /PNSNQ PNSSQ /PNSAQ	-RS -RS-F	H- K 	- V - LS - T - LS - T - LS - T - LS - T - LS - M	-Y		RAVKEINEQA RAVKEINEQA QAIRELNEIE  QAIKELNAQT QQIKELNEVG QQIKEINDTL LQIKELNRDI QAIKDINGDT
Lan Lox2 Nvi Lox2 Pvu Lox2 Hme Lox4 Lan Lox4 Nvi Lox4 Pvu Lox4	/TNGCP /PNSNQ PNSSQ /PNSAQ	-RSS-F	HHHHHHHH	V LS - V - LS - T - LS - T - LS - T - LS - M - LS - M - LS - M - C - M - M	-Y -С н	LLEVELELRM-R-R	RAVKEINEQA RAVKEINEQA QAIRELNEIE  QAIKELNAQT QQIKELNEVG QQIKEINDTL LQIKELNRDI QAIKDINGDT
Lan Lox2 Nvi Lox2 Pvu Lox2 Hme Lox4 Lan Lox4 Nvi Lox4 Pvu Lox4	/TNGCP /PNSNQ  PNSSQ /PNSAQ NPGLHEWT.GQVS /YNPL.EWT.SNVS	-RSRS-FRS-F VRKK-KP-SKF VRKK-KP-K	HH		-Y		RAVKEINEQA RAVKEINEQA QAIRELNEIE  QAIKELNAQT QQIKELNEVG QQIKEINDTL LQIKELNRDI QAIKDINGDT  QRQANQQNNN QKETEKQRQH
Lan Lox2 Nvi Lox2 Pvu Lox2 Hme Lox4 Lan Lox4 Nvi Lox4 Pvu Lox4 Dme AbdB Pca AbdB Cel Y75B87	/TNGCP /PNSNQ  PNSSQ /PNSAQ NPGLHEWT.GQVS /YNPL.EWT.SNVS A.1/GGQL.EWTSSSHA	-R	H	VSKQK-W-L-RYVSKQK-W-L-RY			RAVKEINEQA RAVKEINEQA QAIRELNEIE: QAIKELNAQTI QQIKEINEVG QQIKEINDTL LQIKEINRDI QAIKDINGDTI QROANQQNNNI QKETEKQRQH QRTSGDPTGM
Lan Lox2 Nvi Lox2 Pvu Lox2 Hme Lox4 Lan Lox4 Nvi Lox4 Pvu Lox4 Dme AbdB Pca AbdB Cel Y75B88 Pca HB4	/TNGCP /PNSNQ  PNSSQ /PNSAQ NPGLHEWT.GQVS /YNPL.EWT.SNYS A.1/GGQL.EWTSSSHA /ETPPARHRASF	-RSRS-FRS-F VRKK-KP-SKF VRKK-KPKA VRSK-KP-SKQ-IV	HKK	VSKQK-W-L-KY VSKQK-W-L-KY VSKQK-W-L-KY VSKQK-W-L-KY VSKQK-W-L-KY VSKQK-W-L-KY VSKQK-W-L-KY VSKQK-W-L-KY	-V		RAVKEINEQA RAVKEINEQA QAIRELNEIE QAIKELNAQT QQIKELNEVE QQIKEINDTL LQIKELNRDT QAIKDINGDT QAIKDINGDT QROANQQNNNI QKETEKQRQHQ QRTSGDPTGM LTLRV*
Lan Lox2 Nvi Lox2 Pvu Lox2 Hme Lox4 Nvi Lox4 Pvu Lox4 Dme AbdB Pca AbdB Cel Y75B87 Cel Y75B87	/TNGCP /PNSNQ  PNSSQ /PNSAQ NPGLHEWT.GQVS /YNPL.EWT.SNVS A.1/GGQL.EWTSSNA /ETPPARHRASF A.2AMHLPWAISHD	-RSRS-FRS-F VRKK-KP-SKF VRKK-KPK VRSK-KP-KA VRSK-KP-SKQ-IV G-KKP-KKD-IS	H	VSKQK-W-L-RY VSKQK-W-L-RY VSKQK-W-L-RY VSKQK-W-S VSKQK-W-N-N VSKQK-S-S-S-S-S-S-S-S-S-S-S-S-S-S-S-S-S-S	-Y		RAVKEINEQA RAVKEINEQA QAIRELNEIE QAIKELNAQT QQIKELNEVG QQIKEINDTL LQIKELNRDI QAIKDINGDT QRQANQQNNNI QKETEKQRQH QRTSGDPTGM LTLRV* QRHSGPFPHG
Lan Lox2 Nvi Lox2 Pvu Lox4 Lan Lox4 Nvi Lox4 Pvu Lox4 Pvu Lox4 Pca AbdB Cel Y75B87 Cel Y75B87 Lan Post1	/TNGCP /PNSNQ  PNSSQ /PNSAQ NPGLHEWT.GQVS /YNPL.EWT.SNVS A.1/GGQL.EWTSSSHA /ETPPARHRASF A.2AMHLPWAISHD TTLPAVIH	-RSRSRS VRKK-KP-SKF VRKK-KP-K WRKK-KP-KA VRSK-KP-SKQ-IV MRKK-KP-SK-IA	HK	VSKQK-W-L-RY	-Y		RAVKEINEQAI RAVKEINEQAI QAIRELNEIE  QAIKELNAQTI QQIKELNEVGI QQIKEINDTL LQIKELNRDI QAIKDINGDTI  QRQANQQNNNI QKETEKQRQHQ QRTSGDPTGMI LTLRV* QRHSGPFPHGAI GGKQT*
