Development 128, 779-790 (2001)
Printed in Great Britain © The Company of Biologists Limited 2001
DEV9713

Regulation of ectodermal and excretory function by the *C. elegans* POU homeobox gene *ceh-6*

Thomas R. Bürglin[1,2,*,‡] and Gary Ruvkun[1]

[1]Department of Molecular Biology, Massachusetts General Hospital and Department of Genetics, Harvard Medical School, Wellman 8, Fruit Street, Boston, MA 02114, USA
[2]Division of Cell Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland
*Address from April 2001: Department of Biosciences and Center for Genomics Research, Karolinska Institute, Södertörns Högskola, Alfred Nobels Alle 10, Box 4101, SE-141 04 Huddinge, Sweden
‡Author for correspondence (e-mail: Thomas.Burglin@biosci.ki.se)

Accepted 12 December 2000; published on WWW 7 February 2001

SUMMARY

***Caenorhabditis elegans* has three POU homeobox genes, *unc-86*, *ceh-6* and *ceh-18*. *ceh-6* is the ortholog of vertebrate *Brn1*, *Brn2*, *SCIP/Oct6* and *Brn4* and fly *Cf1a/drifter/ventral veinless*. Comparison of *C. elegans* and *C. briggsae* CEH-6 shows that it is highly conserved. *C. elegans* has only three POU homeobox genes, while *Drosophila* has five that fall into four families. Immunofluorescent detection of the CEH-6 protein reveals that it is expressed in particular head and ventral cord neurons, as well as in rectal epithelial cells, and in the excretory cell, which is required for osmoregulation. A deletion of the *ceh-6* locus causes 80% embryonic lethality. During morphogenesis, embryos extrude cells in the rectal region of the tail or rupture, indicative of a defect in the rectal epithelial cells that express *ceh-6*. Those embryos that hatch are sick and develop vacuoles, a phenotype similar to that caused by laser ablation of the excretory cell. A GFP reporter construct expressed in the excretory cell reveals inappropriate canal structures in the *ceh-6* null mutant. Members of the POU-III family are expressed in tissues involved in osmoregulation and secretion in a number of species. We propose that one evolutionary conserved function of the POU-III transcription factor class could be the regulation of genes that mediate secretion/osmoregulation.**

Key words: *ceh-6*, *Caenorhabditis elegans*, *Caenorhabditis briggsae*, Homeobox gene, POU domain, POU-III family, Osmoregulation

INTRODUCTION

Homeobox genes are a group of transcription factors that play important roles in development (for a review see Bürglin, 1995; Gehring et al., 1994). One particular class, the POU class of homeobox genes, was first defined by the transcription factors Pit1, Oct1, Oct2, and the *Caenorhabditis elegans* developmental control gene *unc-86* (Herr et al., 1988). Since the initial identification of the POU homeobox gene class, more than 70 POU homeobox genes have been identified from a wide range of metazoans (Spaniol et al., 1996). They are expressed in patterns that vary from exquisitely cell type specific (for example, *unc-86*; Finney and Ruvkun, 1990; Finney et al., 1988) to ubiquitous (for example, Oct1; Sturm et al., 1988) and have been shown to regulate transcription of genes involved in a variety of processes (for a review see Ryan and Rosenfeld, 1997). Only three *C. elegans* POU homeobox genes can be identified in the virtually complete *C. elegans* genomic sequence (The *C. elegans* Sequencing Consortium, 1998): (1) *unc-86* was identified genetically and shown to be necessary for the specification of particular neuroblasts and neurons (Finney et al., 1988; Finney and Ruvkun, 1990); (2) *ceh-18*, was identified by molecular probing and shown by reverse genetic analysis to play a distinct role in development, specifying ectodermal and gonadal cell fates (Greenstein et al., 1994; Rose et al., 1997); and (3) *ceh-6*, which was detected in degenerate oligonucleotide screens for *C. elegans* homeobox genes (Bürglin et al., 1989), but has not been functionally characterized. In this paper we examine the developmental role of *ceh-6*.

Of the three *C. elegans* POU homeobox genes, *ceh-6* can be assigned to the POU-III family, *unc-86* to the POU-IV family (Bürglin et al., 1989; He et al., 1989) and *ceh-18* to the POU-II family (see below). We show that *ceh-6* is the bona fide ortholog of the *Drosophila* gene *Cf1a* (Johnson and Hirsh, 1990) and of the four vertebrate genes *Brn1*, *Brn2*, *SCIP* (also known as *Pou3f1*, *Tst1* or *Oct6*) and *Brn4* (Hara et al., 1992; He et al., 1989; Mathis et al., 1992; Monuki et al., 1989). The fly gene *Cf1a*, shown to correspond to the genetic locus *drifter* (*dfr*), also known as *ventral veinless* (*vvl*), regulates a range of developmental processes, including tracheal formation (Anderson et al., 1995; de Celis et al., 1995). The vertebrate genes are expressed in many different areas of the nervous system, in particular in the brain and spinal cord (He

et al., 1989). Mouse knockouts of *Brn2* have shown that this gene is required for endocrine hypothalamus and posterior pituitary development (Nakai et al., 1995; Schonemann et al., 1995). In this paper we demonstrate that *ceh-6* is also expressed in and required for the function of several different types of neurons and epithelial cell types, as well as the excretory cell. This orthology suggests an ancient and perhaps common *ceh-6* function in the specification of these seemingly disparate cell types.

MATERIALS AND METHODS

Molecular techniques

Standard molecular techniques were used for cloning and sequencing (Ausubel et al., 1987; Sambrook et al., 1989). The cloning and sequencing of genomic *ceh-6* has been described previously (Bürglin et al., 1989). A cDNA for *ceh-6* was generated using PCR with the primers PCR6-1 (CCCGGGATCCTCCATGGCCACCAGCTT, containing a *Bam*HI site) and PCR6-2 (TTCGAAGCTTCTATTGTTGTCTCGGGCTCT, containing a *Hin*dIII site) from first strand cDNA. Several clones from an embryonic λgt11 library were isolated with this cDNA probe. The sequence of the longest combined cDNA has been submitted to GenBank under Accession Number AF286377.

For antibody production, the sequenced PCR cDNA insert was subcloned into the expression vectors pTRB0 (Bürglin and De Robertis, 1987) and pGEX1 (Smith and Johnson, 1988). Insoluble full-length fusion protein was SDS PAGE gel purified and used to boost rabbits using RIBI as adjuvant. Polyclonal serum was affinity purified with pGEX1-CEH-6 fusion protein transferred to nitrocellulose. The nitrocellulose strip was incubated with 1:10 diluted polyclonal serum, and washed. Antibodies were eluted with 0.1 M glycine pH 2.0 and immediately neutralized with 1M Tris pH 7.5 (Bürglin et al., 1987). Fixation and whole-mount immunolocalization of worms was performed as previously described (Finney and Ruvkun, 1990).

For reporter constructs, a 11.5 kb genomic *Nco*I fragment of cosmid WWG8 was cloned into pPD22.11ΔI (a modified version of pPD22.11 (Fire et al., 1990), having no synthetic intron), giving pLZ6-1. A further variant, pGFP6-1 was constructed by replacing the *lacZ* gene of pLZ6-1 with *gfp* from TU#61 (Chalfie et al., 1994). For rescue experiments, a 17 kb *Bam*HI/*Sal*I fragment of WWG8 was subcloned to yield pTRB462 (Fig. 1). pTRB462 was digested with *Nco*I and re-ligated, yielding clone pTRB464, carrying the internal *Nco*I fragment in the opposite orientation.

Sequence analysis

C. elegans and *C. briggsae* database searches were performed at the Sanger center in Cambridge and the Genome Sequencing Center in St. Louis (http://www.sanger.ac.uk/Projects/C_elegans/; http://genome.wustl.edu/gsc/) using their Web implementation of BLAST (Gish and Warren, unpublished; Altschul et al., 1990). Genomic organization of the gene structures was analyzed using GENEFINDER within the ACeDB database on a local workstation (Durbin and Thierry Mieg, 1991-). Searches of GenBank were performed using Netblast at NCBI or the Web BLAST servers at NCBI (Altschul et al., 1990). Programs of the GCG package were used such as Fetch and Gap in GCG (Devereux et al., 1984). Phylogenetic analysis was performed with CLUSTAL_X 1.6 (Thompson et al., 1997), trees were visualized NJPLOT by M. Gouy (http://biom3.univ-lyon1.fr/software/njplot.html).

Generation of a *ceh-6* gene disruption

A Tc1 insertion in *ceh-6* was obtained from the library of Dr Plasterk (Zwaal et al., 1993). The strain containing the Tc1 insertion in *ceh-6* was very unhealthy. Thus, we crossed *ceh-6::Tc1* with *dpy-5(e61) unc-29(e1072)*/++ males. *dpy-5(e61) unc-29(e1072)* should not only balance *ceh-6::Tc1*, but also the *mut-2* mutator locus that can be balanced with *unc-13* (M. Finney, PhD thesis, MIT, Cambridge, MA, 1997; J. Collins, personal communication). Animals carrying the Tc1 were detected by PCR. *ceh-6::Tc1*/*dpy-5(e61) unc-29(e1072)* F1 males were independently crossed to *dpy-5(e61) unc-29(e1072)* hermaphrodites; heterozygous *ceh-6::Tc1*/*dpy-5 unc-29* F2 males were again backcrossed to *dpy-5 unc-29* hermaphrodites. Seven independent lines were thus propagated and hermaphrodites of each line were singled to obtain *ceh-6::Tc1* animals, but only two lines yielded healthy hermaphrodites carrying *ceh-6::Tc1*. With the healthy strain, 100 small plates were inoculated using two animals each. After the plates starved, 500 μl of water was added and an aliquot of worms (100 μl containing about 1/5 to 1/3 of the total number of worms) was pipetted off from each plate into 0.5 ml tubes. Samples were then analyzed using an adapted single worm PCR protocol similar to that presented in Greenstein et al. (Greenstein et al., 1994). One deletion was recovered, and PCR analysis of single sibling offspring indicated that *ceh-6* might be lethal, as no homozygous animals could be obtained.

Backcrossing of *ceh-6 (mg60)*

ceh-6 (mg60)/+ was crossed with *dpy-5 unc-29*/++ males. *ceh-6/dpy-5 unc-29* animals were identified by PCR and offspring analysis. Dpy non-Unc recombinant animals were then collected from this line and assayed by PCR for *ceh-6(mg60)*. *dpy-5 ceh-6/dpy-5 unc-29* hermaphrodites were crossed with *lin-28(n719)*/+ males (*lin-28* hermaphrodites had been passed through dauer to allow crossing with wt males). *dpy-5 ceh-6/lin-28* hermaphrodites were identified and crossed with *lin-28(n719) unc-29(e1072)*/++ males. *dpy-5 ceh-6/lin-28 unc-29* offspring were selected, and from these Lin-28 non-Unc recombinant offspring were collected. Ten recombinants were assayed by PCR for *ceh-6(mg60)*. Three were of genotype *lin-28 ceh-6/lin-28 unc-29* and seven were of genotype *lin-28/lin-28 unc-29*. *lin-28 ceh-6/lin-28 unc-29* hermaphrodites were passed through dauer and mated with *dpy-5(e61) unc-29(e1072)* males. From the resulting *lin-28 ceh-6/dpy-5 unc-29* hermaphrodites, Dpy non-Unc recombinants were selected (approximately 60 independent recombinants). Each recombinant was crossed with *lin-28*/+ males to test for the absence of *lin-28* linked to *ceh-6*, whose presence was checked by PCR. Two *dpy-5 ceh-6/lin-28* lines were obtained. Single offspring of one of these lines was collected (about 100 lines), and examined for the absence of Dpy animals, which arise as recombinants. Several non-*dpy-5 ceh-6/lin-28* hermaphrodites were crossed with *dpy-5/unc-29* males and 20 single wt male progeny were then crossed to *dpy-5/unc-29* hermaphrodites. Only crosses giving Dpy Unc males were used further: *ceh-6/dpy-5 unc-29* males were backcrossed three more times to *dpy-5/unc-29* hermaphrodites to give strain TB1 (*ceh-6(mg60)/dpy-5(e61) unc-29(e1072)*).

Rescue of *ceh-6(mg60)*

Transgenic lines carrying arrays of pTRB462 and *rol-6* were crossed to TB1. About 100 rolling animals of the F_2 generation were singled onto plates. One quarter of these we expected to be homozygous for *ceh-6(mg60)*. Indeed, only 60% were either homozygous for wild-type or heterozygous. However, the picking was biased, as we could detect roller animals with variable abnormalities, which we specifically picked. Observed variable abnormalities were clear: small, and/or skinny appearance and defects in the tail area around the rectum, and some animals had an Egl phenotype. These animals produced few offspring, most of which died during embryogenesis. Seven lines carrying the control construct pTRB464 were examined, but did not show any rescue.

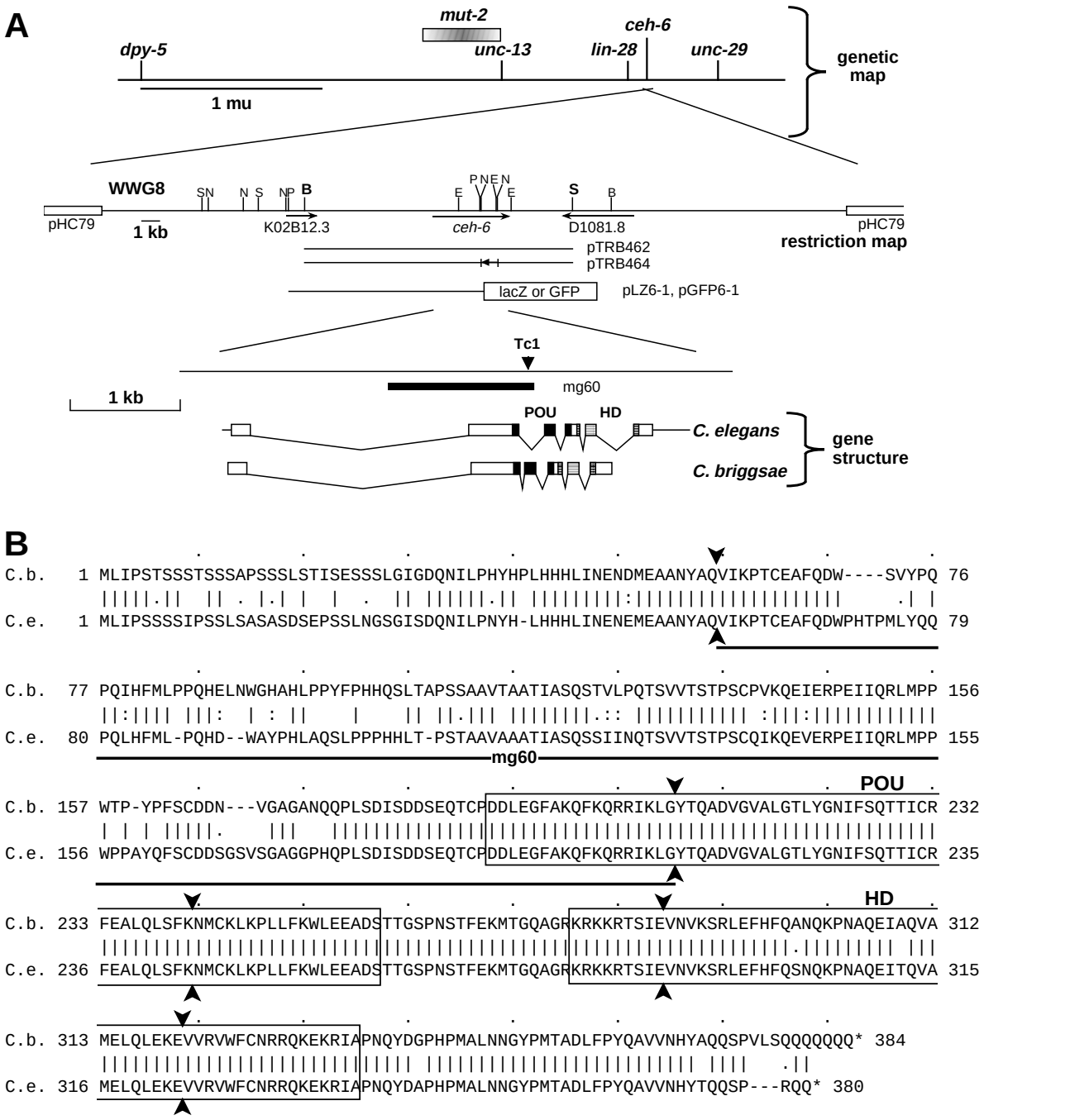

Fig. 1. Genetic map, physical map, structure and sequence of *ceh-6*. (A) The top part shows the genetic map of *ceh-6* with selected flanking genes. The approximate position of the mutator locus *mut-2* is indicated (J. Collins, personal communication). The center part shows the position of *ceh-6* and the Tc1 insertion within cosmid WWG8 and the constructs used. *ceh-6* is transcribed from right to left on the chromosome. At the bottom is a schematic representation of the gene structure of *C. elegans* and *C. briggsae ceh-6.* The POU-specific domain and homeodomain are indicated. (B) The deduced protein sequence for *C. elegans ceh-6* is compared with that of *C. briggsae*. Splice sites are marked with arrowheads. The *ceh-6(mg60)* gene disruption removes 1434 bases, including exon 2. The POU-specific domain and the homeodomain (HD) are boxed.