Lan Lox2 Nvi Lox2 Pvu Lox2 Hme Lox4 Lan Lox4 Nvi Lox4 Pvu Lox4 Dme AbdB Pca AbdB Cel Y75B87 Pca HB4 Cel Y75B87 Lan Post1 Nvi Post1	/TNGCP /PNSNQ  PNSSQ /PNSAQ NPGLHEWT.GQVS /YNPL.EWT.SNVS A.1/GGQL.EWTSSSHA /ETPPARHRASF A.2AMHLPWAISHD TTLPAVIH	-RSRS-FRS-FN-KK-KP-SKF VRKK-KP-KA VRSK-KP-KA VRSK-KP-SKQ-IV G-KK-P-KKD-IS MRKK-KP-SK-IA	HKKQH	V			RAVKEINEQAI RAVKEINEQAI QAIRELNEIEI QAIKELNAQTI QQIKELNEVGI QQIKEINDTLI LQIKELNRDI' QAIKDINGDTI QRQANQQNNNI QKETEKQRHG QRTSGDPTGM: LTLRV* QRHSGPFPHGGGGKQT* DKKCDD
Lan Lox2 Nvi Lox2 Pvu Lox2 Hme Lox4 Lan Lox4 Nvi Lox4 Pvu Lox4  Dme AbdB Pca AbdB Cel Y75B8F Pca HB4 Cel Y75B8F Lan Post1 Lvi Post1 Lan Post2	/TNGCP /PNSNQ  PNSSQ /PNSAQ NPGLHEWT.GQVS /YNPL.EWT.SNVS A.1/GQQL.EWTSSSHA /ETPPARHRASF A.2AMHLPWAISHD TTLPAVIHTPRHINIGPTILH	-RSRS-FRS-F VRKK-KP-SKF WRKK-KP-KA WRSK-KP-SKQ-IV G-KK-P-KKD-IS MRKK-KP-SK-IA MRKK-KP-SK-IA	H	VSKQK-W-L-RY VSKQK-W-L-KY SKQK-W-L-KY SKQK			RAVKEINEQA RAVKEINEQA QAIRELNEIE QAIKELNAQT QQIKELNEVG QQIKEINDTL LQIKELNRDI QAIKDINGDT QRQANQQNNNI QKETEKQRHQ QRTSGDPTGM LTLRV* QRHSGPFPHG GGKQT* DKKCDD ERAKALFKSE
Lan Lox2 Nvi Lox2 Pvu Lox4 Lan Lox4 Lox4 Lox4 Pvu Lox4 Pvu Lox4 Pvu Lox4 Lox4 Lox4 Pvu Lox4 Lox4 Lox4 Lox4 Lox4 Lox4 Lox4 Lox4	/TNGCP /PNSNQ  PNSSQ /PNSAQ NPGLHEWT.GQVS /YNPL.EWT.SNVS A.1/GGQL.EWTSSSHA A.2AMHLPWAISHD TTLPAVIHTPRHINIGPTILH	-RSRSRSRSRSRSRSRSR	H	VSKQK-W-L-RYVSKQK-W-L-KYSLSQF K-S-S-S-S-S-S-S-S-S-S-S-S-S-S-S-S-S-S			RAVKEINEQA RAVKEINEQA QAIRELNEIE  QAIKELNAQT QQIKELNEVG QQIKEINDTL LQIKELNRDII QAIKDINGDTI QKETEKQRQHI QRTSGDPTGM LTLRV* QRHSGPFPHG GGKQT* DKKCDD ERAKALFKSE ERAKTLIKSD
Lan Lox2 Nvi Lox2 Pvu Lox2 Hme Lox4 Lan Lox4 Pvu Lox4 Pvu Lox4 Pvu Lox4  Dme AbdB Cel Y75B87 Cal HB4 Cel Y75B87 Lan Post1 Nvi Post1 Lan Post2 Lsa Hox9	/TNGCP /PNSNQ  PNSSQ /PNSAQ NPGLHEWT.GQVS /YNPL.EWT.SNVS A.1/GGQL.EWTSSSHA /ETPPARHRASF A.2AMHLPWAISHD TTLPAVIHTPRHINIGPTILH /QPR /PR	-RSRS-FRS-FN	HK	VSKQK-W-L-RY VSKQK-W-L-RY VSKQK-W-L-S-LSQG ISKPK-W-LSQG I-KPK-W-SCK I-QK-W-SCK I-QK-W-SCK I-QK-W-SCK I-QK-W-SCK I-QK-W-SCK I-QK-W-SCK I-QK-W-SCK I-QK-W-SCK			RAVKEINEQA RAVKEINEQA QAIRELNEIE  QAIKELNAQT QQIKELNEVG QQIKEINDTL LQIKEINBDT QAIKDINGDT  QRQANQQNNNI QKETEKQRQH QRTSGDPTGM LTLRV* QRHSGPFFHG. GGKQT* DKKCDD ERAKALFKSE ERAKTLIKSD. ARSKVKTRMT