RESULTS

Structure of *ceh-6*

ceh-6 was identified by probing *C. elegans* genomic libraries with the highly degenerate oligonucleotides PRD-1 and PRD-2, which detect conserved amino acid positions in helix 3 of the homeodomain (Bürglin et al., 1989). *ceh-6* is located in the middle of cosmid WWG8 on chromosome I (Fig. 1A). A *ceh-6* cDNA was generated using PCR from total cDNA, and this cDNA was used to express recombinant CEH-6 protein. In addition, five distinct cDNAs were isolated from λgt11 libraries provided by P. Okkema. Based on these data, we find that *ceh-6* has 6 exons and 5 introns (Fig. 1A). A 300 base long 3′ untranslated region follows the coding region.

Blast searches revealed a *ceh-6* homolog on clone CB018A07 from the *C. briggsae* genome project (Genome Sequencing Center, personal communication). The *C. briggsae* gene has the same structure as the *C. elegans* gene. The major differences are in the length of intron 2 and intron 5, which are markedly smaller in *C. briggsae* (Fig. 1A). The large intron 1 is about the same size in *C. briggsae* and several small conserved sequence elements are present (data not shown), suggesting that regulatory elements are located in that intron. Sequence conservation is also found upstream of the predicated start methionine, as expected for regulatory regions, but no sequence conservation was found in the 3′ untranslated region (data not shown).

ceh-6 orthologs and phylogenetic analysis

The predicted amino acid sequence of CEH-6 is 380 amino acids, while that of the *C. briggsae* homolog is 384 residues (Fig. 1B). The *C. briggsae* gene is clearly the ortholog of *ceh-6*, because there is synteny of the ORFs upstream and downstream and the overall sequence identity between the two genes is 85%. Within the POU-specific domain they are 100% identical, and within the homeodomain they are 97% identical. This degree of similarity is in the appropriate range reported for other orthologs in *C. elegans* and *C. briggsae* (de Bono and Hodgkin, 1996; Robertson, 1998; Sluder et al., 1999). The major regions that differ are in a serine-rich stretch at the N terminus and in some patches in exon 2, which is upstream of the POU-specific domain.

The sequence identity in the conserved domains between *C. elegans* CEH-6 and *Drosophila* Cf1a/Dfr/Vvl, as well as between CEH-6 and mouse Brn1, is 93% (64/69) in the POU-specific domain and 75% (45/60) in the POU homeodomain. For comparison, the identity between fly Cf1a and mouse Brn1 is 97% (67/69) in the POU-specific domain, and 93% (56/60) in the homeodomain. However, outside the POU domain there