Lan Lox2 Nvi Lox2 Pvu Lox2 Hme Lox4 Lan Lox4 Pvu Lox4 Pvu Lox4  Dme AbdB Pca AbdB Cel Y75B8F Ca HB4 Cel Y75B8F Lan Post1 Nvi Post2 Nvi Post2 Lsa Hox9 Cel egl-5	/TNGCP /PNSNQ  PNSSQ /PNSAQ NPGLHEWT.GQVS /YNPL.EWT.SNVS A.1/GGQL.EWTSSSHA /ETPPARHRASF A.2AMHLPWAISHD TTLPAVIHTPRHINIGPTILF /QPR /PR /S	-RSRS-FRS-FN-KK-KP-KF WRKK-KP-KA WRSK-KP-SKQ-IV G-KK-P-KKD-IS MRKK-KP-SK-IA /-KP	HK	VSKQK-W-L-RN VSKQK-W-L-RY VSKQK-W-L-RY T-Q-F-M-N -NK-S-LSAC T-KFK-W-LSQF T-KFK-W-LSQF T-QK-W-SCK T-QK-W-SCK VSKKQ-E-LRLC VSKKK-S-LRLC VSKQK-W-LSQF T-KFK-W-LSQF T-KFK-W-LSQF T-KFK-W-LSQF T-KFK-W-SCK T-QK-W-SCK T-QK-W-SCK	-V		RAVKEINEQAI RAVKEINEQAI RAVKEINEQAI QAIRELNEIE  QAIKELNAQT QQIKELNEVGI QQIKEINDTLI LQIKELNRDI' QAIKDINGDTI  QRQANQQNNNI QKETEKQRQHQ QRTSGDPTGM: LTLRV* QRHSGPFPHG: GGKQT* DKKCDD ERAKALFKSEI ERAKTLIKSD: ARSKVKTRMTI QRVDDHTEHT:
Lan Lox2 Nvi Lox2 Pvu Lox2 Hme Lox4 Lan Lox4 Nvi Lox4 Pvu Lox4  Dme AbdB Pca AbdB Cel Y75B8F Pca HB4 Cel Y75B8F Lan Post1 Nvi Post1 Lan Post2 Nvi Post2 Lsa Hox9 Cel eg1-5 Bf1 Hox9	/TNGCP /PNSNQ  PNSSQ /PNSAQ NPGLHEWT.GQVS /YNPL.EWT.SNVS A.1/GGQL.EWTSSNAA /ETPPARHRASF A.2AMHLPWAISHD TTLPAVIHTPRHINIGPTILH /QPR /PR /PR /S	-RSRS-FRS-FRS-FR	HK	VSKQK-W-L-RN VSKQK-W-L-RT VSKQK-W-L-RT VSKQK-W-L-ST I-Q-F-M-N -NK-S-LSAC ISKPK-W-LSQF I-QK-W-SCK I-QK-W-SCK VSKKQE-LRLC VSKQK-W-L-SY I-QK-W-SCK VSKKQE-LRLC VSKKQE-LRLC			RAVKEINEQA RAVKEINEQA QAIRELNEIE  QAIKELNAQT QQIKELNEVG QQIKEINDTL LQIKELNRDII QAIKDINGDTI QKETEKQRQHI QRTSGDPTGM LTLRV* QRHSGPFPHG, GGKQT* DKKCDD ERAKALFKSE ERAKTLIKSD ARSKVKTRMT QRVDDHTEHT KQRQEQQQSP
Lan Lox2 Nvi Lox2 Hme Lox4 Lan Lox4 Nvi Lox4 Pvu Lox4 Pvu Lox4 Pvu Lox4  Dme AbdB Cel Y75B87 Cel Y75B87 Lan Post1 Nvi Post1 Lan Post2 Nvi Post2 Lsa Hox9 Mmu HoxA9	/TNGCP /PNSNQ  PNSSQ /PNSAQ NPGLHEWT.GQVS /YNPL.EWT.SNVS A.1/GQQL.EWTSSSHA /ETPPARHRASF A.2AMHLPWAISHD TTLPAVIHTPRHINIGPTILH /QPR /PR /SVVANQPGWMNNS	-RSRS-FRS-FRS-FNRKK-KP-SKF VRKK-KP-KA VRSK-KP-SKQ-IV G-KK-P-KND-IS MRKK-KP-SK-IA /-KP MRKK-KP-SK-IA /-KP	H	VSKQK-W-L-RY VSKQK-W-L-RY VSKQK-W-L-RY VSKQK-W-L-SY I-QK-W-LSQF I-KPK-W-LSQF I-QK-W-SCK I-QK-W-SCK VSKKQ-W-SCK VSKKQ-W-SCK VSKKQ-W-SCK VSKKQ-W-SCK VSKKQ-S-LRLC VSKY-SQF VSKKQ-S-LRLC VSKKQ-SQF			RAVKEINEQAI RAVKEINEQAI RAVKEINEQAI QAIRELNEIE  QAIKELNAQTI QQIKEINEU QQIKEINEU QQIKEINEU QAIKDINGDTI QRQANQQNNNI QKETEKQRQHI QKETEKQRQHI QKTSGDPTGMI LTLRV* QRHSGPFFHGMI GGKQT* DKKCDD ERAKALFKSEI ERAKTLIKSDI ARSKVKTRMTI QRVDDHTEHTI KQRQEQQSPI KQRQEQQSPI KQRQEQQSPI
Lan Lox2 Nvi Lox2 Pvu Lox2 Hme Lox4 Lan Lox4 Pvu Lox4 Pvu Lox4 Pvu Lox4  Dme AbdB Pca AbdB Cel Y75B8F Pca HB4 Cel Y75B8F Lan Post1 Nvi Post2 Lsa Hox9 Cel egl-5 Bfl Hox9 Mmu HoxA9 Bfl Hox10	/TNGCP /PNSNQ  PNSSQ /PNSAQ NPGLHEWT.GQVS /YNPL.EWT.SNVS A.1/GGQL.EWTSSSHA /ETPPARHRASF A.2AMHLPWAISHD TTLPAVIHTPRHINIGPTILH /QPR /PR /SVVANQPGWMNNHS /SVVANQPGWMNHS /SCGATSWMAPRV	-RSRS-FRS-FRS-FN-KK-KP-SKFVRKK-KP-KAVRSK-KP-SKQ-IV G-KK-P-KKD-IS MRKK-KP-SK-IA /-KP QRKK-KP S-K	HK	V			RAVKEINEQAI RAVKEINEQAI RAVKEINEQAI QAIRELNAGTI QAIKELNAGTI QQIKELNEVGI QQIKEINDTLI LQIKEINDTLI LQIKEINDTLI QAIKDINGDTI QAIKDINGDTI QAIKDINGDTI CRETEKQRHG QRTSGDPTGM: LTLRV* QRHSGPFPHG: GGKQT* DKKCDD ERAKALFKSE: ERAKTLIKSD: ARSKVKTRMTI QRVDDHTEHT: KQRQEQQQSP: KDRAKDE* KAREEQIRNQG
Lan Lox2 Nvi Lox2 Pvu Lox2 Hme Lox4 Lan Lox4 Pvu Lox4 Pvu Lox4 Pvu Lox4  Dme AbdB Pca AbdB Cel Y75B8F Pca HB4 Cel Y75B8F Lan Post1 Nvi Post2 Lsa Hox9 Cel egl-5 Bfl Hox9 Mmu HoxA9 Bfl Hox10	/TNGCP /PNSNQ  PNSSQ /PNSAQ NPGLHEWT.GQVS /YNPL.EWT.SNVS A.1/GGQL.EWTSSSHA /ETPPARHRASF A.2AMHLPWAISHD TTLPAVIHTPRHINIGPTILH /QPR /PR /SSVVANQPGWMNNHS /NPAANWLHARS /SCGATSWMAPRV OSKGENAANWLTAKS	-RSRS-FRS-FRS-FN-KK-KPKFN-KK-KPKAN-KK-KPKAN-KK-KP-SKQ-IV -K	HKKK	V			RAVKEINEQAI RAVKEINEQAI QAIRELNEIE  QAIKELNAQTI QQIKELNEVG QQIKEINDTLI LQIKELNRDII QAIKDINGDTI  QRQANQQNNNI QKETEKQRQHO QRTSGDPTGMI LTLRV* DKKCDD ERAKALFKSEI ERAKTLIKSDI ARSKVKTRMTI QRVDDHTEHTI KQRQEQQQSPI KDRAKDE* KAREEQIRNQG
Lan Lox2 Nvi Lox2 Pvu Lox2 Hme Lox4 Lan Lox4 Nvi Lox4 Pvu Lox4  Dme AbdB Pca AbdB Ccl Y75B8F Pca HB4 Ccl Y75B8F Lan Post1 Lan Post2 Nvi Post2 Nvi Post2 Lsa Hox9 Ccl egl-5 Bfl Hox9 Mmu HoxA9 Bfl Hox10 Mmu HoxA10	/TNGCP /PNSNQ  PNSSQ /PNSAQ NPGLHEWT.GQVS /YNPL.EWT.SNVS A.1/GGQL.EWTSSSHA A.2AMHLPWAISHD TTLPAVIHTPRHINIGPTILH  /QPR /PR /PR /S /SVVANQPGWMNHS /NPAANWLHARS /SCGATSWMAPRV /SKGENAANWLTAKS /GQR	-RSRS-FRS-FRS-FN-KK-KP-SKFVRKK-KP-KAVRSK-KP-SKQ-IV G-KK-P-KKD-IS MRKK-KP-SK-IA /-KP QRKK-KP S-K	HK	VSKQK-W-L-RY L-QC-F-M-N LSQF L-KPK-W-LSQF L-KPK-W-LSQF L-QK-W-SCK L-QK-W-SCK VSKKQ-E-LRLC VSKKQ-E-LRLC VSKKQ-E-LRLC SCE-Y-SQF -D-Y-V-RI VS-E-Q-SRS INKEK-LQLSRM			RAVKEINEQAI RAVKEINEQAI RAVKEINEQAI QAIRELNEIEI  QAIKELNAQTI QQIKEINEU QQIKEINEU QQIKEINEU QAIKDINGDTI  QRQANQQNNNI QKETEKQRQH QRTSGDPTGMI LTLRV* QRHSGPFPHGA GGKQT* DKKCDD ERAKALFKSEI ERAKTLIKSDARSKVKTRMTI QRVDDHTEHTI KQRQEQQQSPI KQRQEQQQSPI KDRAKDE* KAREEQIRNQARENIRELTAI RDRLQYYSANI