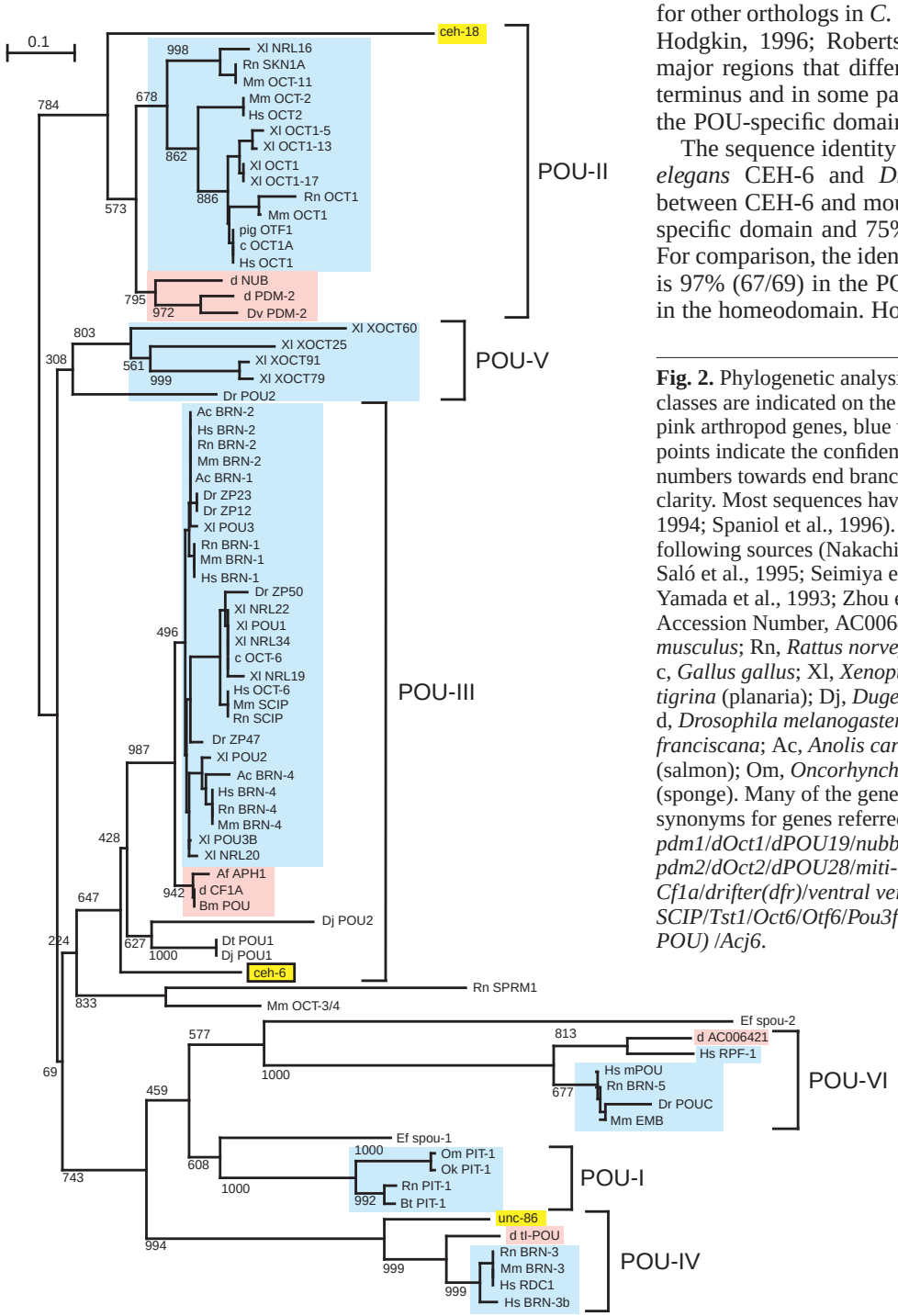

Fig. 2. Phylogenetic analysis of POU homeobox genes. The different classes are indicated on the right. Yellow indicates *C. elegans* genes, pink arthropod genes, blue vertebrates genes. Numbers at branch points indicate the confidence limits of 1000 bootstrap trials; numbers towards end branches (to the right) are not shown for clarity. Most sequences have been compiled previously (Bürglin, 1994; Spaniol et al., 1996). Additional sequences are from the following sources (Nakachi et al., 1997; Ono and Takayama, 1992; Saló et al., 1995; Seimiya et al., 1997; Sumiyama et al., 1996; Yamada et al., 1993; Zhou et al., 1996). Drosophila POU-VI gene Accession Number, AC006421; Hs, *Homo sapiens*; Mm, *Mus musculus*; Rn, *Rattus norvegicus*; pig, *Sus scrofa*; Bt, *Bos taurus*; c, *Gallus gallus*; Xl, *Xenopus laevis*; zf:, *Danio rerio*; Dt, *Dugesia tigrina* (planaria); Dj, *Dugesia japonica*; Bm, *Bombyx mori;* d, *Drosophila melanogaster*; Dv, *Drosophila virilis*; Af, *Artemia franciscana*; Ac, *Anolis carolinensis*; Ok, *Oncorhynchus keta* (salmon); Om, *Oncorhynchus mykiss*; Ef, *Ephydatia fluviatilis* (sponge). Many of the genes are known under multiple names, some synonyms for genes referred to in the figure are *Drosophila pdm1/dOct1/dPOU19/nubbing(nub)*; *Drosophila pdm2/dOct2/dPOU28/miti-mere(miti)*; *Drosophila Cf1a/drifter(dfr)/ventral veinless (vvl)*; vertebrate *SCIP/Tst1/Oct6/Otf6/Pou3f1*; vertebrate *I-POU/twin of I-POU(tI-POU) /Acj6*.

is no significant sequence similarity between CEH-6 and POU-III proteins from other phyla.

Given that the complete *C. elegans* and *Drosophila* genomic sequences are now available, we performed a phylogenetic analysis to reassess the worm POU genes in the context of vertebrate and fly genes (Fig. 2). We used neighbor-joining over the complete POU domains of a representative set of POU homeobox genes. The analysis reveals most of the families of genes that have been defined previously (Bürglin, 1994; Spaniol et al., 1996). However, several new conclusions can be drawn. The *C. elegans* protein CEH-18, previously thought to form a novel family, clusters with the POU-II genes, i.e. the transcription factors Oct1 and Oct2. The highly divergent rat gene Sprm1 clusters with mouse Oct3/4. Zebrafish Pou2, which has been assigned to a new family (Spaniol et al., 1996), appears to belong to family POU-V. Family IV is a distinct subclass comprising Unc-86, *Drosophila* I-POU (recently shown to correspond to Acj6, Clyne et al., 1999), and the vertebrate Brn3 genes. The family POU-VI, defined by Brn5/mPOU genes, is now proven to be highly conserved, since an ortholog has been found in the *Drosophila* genome project. The POU-I family is presently only defined by vertebrate Pit1 genes. Two sponge POU homeobox genes (spou-1 and spou-2) appear to cluster with POU-I and POU-VI, respectively. Overall, most higher metazoan POU homeobox genes can be assigned to one of the six subclasses.

The phylogenetic and sequence analysis shows that *ceh-6* is the ortholog of the vertebrate genes *Brn1, Brn2, SCIP/Tst1/Oct6/Pou3f1* and *Brn4*, and the *Drosophila* gene *Cf1a*/*dfr*/*vvl*, since *ceh-6* is the only POU-III gene in *C. elegans*.

Expression pattern of *ceh-6*

To examine the expression pattern of *ceh-6*, two independent approaches were taken using both anti-CEH-6 antibodies and reporter constructs. For antibodies, CEH-6 protein expressed in *E. coli* was used to raise rabbit antisera, which were affinity purified. For the gene fusions, a 11.5 kb *Nco*I fragment that extends from the *Nco*I site in the POU-specific domain towards the 5′ region (Fig. 1A) was cloned into *lacZ* (clone LZ6-1) or green fluorescent protein (GFP) (clone GFP6-1) expression vectors (see Materials and Methods).

Whole-mount immunolocalization of CEH-6 and analysis of *ceh-6* gene fusions resulted in the following expression pattern. At the L1 larval stage CEH-6/*ceh-6* is expressed in four pairs of bilaterally symmetric neurons in the lateral ring ganglion of the animal. These neurons are the RMDDL/R, RMDVL/R, AUAL/R and AVHL/R neurons (Fig. 3A-G). The expression in RMDD and RMDV is weaker than in AUA and AVH. *ceh-6* is also expressed in the excretory cell, very strongly with the *lacZ* reporter construct but more weakly with the antibody, which is probably due to the large volume of the nucleus (Fig. 3A,F). Despite the nuclear localization signal in the *lacZ* construct, the β-galactosidase was occasionally expressed at such high levels that the excretory canals were also stained (Fig. 3F,G). Posterior to the excretory cell, the neurons SABVL/R in the retro-vesicular ganglion express CEH-6/*ceh-6* (Fig. 3A,F). Additional CEH-6-expressing cells in the body and tail were observed only with the antiserum, indicating that the *ceh-6* reporters may not contain all promoter elements. In the body region, expression was observed in dividing P.na cells in the ventral nerve cord in L1 animals (Fig. 3J,L). The Pn.a expression is transient, appearing before the cell division of Pn.a and fading in the daughter cells. During the cell division, CEH-6 is localized to the cytoplasm of the dividing cells (Fig. 3L). It appears that the posterior daughters lose CEH-6 before that of the anterior daughter, as seen in the anterior Pn.ap cells (Fig. 3J,L). In the tail, CEH-6 expression in L1 animals is seen around the rectum in the five rectal cells B, Y, U, F and K (Fig. 3H,I). After K has divided, only the rectal cell K.a expresses CEH-6, though expression during the division was not monitored. At the L2 stage, a sixth cell becomes apparent that, based on its position at the very bottom of the rectum, is deduced to be P12.pa. Head and rectal expression of CEH-6 persists into adulthood. In addition, in adult animals four symmetric CEH-6 expressing cells are seen around the vulva, possibly one set of the vul cells, though we did not determine their identity (data not shown).

During embryogenesis *ceh-6* is expressed at the comma stage in two clusters: one cluster corresponds to the ectodermal cells surrounding the anus (B, Y, U, F and K), and the other cluster corresponded to the cells in the head described above, many of which are located relatively close to each other at that stage in development (Sulston et al., 1983; Fig. 3M,N).

Phenotypic defects of a *ceh-6* gene disruption

To study the function of *ceh-6*, we disrupted the gene using Tc1 transposon-mediated deletion (Zwaal et al., 1993). We identified a 1.4 kb deletion in *ceh-6* (Fig. 1B, see Materials and Methods). This *ceh-6* allele (*mg60*) lacks exon 2, which encodes the N-terminal region of the highly conserved POU domain as well as other coding regions conserved between *C. elegans* and *C. briggsae*. The other exons, including those encoding the homeodomain, are still present. The *ceh-6 (mg60)* transcript is predicted to splice from exon 1 to exon 3, which would cause a frameshift, resulting in a lack of both the POU-specific domain and the homeodomain. However, splicing from exon 1 to exon 4 cannot be excluded, and would result in a protein with a functional homeodomain that lacked a functional POU-specific domain. The POU-specific domain regulates DNA-binding specificity of POU proteins. Thus, *ceh-6(mg60)* is predicted to be a strong and perhaps a null allele. *ceh-6(mg60)* was backcrossed ten times, which included recombination with *lin-28* to replace the chromosome arm to the left of *ceh-6*. This recombination also removes the *mut-2* mutator locus that is located to the left of *ceh-6* (J. Collins, personal communication).