**Figure 1** Alignment of Hox homeodomains and flanking sequences. Sequences first reported here are indicated by an arrowhead, aligned in comparison to previously published Hox genes. Colour code for arrowheads and dots: green, ecdysozoans; orange, lophotrochozoans; blue, deuterostomes. Within the homeodomain, dashes indicate identity to *Drosophila melanogaster Antp* (at top). Anterior and central genes that can be grouped by orthology across metazoan phyla are enclosed within black boxes (*Dfd* peptide motif highlighted in

pink). Sequences boxed in orange contain the 'Lox6' peptide motif<sup>4</sup> (highlighted in orange) and sequences boxed in blue contain the 'Ubd-A' peptide motif<sup>4</sup> (highlighted in blue). Posterior genes are boxed in grey when orthology is well supported by phylogenetic analysis (see Fig. 2). Yellow highlighting shows conserved residues (>50% majority) that differ from Antp within each group of orthologues. See Table 1 for species abbreviations.

## letters to nature

Table 1 Species names and abbreviations used in Figures							
Common name	Species name	Abbreviation	References				
Amphioxus	Branchiostoma floridae	Bfl	26, 27				
Ascidian	Ciona intestinalis	Cin	*				
Brachiopod	Lingula anatina	Lan	This paper				
Crustacean	Artemia franciscana	Afr	*				
Flatworms	Girardia tigrina	Gti	*				
	Polycelis nigra	Pni	*				
Fruit fly	Drosophila melanogaster	Dme	18				
Gastropod	Patella vulgata	Pvu	This paper				
Leeches	Helobdella robusta	Hro	*				
	Helobdella triserialis	Htr	*				
	Hirudo medicinalis	Hme	*				
Locust	Schistocerca gregaria	Sgr	*				
Mouse	Mus musculus	Mmu	28				
Nematode	Caenorhabditis elegans	Cel	19, 29				
Nemertean	Lineus sanguineus	Lsa	14				
Onychophoran	Acanthokara kaputensis	Aka	5				
Polychaete	Nereis virens	Nvi	This paper				
Priapulid	Priapulus caudatus	Pca	This paper				
Sea urchins	Strongylocentrotus purpuratus	Spu	30				
	Heliocidaris erythrogramma	Her	*				
	Tripneustes gratilla	Tgr	*				

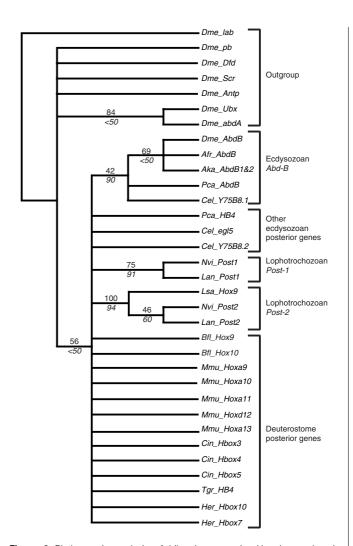
<sup>\*</sup> See Supplementary Information for complete reference list and accession numbers.

new tool for addressing long-standing issues in metazoan phylogeny<sup>4,5</sup>. For both these purposes data need to be obtained from a broad sample of phyla, particularly those that help resolve the early branching pattern of the bilaterians. For example, if the proposed assemblages of lophotrochozoan phyla (lophophorates, annelids, molluscs, flatworms and others) and ecdysozoan phyla (arthropods, onychophorans, nematodes, priapulids and others) are correct<sup>1,2,6</sup>, Hox gene identities ought to reflect this phylogeny.

We have identified Hox genes from two previously unsampled phyla, a lophophorate (the brachiopod *Lingula anatina*) and a priapulid (*Priapulus caudatus*). Brachiopods were once thought to be deuterostomes<sup>7</sup>, whereas priapulids were often associated with the other 'pseudocoelomates' basal to the great majority of bilaterian phyla<sup>8</sup>. Both these phyla are among the most ancient bilaterians to appear in the Cambrian fossil record<sup>9</sup>. We have also collected an extensive Hox gene data set from a polychaete annelid (*Nereis virens*), and partial data from a gastropod mollusc (the limpet *Patella vulgata*).

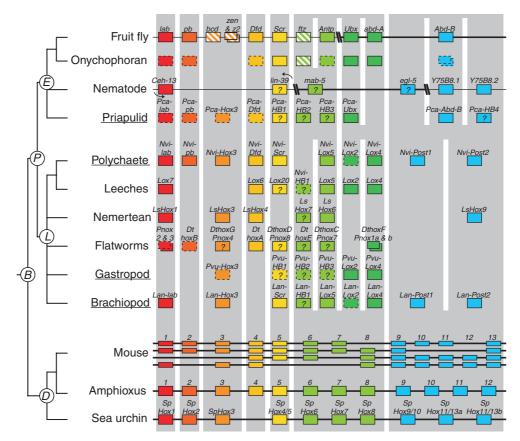
Many protostome Hox genes can clearly be assigned to one of five orthology groups that are also recognizable in deuterostomes: labial (lab)/Hox1, proboscipedia (pb)/Hox2, zerknüllt (zen)/Hox3, Deformed (Dfd)/Hox4 and Sex combs reduced (Scr)/Hox5. These orthology groups are defined by conserved residues unique to each group that are shared across bilaterian phyla, and they are supported by phylogenetic analysis (Fig. 1 and Supplementary Information). The remaining protostome Hox genes cannot be identified as orthologues of specific deuterostome genes. Many can, however, be placed into groups on the basis of conserved amino acids within the homeodomain and distinct peptide motifs in the flanking sequences (Fig. 1) that are found in subsets of protostome phyla. In this way, we identify three conserved central (Antennapedia-like) genes shared by the brachiopod, the polychaete annelid and the gastropod mollusc that have been previously characterized in annelids and named Lox5 (ref. 10), Lox2 (ref. 12) and Lox4 (refs 12, 13) (we continue to use these names until unambiguous linkage data allows rationalization of the nomenclature). Similarly, on the basis of conserved residues within and flanking the homeodomain, we group one of the priapulid central Hox genes with the *Ultrabithorax* (*Ubx*) gene. This gene has been described previously only in arthropods and onychophorans<sup>5</sup>. Additional central genes found in each organism cannot be clearly assigned to an orthology group because sequence divergence obscures relationships across phyla. Some of these may be new genes that have arisen from more recent gene duplication events.

Previous studies of protostome Hox genes have identified at most



**Figure 2** Phylogenetic analysis of bilaterian posterior Hox homeodomain sequences. Numbers above branches indicate per cent support in MP bootstrap analyses. Italicized numbers below branches indicate per cent reliability in a Puzzle ML tree. Note high support for the ecdysozoan *Abd-B* genes, the lophotrochozoan *Post-1* genes, and the lophotrochozoan *Post-2* genes. Nodes with low support (<50%) in all analyses have been collapsed to polytomies. See Supplementary information for phylogenetic analyses that include *caudal/cdx* genes.

a single posterior (Abdominal-B-like) gene<sup>5,14–18</sup>, but we have found multiple posterior genes in several protostome phyla. We discovered two posterior genes in the lophophorate, the polychaete and the priapulid, and the C. elegans genome sequencing project has identified two previously uncharacterized posterior genes (in addition to egl-5) in this nematode<sup>19</sup> (Fig. 1). On the basis of the homeodomain sequence, the brachiopod and polychaete genes fall into two new orthology groups that we term Post-1 and Post-2; Post-2 has also been found in the nemertean Lineus sanguineus (LsHox9; ref. 14) and a fragment of it in an oligochaete<sup>15</sup>. One of the priapulid posterior genes and one of the newly sequenced nematode genes (Y75B8A.1) are orthologous to the arthropod Abdominal-B (Abd-B) gene (Fig. 1). These orthology assignments are well supported by phylogenetic analyses of the homeodomain sequences (Fig. 2). However, none of the posterior Hox genes can be individually related to any of the deuterostome genes, as shown by the basal polytomy in Fig. 2; nor are any of these genes orthologous to the ParaHox caudal/cdx genes<sup>20</sup> (see Supplementary Information). Sequence divergence has obscured the origins of these posterior genes, and it is not possible to determine whether they arose from



**Figure 3** Distribution of Hox genes in bilaterians. The tree on the left summarizes the 18S rDNA phylogeny of Aguinaldo *et al.*?. B, common bilaterian ancestor; D, common deuterostome ancestor; E, stem ecdysozoan; L, stem lophotrochozoan; P, common protostome ancestor. Horizontal black lines indicate mapping data (when available). Vertical white bars delineate orthologous genes or groups of

genes. Uncertain orthology relationships are indicated by question marks, dashed boxes indicate short PCR fragments, solid boxes indicate complete or almost complete homeoboxes, and striped colours indicate fast-evolving arthropod Hox sequences. The organisms studied in this paper are underlined.

independent duplications in each of the three clades, or whether multiple posterior Hox genes were inherited from a common bilaterian ancestor.

The identification of distinct Hox genes that are shared by some, but not all, protostome phyla constitutes independent molecular evidence for the division of protostomes into two great clades—the lophotrochozoans and the ecdysozoans. The animals within each of these lineages have central and posterior Hox genes specific to their clade, which provides a surprisingly strong signal of the deep division within the protostomes. The lophotrochozoans are characterized by the presence of Lox5, Lox2, Lox4, Post-1 and Post-2, and the ecdysozoans by Ubx and Abd-B. Each of these genes can be identified by characteristic shared, derived residues in and near the homeodomain, which became fixed before the radiation of the crown phyla of each lineage. Not all of the proposed ecdysozoans definitively have all of the 'ecdysozoan' Hox genes (most notably C. elegans), but the unambiguous identification of an 'ecdysozoan' Hox gene in any animal, such as the C. elegans Abd-B gene (Y75B8A.1), supports the inclusion of that animal within the ecdysozoans. Thus, these shared sets of Hox genes provide signatures for each protostome clade, and can be used to infer the phylogenetic placement of other protostome phyla.

A comparison of the Hox genes that are shared by descendants of a particular lineage enables estimation of the number of Hox genes present in earlier animal ancestors. On the basis of shared Hox genes, we conclude that the stem lophotrochozoan (Fig. 3, Node L) had at least ten Hox genes and the stem ecdysozoan (Fig. 3, node E) at least eight. Following the same logic, we infer that the common protostome ancestor (Fig. 3, node P) also had at least eight Hox

genes and the common bilaterian ancestor (Fig. 3, node B) had at least seven (lab/Hox1, pb/Hox2, Hox3, Dfd/Hox4, Scr/Hox5, one additional central gene and one posterior gene). Indeed, seven Hox genes is a minimum estimate for the bilaterian ancestor, and the true number might have been higher. As more phyla have been sampled, we are struck by the fact that all clusters (except for the degenerate C. elegans cluster) contain nine or more genes. This indicates that most or all of the Hox genes that are present in extant bilaterians (Fig. 3) may have been present in the common ancestor, but that some orthology relationships have become obscured. If this new 'early expansion' hypothesis is true, then at least ten Hox genes could have existed in the protostome ancestor (lab/Hox1, pb/Hox2, Hox3, Dfd/Hox4, Scr/Hox5, three to four central genes and at least two posterior genes), or even more if the deuterostome condition of multiple posterior Hox genes is ancestral. The subsequent bilaterian history of Hox genes would have been primarily one of functional divergence and gene loss, rather than gene duplication. Regardless of the exact number of Hox genes in the bilaterian ancestor, the major period of progressive expansion of the Hox cluster due to tandem duplication events<sup>21</sup> predated the radiation that generated the bilaterian crown phyla, concurrent with radical evolutionary changes in body architecture and development.