The following phenotypes were observed in the backcrossed mutant: More than 80% of *ceh-6(mg60)* animals die during embryogenesis. These embryos died with variably disorganized and disrupted posterior regions. More detailed observation revealed that abnormal cell protrusions started to appear around the twofold stage of embryogenesis. For example, in one embryo a small protrusion of a few cells could be seen in the presumptive rectal area (Fig. 4A). After 30 minutes, the whole rectal region of the same embryo was a chaotic mass of loose cells (Fig. 4B,C). We investigated this further using a 4D microscopy system to record 3D stacks through time (Bürglin, 2000). In many embryos, cells start to extrude in the rectal area. The time point of the event varies: in one case, the tail area of the embryo started to swell at the late twofold stage until, after about 30 minutes, the first cells started to extrude in the rectal area (Fig. 4D). Another 30 minutes later, a large mass of cells had extruded, resulting in

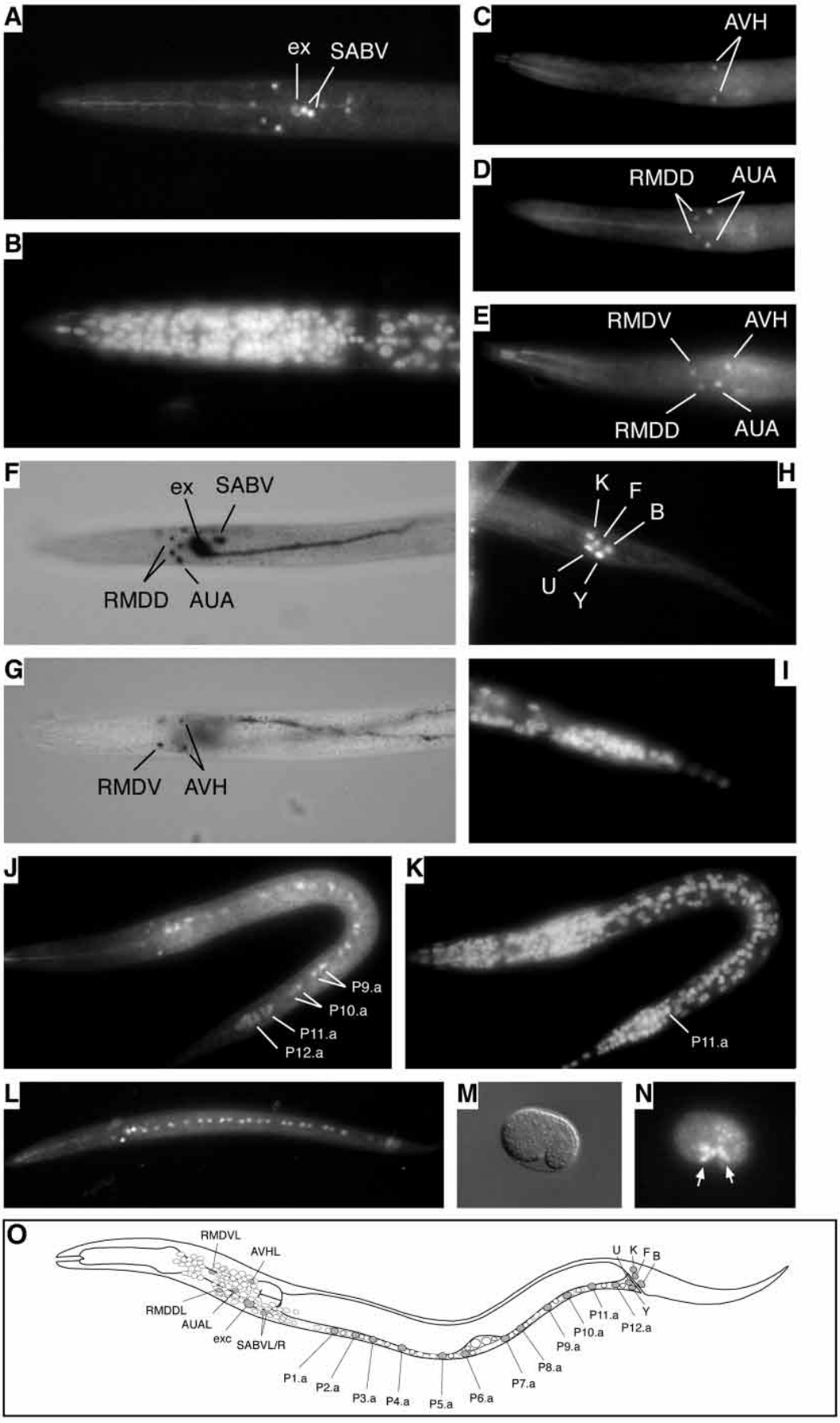
A
ex
SABV
B
C
AVH
D
RMDD
AUA
E
RMDV
AVH
RMDD
AUA
F
ex
SABV
RMDD
AUA
H
K
F
B
U
Y
G
RMDV
AVH
I
J
P9.a
P10.a
P11.a
P12.a
K
P11.a
L
M
N
O
RMDVL
AVHL
RMDDL
AUAL
exc
SABVL/R
U
K
F
B
Y
P11.a
P12.a
P10.a
P9.a
P8.a
P7.a
P6.a
P5.a
P4.a
P3.a
P2.a
P1.a

Fig. 3. Expression pattern analysis of *ceh-6* using anti-CEH-6 antibodies (A-E,H-L). Affinity-purified anti-CEH-6 antibodies were incubated with whole-mount fixed mixed-stage *C. elegans* and visualized using anti-rabbit FITC-coupled secondary antibodies. Expression was also analyzed using reporter constructs: transgenic animals with the *lacZ* reporter construct LZ6-1 were visualized by staining with X-gal (F ,G) using differential interference contrast (DIC), the GFP reporter construct GFP6-1 was visualized using fluorescent microscopy (M,N). (A) Ventral view of the head of an L1 animal: expression of CEH-6 in the ring ganglia of the head. Specifically marked are the excretory cell (ex) and the SABV neurons. The 'dots' to the right of the SABV neurons are background from the pharyngeal lumen. (B) DAPI of A. (C) Dorsal focal plane; CEH-6 expression in AVH. (D) Same animal as in C; subventral focal plane; CEH-6 expression in RMDV and AUA. (E) Lateral view of CEH-6 expression. (F,G) *ceh-6::lacZ* expression in head region, two different focal planes are shown. Excretory canals are stained despite a nuclear localization signal; twisted appearance is due to the *rol-6* marker used to generate transgenic animals. The dorsal focal plane shows the neurons RMDDL/R and AVHL/R, the excretory canals are also visible as twisting lines. The ventral focal plane shows RMDVL/R, AVHL/R, the excretory cell and canals, and SABVL/R. (H) CEH-6 expression in the rectal cells. (I) DAPI of same animal as in H. (J) CEH-6 expression in the ventral nerve cord. Transient expression is seen in the dividing Pn.a cells. Note that P11.a and P12.a are just undergoing cell division, CEH-6 is localized in the cytoplasm in these cells. (K) DAPI of same animals as in J. (L) CEH-6 expression in the ventral nerve cord at a stage slight prior to that in J; the posterior Pn.a cells have not yet divided. (M,N) *ceh-6::gfp* expression in the comma stage embryo, DIC (M), and fluorescent channel (N). Note that most cells that express *ceh-6* in the head are localized where close to each other at that stage (arrows point to the clusters in the head region and the tail region). (O) Schematic summary overview of the *ceh-6* expressing cells in a late L1 larva.

a totally disrupted posterior region (Fig. 4D). In some cases, the embryos hatch, but die soon afterwards with grossly malformed rectal areas (Fig. 4E,F). A number of animals that are able to hatch have a less severe malformed rectal structures (Fig. 4G).

During elongation at late stages of embryogenesis a strong pressure is exerted on the cells in the interior of the embryo when the hypodermal cells shape the worm. Laser ablation of hypodermal cells and their precursors, as well as rectal cells has shown that microsurgically treated embryos die during elongation, because cells ooze out through the missing area or embryos rupture at the point where the cells have been destroyed (Sulston et al., 1983; Priess and Hirsh, 1986). A similar phenomenon seems to occur in *ceh-6(mg60)* when the embryos rupture in the rectal region. In those instances, where the embryos and early larvae do not extrude cells but only show big enlargements of the posterior, the defects may be due to a lack of appropriate internal attachments of the rectal cells to their neighbors.

Those animals that are able to hatch have a variable penetrance of malformed rectal structures (Fig. 4G). Additionally, however, they start to develop vacuoles throughout the body area and move very little (Fig. 4H-J). We could observe the formation of vacuoles within 30 minutes of hatching. Most animals die heavily vacuolated during L1/L2, although some can persist as scrawny, clear animals for 1-2 days with variously detached tissues (Fig. 4I,J).

The most parsimonious explanation for the larval lethality is that *ceh-6* is essential for the proper function of the excretory cell. Laser ablation of the excretory cell results in similar phenotypes, with accumulation of vacuoles and fluid, and lethality after about 5 days (Nelson and Riddle, 1984). Improper structure or function of the excretory cell might also prevent removal of toxic substances and would contribute to the lack of movement and general sickly appearance (M. Buechner, personal communication).