#### Methods

L. anatina DNA was made available by the authors of a previous study<sup>22</sup>. N. virens specimens were collected in the Marine Biological Station of St Petersburg University, Chupa Inlet, White sea, Russia. P. vulgata specimens were collected in the Station Biologique, Roscoff, France. P. caudatus specimens were collected from Lopez Sound, Washington, USA and from Kasitsna Bay, Alaska,

### letters to nature

USA. Short homeobox fragments were amplified by PCR from genomic DNA using degenerate primers (see Supplementary Information for details). Homeodomain sequences were extended using either ligation mediated-PCR<sup>5,23</sup> or inverse-PCR<sup>17</sup>. Phylogenetic trees were built from aligned complete homeodomain sequences using unweighted maximum parsimony (MP) with PAUP\*4.0b1 (ref. 24) and maximum likelihood (ML) with Puzzle<sup>25</sup>. The MP analysis was performed with the following settings: heuristic search over 100 bootstrap replicates, MAXTREES set to 10,000 due to computer limitations, other parameters set to default values. The ML analysis was performed using the quartet puzzling tree search procedure with 10,000 puzzling steps and using a gamma-distributed model of rate heterogeneity with eight categories.

Received 11 February 1999; accepted 20 April 1999.

- Halanych, K. M. et al. Evidence from 18S ribosomal DNA that the lophophorates are protostome animals. Science 267, 1641–1643 (1995).
- Aguinaldo, A. M. et al. Evidence for a clade of nematodes, arthropods and other moulting animals Nature 387, 489–493 (1997).
- Carroll, S. B. Homeotic genes and the evolution of arthropods and chordates. Nature 376, 479

  –485
  (1995).
- Balavoine, G. Are platyhelminthes coelomates without a coelom? An argument based on the evolution of Hox genes. Am. Zool. 38, 843–858 (1998).
- Grenier, J. K., Garber, T. L., Warre, R., Whitington, P. M. & Carroll, S. Evolution of the entire arthropod *Hox* gene set predated the origin and radiation of the onychophoran/arthropod clade. *Curr. Biol.* 7, 547–553 (1997).
- Adoutte, A., Balavoine, G., Lartillot, N. & de Rosa, R. Animal evolution: the end of the intermediate taxa? Trends Genet. 15, 104–108 (1999).
- 7. Nielsen, C. Animal Evolution (Oxford Univ. Press, Oxford, 1995).
- Lorenzen, S. in The Origins and Relationships of Lower Invertebrates (ed. Conway-Morris, S.) 210–223 (Clarendon, Oxford, 1985).
- Boardmann, R. S., Cheetham, A. H. & Rowell, A. J. Fossil Invertebrates (Blackwell Scientific, Oxford, 1987)
- Shankland, M., Martindale, M. Q., Nardelli-Haefliger, D., Baxter, E. & Price, D. J. Origin of segmental identity in the development of the leech nervous system. *Development* (Suppl. 2) 29–38 (1991).
- Wysocka-Diller, J. W., Aisemberg, G. O., Baumgarten, M., Levine, M. & Macagno, E. R. Characterization of a homologue of bithorax-complex genes in the leech Hirudo medicinalis. Nature 341, 760– 763 (1989).
- Dick, M. H. & Buss, L. W. A PCR-based survey of homeobox genes in Ctenodrilus serratus (Annelida: Polychaeta). Mol. Phylogenet. Evol. 3, 146–158 (1994).
- Wong, V. Y., Aisemberg, G. O., Gan, W. B. & Macagno, E. R. The leech homeobox gene Lox4 may determine segmental differentiation of identified neurons. J. Neurosci. 15, 5551–5559 (1995).
- Kmita-Cunisse, M., Loosli, F., Bièrne, J. & Gehring, W. J. Homeobox genes in the ribbonworm *Lineus sanguineus*: evolutionary implications. *Proc. Natl Acad. Sci. USA* 95, 3030–3035 (1998).
- Snow, P. & Buss, L. W. HOM/Hox type homeoboxes from Stylaria lacustris (Annelida: Oligochaeta). Mol. Phylogenet. Evol. 3, 360–364 (1994).
- Degnan, B. M. & Morse, D. E. Identification of eight homeobox-containing transcripts expressed during larval development and at metamorphosis in the gastropod mollusc *Haliotis rufescens*. Mol. Mar. Biol. Biotechnol. 2, 1–9 (1993).
- Averof, M. & Akam, M. HOM/HOX genes in a crustacean: implications for the origin of insect and crustacean body plans. Curr. Biol. 3, 73–78 (1993).
- Akam, M. Hox and HOM: homologous gene clusters in insects and vertebrates. Cell 57, 347–349 (1989).
- Ruvkun, G. & Hobert, O. The taxonomy of developmental control in *Caenorhabditis elegans. Science* 282, 2033–2041 (1998).
- Brooke, N. M., Garcia-Fernàndez, J. & Holland, P. W. H. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. *Nature* 392, 920

  –922 (1998).
- Schubert, F. R., Nieselt-Struwe, K. & Gruss, P. The Antennapedia-type homeobox genes have evolved from three precursors separated early in metazoan evolution. Proc. Natl Acad. Sci. USA 90, 143–147 (1993)
- Mackey, L. Y. et al. 18S rNA suggests that Entoprocta are protostomes, unrelated to Ectoprocta. J. Mol. Evol. 42, 552–559 (1996).
- Balavoine, G. Identification of members of several homeobox gene classes in a planarian using a ligation-mediated polymerase chain reaction technique. Nucleic Acids Res. 24, 1547–1553 (1996).
- Swofford, D. L. PAUP\* Phylogenetic Analysis Using Parsimony (\* and Other Methods), Version 4 (Sinauer, Sunderland, MA, 1998).
- Strimmer, K. & von Haeseler, A. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13, 964–969 (1996).
- Garcia-Fernandez, J. & Holland, P. W. H. Archetypal organization of the amphioxus Hox gene cluster. Nature 370, 563–566 (1994).
- 73. Holland, P. W. H. & Garcia-Fernàndez, J. Hox genes and chordate evolution. Dev. Biol. 173, 382–395 (1996).
- Duboule, D. et al. An update of mouse and human HOX gene nomenclature. Genomics 7, 458–459 (1990).
- Wang, B. B. et al. A homeotic gene cluster patterns the anteroposterior body axis of C. elegans. Cell 74, 29–42 (1993).
- Martinez, P., Rast, J. P., Arenas-Mena, C. & Davidson, E. H. Organization of an echinoderm Hox gene cluster. Proc. Natl Acad. Sci. USA 96, 1469–1474 (1999).

Supplementary information is available on Nature's World-Wide Web site or on paper copy from the London editorial office of Nature.

Acknowledgements. We thank B. Winnepenninckx for L. anatina DNA; A. Knowlton, R. Highsmith, and P. Reynolds for P. caudatus tissue; E. Davidson for his interest and for sharing unpublished data; V. Kassnad N. Lartillot for their help; and A. Friday and G. Budd for comments on the manuscript. This work was supported by grants from the BBSRC and the Wellcome trust (M.E.A.); the Russian Basic Research Foundation and the Royal Society (T.A.); the CNRS, program 'Genome' and the Université Paris-Sud (A.A.); and the Howard Hughes Medical Institute (J.K.G. and S.B.C.).

Correspondence and requests for materials should be addressed to G.B. (e-mail: Guillaume.Balavoine@cgm.cnrs-gif.fr.) Accession numbers: Lan: AF144672–AF144680; Nvi: AF151663–AF151673; Pca AF144884–AF144893; Pvu: AF144666–144671.

## A stop-codon mutation in the BRI gene associated with familial British dementia

Ruben Vidal\*, Blas Frangione\*, Agueda Rostagno\*, Simon Mead†, Tamas Révész‡, Gordon Plant†‡ & Jorge Ghiso\*

- \* Department of Pathology, New York University School of Medicine, 550 First Avenue, New York 10016, USA
- † The National Hospital for Neurology and Neurosurgery, Queen Square, London WC1N 3BG, UK
- ‡ The Department of Neuropathology, Institute of Neurology, Queen Square, London WC1N 3BG, UK

Familial British dementia (FBD), previously designated familial cerebral amyloid angiopathy-British type<sup>1</sup>, is an autosomal dominant disorder of undetermined origin characterized by progressive dementia, spasticity, and cerebellar ataxia, with onset at around the fifth decade of life. Cerebral amyloid angiopathy, non-neuritic and perivascular plaques and neurofibrillary tangles are the predominant pathological lesions<sup>1-4</sup>. Here we report the identification of a unique 4K protein subunit named ABri from isolated amyloid fibrils. This highly insoluble peptide is a fragment of a putative type-II single-spanning transmembrane precursor that is encoded by a novel gene, BRI, located on chromosome 13. A single base substitution at the stop codon of this gene generates a longer open reading frame, resulting in a larger, 277-residue precursor. Release of the 34 carboxy-terminal amino acids from the mutated precursor generates the ABri amyloid subunit. The mutation creates a cutting site for the restriction enzyme XbaI, which is useful for detecting asymptomatic carriers. Antibodies against the amyloid or homologous synthetic peptides recognize both parenchymal and vascular lesions in FBD patients. A point mutation at the stop codon of BRI therefore results in the generation of the ABri peptide, which is deposited as amyloid fibrils causing neuronal disfunction and dementia.

The uninterrupted transmission from one generation to the next and analyses of segregation and sex ratios have previously confirmed the autosomal dominant mode of inheritance of FBD in a large family of more than 200 members encompassing seven generations<sup>1</sup>. However, the biochemical basis of the disorder has remained elusive. It has been reported as an atypical form of familial Alzheimer's disease<sup>5-7</sup>, an example of spongiform encephalopathies<sup>8–10</sup> and a specific form of primary congophilic angiopathy<sup>11</sup>. Classification attempts based on immunohistochemical analysis failed to demonstrate specific staining of the amyloid deposits with a large set of antibodies directed toward known amyloid molecules, although the lesions were immunoreactive for several amyloid-associated proteins (for example, apoliproproteins E and J and serum amyloid P-component<sup>12</sup>). C-terminal fragments of  $\alpha$ - and  $\beta$ -tubulin have been reported to be associated with the amyloid lesions, although the identification of the major component of the amyloid deposits remained unknown<sup>13</sup>.

Using a combination of formic-acid solubilization and gel-filtration chromatography, we have isolated amyloid fibrils from case V41, a female member of this pedigree<sup>1</sup> who developed disease symptoms at age 56 years and died at age 65. Post-mortem examination revealed severe widespread cerebrovascular amyloidosis in the brain and spinal cord, non-neuritic amyloid plaques affecting the cerebellum, hippocampus, amygdala and occasionally the cerebral cortex, changes in the periventricular white matter, perivascular amyloid plaques and neurofibrillar degeneration in hippocampal neurons<sup>1,14</sup>. A prominent component of relative molecular mass 4,000 ( $M_{\rm r}$  4K) was defined in SDS–polyacrylamide gels in preparations from leptomeningeal as well as parenchymal