To demonstrate that the observed phenotypes are indeed due to *ceh-6(mg60)*, we performed several rescue experiments. Cosmid WWG8, which bears the complete *ceh-6* gene, rescues the *ceh-6(mg60)* lethal phenotype. Fully viable lines could be maintained, although they always showed sick and clear animals, probably owing to mosaic or partial rescue. Clone pTRB462 (Fig. 1A) was used to generate transgenic lines that were crossed into *ceh-6(mg60)*. These crossed arrays showed weaker rescuing activity. Many embryos showed rectal defects and many animals were sick and clear and had very low brood sizes. This may also be due to mosaic effects of the array. Control construct pTRB464 did not show any rescue.

Expression of the HOX cluster gene *egl-5* in *ceh-6(mg60)*

The HOX cluster gene *egl-5*, a divergent homolog of the *Drosophila* gene *Abdominal-B*, acts and is expressed in the rectal cells, as well as in other cells in the tail of *C. elegans* (Chisholm, 1991; Wang et al., 1993). It is unlikely that EGL-5 regulates *ceh-6*, since *egl-5* mutants have no lethal phenotype and the defects are mostly restricted to abnormal cell lineages during cell divisions in larval stages, in particular in males (Chisholm, 1991).

We used *egl-5* as a marker for the rectal cells and tested whether CEH-6 regulates *egl-5*. An *egl-5::gfp* reporter construct was crossed into *ceh-6(mg60)/dpy-5(e61) unc-29(e1072)*. Analysis of *ceh-6(mg60)* homozygous animals revealed that *egl-5::gfp* is still expressed in the rectal area in several cells. In embryos, the GFP-expressing cells were distributed over large regions in the disrupted rectal area (Fig. 5D-F), supporting the notion of a break-up in this area with a concomitant uncontrolled dispersal of the cells. The normally adjacent rectal cells (Wang et al., 1993, Fig. 5A,B) are widely separated, suggesting that they do not properly adhere to each other. Thus, *ceh-6* may directly or indirectly control cell surface or cell adhesion molecules. These results also show that the cells are apparently still present and that *ceh-6* is not required for the formation of the rectal cells, but is necessary for their proper differentiation.

From the *egl-5::gfp* analysis and the 4D data we conclude that *ceh-6* affects the generation or the differentiation of the rectal cells B, F, Y, U and K, where *ceh-6* is expressed. As a consequence, hypodermal cells and internal cells probably do not properly attach to the rectal cells. This leads to cell adhesion defects in the rectal area, resulting in either ruptured embryos, or gross abnormalities in the posterior of embryos and hatched L1 larvae.

Excretory canal defects

To examine the morphology of the excretory cell, a *vha-1::gfp* reporter construct was crossed into *ceh-6(mg60)*. *vha-1* encodes a proteolipid of the vacuolar-type H^+-ATPase (Oka et al., 1997) and is expressed in the excretory cell (Fig. 5G,H).

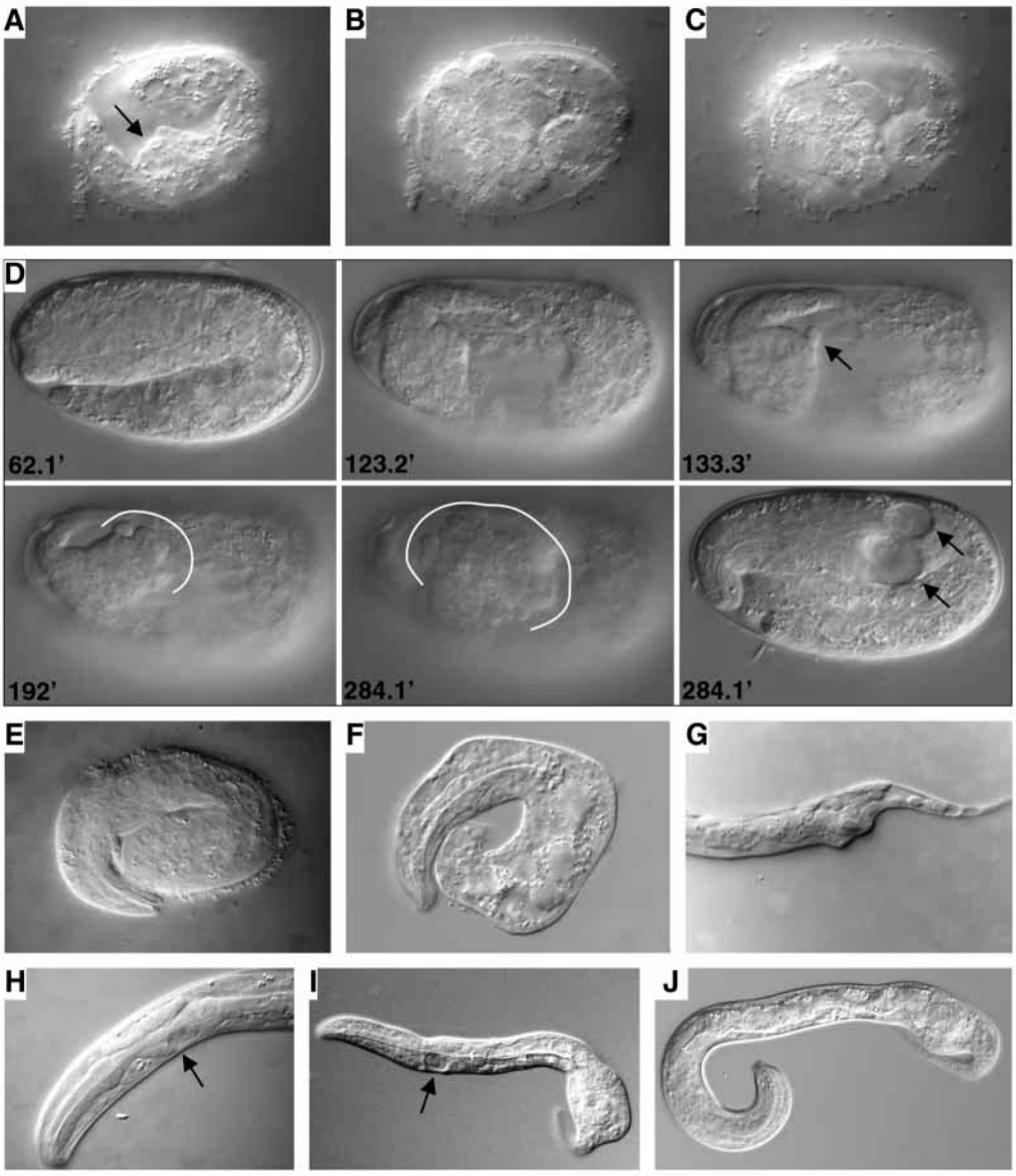

Fig. 4. Phenotypic analysis of the *ceh-6(mg60)* gene disruption. DIC photographs. (A) Twofold embryo with an abnormal protrusion of cells in the rectal area (arrow). (B,C) Same embryo about 30 minutes later; the protrusion has grown to a large unorganized mass of cells. Two different focal planes are shown in B and C. (D) Sample images from a 4D recording through late embryogenesis. The numbers are minutes that refer to the start of the recording session at the early twofold stage. Initially (62.1 minutes) the embryo is normal. From around 110 to 130 minutes (sample at 123.2′) the rectal area enlarges and bulges. At 133.3 minutes, the first cell (arrow) oozes out in the region of the anus. Subsequently, more and more cells extrude (192′, 284.1′, marked with a white line). From about 130 to 150 minutes the embryo stops most movements and does not elongate anymore, probably because of the loss of internal pressure. Also, large vacuoles develop in the head region (arrows) starting at around 225 minutes (284.1′). This could be related to some excretory cell defect, or perhaps be the consequence of loss of cell mass from the interior. (E) Hatching embryo, with abnormal body and tail region. (F) Embryo, arrested at hatching, with abnormal tail and body region, and detached gut. (G) L1 larval animal with misformed rectal area. (H,I) L1 larval animals with vacuoles in the region of the excretory cell (arrows), and separation of tissues, such as nerve cells above pharynx. (J) L1 larval animal with detached cells throughout the body and misformed rectal area.

Homozygous *ceh-6(mg60)* larvae were recognized by their aberrant rectal structures (Fig. 5I,L). In many GFP-positive *ceh-6* mutant animals no excretory cell could be discerned in differential interference contrast (DIC), and no excretory cell with its canals could be seen using the *vha-1::gfp* reporter gene, although GFP is present in many vesicular type structures throughout the animal (data not shown). In some cases, the excretory cell could be clearly identified in DIC, and abnormally short and enlarged excretory canals were revealed with *vha-1::gfp* (Fig. 5J-M). These data show that *ceh-6* gene activity is necessary for the differentiation of the excretory cell.

DISCUSSION

Evolution of POU genes

C. elegans has only three POU genes, *unc-86*, *ceh-6* and *ceh-18*. *ceh-18* is the ortholog of the fly *pdm* and vertebrate *Oct1/Oct2/Skn1* genes. *ceh-6* and *Cf1a/dfr/vvl* can clearly be assigned to POU-III, and *unc-86* and *I-POU/Acj6* to POU-IV. The POU-VI family is also highly conserved in evolution from flies to vertebrates. Strikingly, no homolog for this gene seems to exist in *C. elegans*.

Two other families, POU-I and POU-V, are defined only by vertebrate genes so far. POU-V is a ‘loose’ group of divergent genes. For example, zebrafish *pou2* has been classified as a separate group (Spaniol et al., 1996). Two other divergent genes may belong to POU-V, *Sprm1* and *Oct3/4*, although in this analysis they do not cluster with the other POU-V genes. *Oct3/4* may be derived from the POU-III family, since *Oct3/4* and *ceh-6* share the same intron position in the homeodomain. The POU-I family has only a single member, *Pit1*, which plays a role in neuroendocrine development (Bodner et al., 1988; Ingraham et al., 1988). Given this specialized function

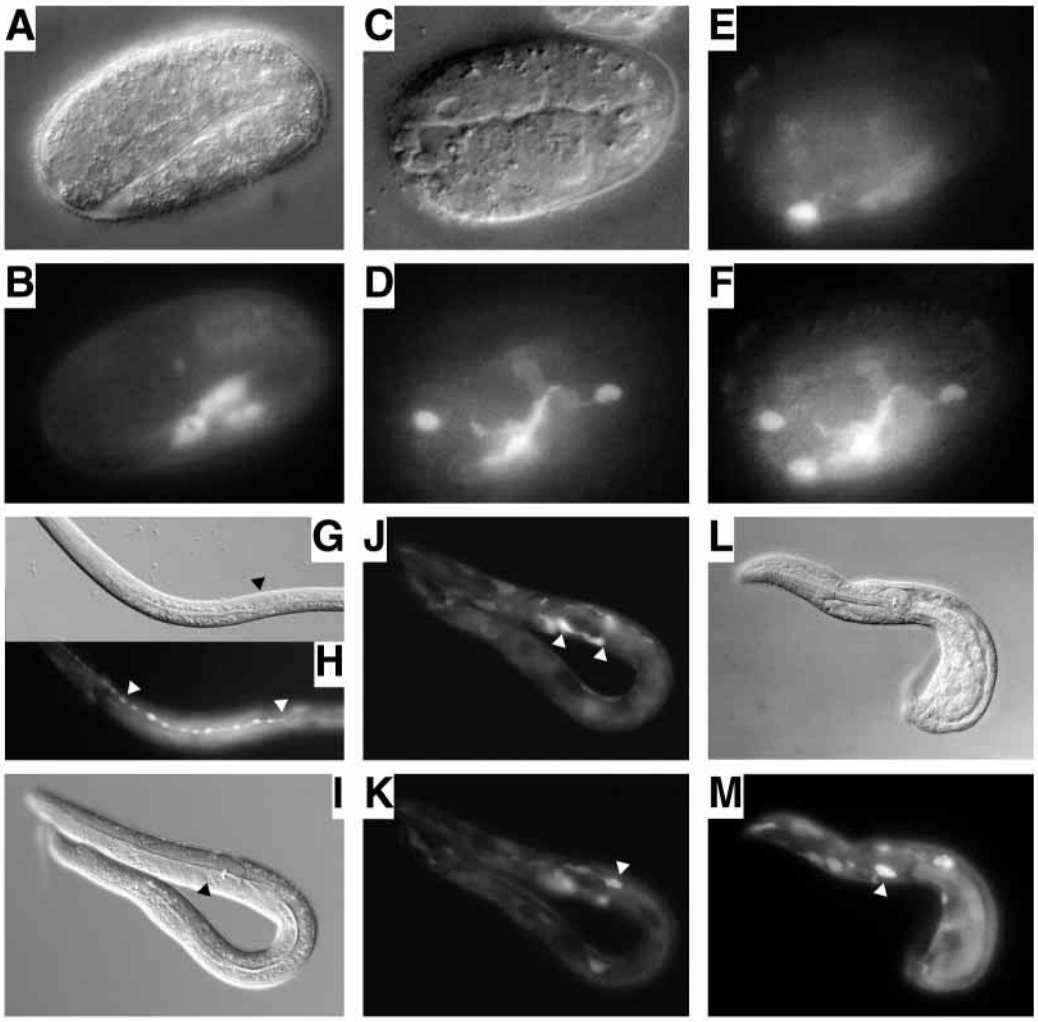

Fig. 5. Expression of *egl-5::gfp* (A-E) and *vha-1::gfp* (F-K) reporter constructs in wild type (A,B,G,H) or *ceh-6(mg60)* (C-F,I-M) embryos and L1 larvae. GFP was visualized using fluorescent microscopy. (A) Wild-type twofold embryo. (B) Fluorescent image of A, showing the expression in the rectal region. (C) DIC, twofold embryo with disrupted tail area. (D,E) GFP-expressing cells in two focal plains of the embryo shown in C. (F) Composite of C-E. Note that the GFP-expressing cells are distributed over wide area within the tail region. (G) Wild-type L1 larva; an arrowhead marks the gonad precursor. (H) GFP expression in the left excretory canal, which extends to the middle of the animal, can be seen (marked with arrowheads). (I) DIC of *ceh-6* larva with malformed rectum; the arrowhead indicates the excretory cell. (J,K) GFP expression of I, two focal plains. The arrowheads show the excretory cell body and the premature end of the left canal (J), and the premature end of the right canal (K). (L) DIC of *ceh-6* larva with malformed posterior body region, no excretory cell can be discerned. (M) Fluorescent image of L, showing GFP in the region where the excretory cell should be (arrowhead). No canals can be seen.

and the single copy status, it is possible that *Pit1* is a relatively recent divergent POU gene in vertebrate evolution. However, there are two POU genes from sponges, *spou-1* and *spou-2* (Seimiya et al., 1997). *spou-1* clusters with the *Pit1* genes, suggesting that this family could be ancient. *spou-2* clusters with the POU-VI family. If these two POU families were truly that ancient, it would indicate that *C. elegans* has lost these genes. In absence of further data from sponges, one should remain cautious, as sponges may have a smaller complement of POU genes, and *spou-1* and *spou-2* could represent several POU families.

Members of the POU-III family have been isolated from many species, in particular from mammals (*Brn1*, *Brn2*, *SCIP/Oct6/Tst1/Pou3f1*, *Brn4/RHS4*), *Xenopus* (*XlPOU2*, *Brn4*), zebrafish, insects, planaria. In mammals, the genomic structure has been determined, and it was found that no introns are present in the four POU-III genes (Hara et al., 1992). In contrast, five introns interrupt the open reading frame of *ceh-6*, two in the POU-specific domain and two in the homeodomain. One intron position in the homeodomain is in exactly the same place as in the mammalian Oct3/4 genes (Bürglin, 1995), thus this intron position may be ancient. The proposal by Hara et al. that the vertebrate POU-III genes arose by reverse transcription (Hara et al., 1992) seems unlikely, as the POU-III family is very ancient. The four POU-III subfamilies in vertebrates are most likely a result of the same polyploidisation events during vertebrate evolution that also gave rise to the four Hox clusters (Sharman and Holland, 1996).

ceh-6 displays a complex expression pattern

The expression pattern of *ceh-6* is complex. It is expressed in a series of neurons in the ring ganglia, in the excretory cell, in dividing neuroblasts in the ventral cord and in the rectal cells (Fig. 3O). Furthermore, additional cells express *ceh-6* in adult hermaphrodite and male animals (not shown). The *ceh-6*-expressing cells are neither related by cell lineage nor by function. The different cell types represented are motoneurons/interneurons (RMDD, RMDV), putative sensory neurons (AUA), interneurons (SABV), neuroblasts (Pn.a), epidermal cells (B, F, K, U, Y) and an osmoregulatory cell. In several instances, CEH-6 expression differentiates subtypes of neurons. There are three pairs of neurons classified as RMD motoneurons: RMD, RMDD, RMDV. The RMDD and RMDV neurons are distinguished by CEH-6. Similarly, CEH-6 is expressed in the SABV neurons, but not in SABD neurons. The

transient expression in the dividing Pn.a cells of the ventral nerve cord is intriguing. Detailed studies of other homeobox genes, such as *mab-5* and *egl-5* have shown that precise control of expression in particular cell lineages is an important aspect of their function (Ferreira et al., 1999; Salser and Kenyon, 1996).

How is the complex expression pattern achieved? The *ceh-6::lacZ* reporter construct is only expressed in the head region and does not recapitulate the complete *ceh-6* expression pattern. This suggests that the promoter of *ceh-6* is large and complex. Such a complex promoter has also been observed in the case of *unc-86* gene, the *C. elegans* POU-IV homeobox gene (Baumeister et al., 1996). A further suggestion how the expression pattern is set up comes from the expression during embryogenesis. At the late comma stage, the *ceh-6* expressing cells in the tail (the rectal cells) and the cells in the head form a relatively compact cluster in the head and tail, respectively (Fig. 3N). It is, thus, feasible that these two groups of cells might respond to external local signals or to localized determinants in these areas of the embryo. However, a detailed study of the dynamics of *ceh-6* expression during embryogenesis is needed to resolve this issue.

While in other organisms overlapping gene expression between different POU families has been observed, we do not see any overlapping expression between *unc-86*, *ceh-6* and *ceh-18*. It suggests that the most basic function of the POU homeobox genes is to specify distinct neuronal and epidermal cell types during development.

Evolutionary considerations about POU-III genes in the nervous system

ceh-6 is expressed in many different neuronal cell types. This is comparable with POU-III genes in *Drosophila* and vertebrates. In these organisms, the genes *Brn1*, *Brn2*, *SCIP* and *Brn4* are also expressed in many different tissue types in the brain (see Alvarez-Bolando et al., 1995, Hauptmann and Gerster, 2000). Owing to the complexity of the vertebrate brain, it is difficult to make direct comparisons to the precisely defined neurons in *C. elegans*, but it seems clear that the vertebrate orthologs are expressed in some of the same different types of neurons, i.e. motoneurons, sensory neurons and interneurons. Furthermore, in vertebrates, expression is observed in the spinal cord, a structure for which there is more and more evidence to indicate that it corresponds to the ventral cord in flies and nematodes (Arendt and Nübler-Jung, 1997; Arendt and Nübler-Jung, 1999; De Robertis, 1997). Both, *ceh-6* and *dfr/vvl* are expressed in the ventral cord, while several of the vertebrate POU-III genes are expressed in the developing neural tube and the spinal cord (Ryan and Rosenfeld, 1997). *ceh-6* is transiently expressed in the ventral cord, and, for example, the vertebrate factor SCIP has also been shown to be transiently expressed in Schwann cells of the PNS (Monuki et al., 1990). Although we did not examine the phenotypes of *ceh-6(mg60)* in the nervous system, due to the early defects, we nevertheless suspect that *ceh-6* plays an important role in that system, like its cousin, *unc-86*.

Involvement of POU-III genes in the epidermal cells

The studies in *Drosophila* of *dfr/vvl* have mainly focused on the role in the developing tracheal system, since, similar to *C. elegans*, the mutants did not live long enough to examine the nervous system defects. *dfr/vvl* plays an important role during the development of the trachea, in particular for the differentiation and migration of these cells, and thus it might directly or indirectly regulate cell surface molecules (Anderson et al., 1995; de Celis et al., 1995). *dfr/vvl* is also expressed in the Filzkörper, the area where the tracheal system attaches to the cuticle in the rectal area of the fly. Our data suggest that *ceh-6* is required for the proper differentiation of the rectal cells, and that it may directly or indirectly regulate cell surface molecules. Further, *ceh-6* seems to regulate the outgrowth of the excretory canals. Thus, *ceh-6* may regulate the same types of molecules in epithelial cells that are regulated by *dfr/vvl*. In vertebrates, POU-III genes have also been implicated in regulating cell surface molecules. For example, *Brn2* directly regulates the expression of the cell surface adhesion molecule Po (He et al., 1991). In addition, *Oct6* has been shown to be expressed in the epidermis, as well as other squamous epithelia (Faus et al., 1994, Andersen et al., 1997). Thus, expression in subsets of epithelial cells appears to be another conserved feature of POU-III homeobox genes.

The excretory system, an evolutionary conserved system?

ceh-6 is expressed in the excretory cell and the morphology of this cell is abnormal in *ceh-6* mutants. Thus, *ceh-6* is essential for the proper function of this cell. In vertebrates, the POU-III gene *Brn2* plays a key role for the development and survival of the endocrine hypothalamus and posterior pituitary gland, as demonstrated by knockout mutations (Nakai et al., 1995; Schonemann et al., 1995). It is not required for the initial development, but rather for the differentiation of many neurosecretory neurons in this system. One of the molecules secreted by the endocrine system is vasopressin, a molecule that regulates the permeability of the kidney tubules for water. Another POU-III gene, *Brn1*, is expressed in kidney tubules (He et al., 1989). The kidney performs the function of osmoregulation in vertebrates; thus, despite the more complex organization of the osmoregulatory system in vertebrates, it is under the control of the vertebrate *ceh-6* orthologs. In the brine shrimp *Artemia franciscana* the single POU-III gene *APH-1* is expressed in the salt gland, which is necessary for osmoregulation (Chavez, et al., 1999). Another gland system, where a POU-III gene is expressed, is the silk gland in *Bombyx mori* (Fukuta et al., 1993), although this system is probably not related to osmoregulation.

Nothing is known about the evolution of secretory/osmoregulatory systems as no fossil records exist. We suggest that one of the roles of the POU-III class genes is in secretory/osmoregulatory systems and that this function is conserved in evolution. Additional support for the conserved nature of such a system stems from the observation that a LIM homeobox plays a role in the excretory system in *C. elegans* and humans. The *C. elegans* LIM homeobox gene *lim-6* is expressed in the excretory glands, which are associated with the excretory cell (Hobert et al., 1999), while mutations in the human ortholog *LMX1B* cause renal displasia (Dreyer et al., 1998). Other systems conserved in evolution have been described before. For example, the many eye types found in the different animal phyla have an underlying common molecular mechanism in the form of the *Pax6* gene (Callaerts et al., 1997; Quiring et al., 1994). Similarly, it seems logical to

assume that an osmoregulatory/excretory system has developed only once during metazoan evolution.

We thank Gudrun Aspöck, Joe Gatto and Gisela Niklaus for excellent technical assistance; and Ann Sluder, David Greenstein, Mike Finney, Matthew Buechner, John Collins, Guy Rousseau, Joseph Martial and Tom Barnes for stimulating discussions and comments. We are indebted to Michael Finney for a gift of first strand cDNA and helpful discussions, to Toshihiko Oka for clone pCV01, to Andrew Chisholm for the *egl-5::gfp* strain, to Pete Okkema for the λgt11 library, to Cori Bargmann for confirming cell identifications, and to Victor Ambros for advice and *lin-28* strains. We also thank the Genome Sequencing Center, Washington University, St Louis for communication of DNA sequence data prior to publication. T. R. B. was a recipient of a postdoctoral Swiss National Science Foundation Fellowship, and is currently supported by a START Fellowship (NF. 3130-038786.93) and grants from the Swiss National Science Foundation. This work was supported by Hoechst and the Human Frontiers Science Program. We gratefully thank Ron Plasterk for providing *ceh-6::Tc1*, which was isolated from a Tc1 mutant library supported by grant 5 RO1 RR10082-02 of the NIH-NCRR to R. Plasterk.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.

Alvarez-Bolado, G., Rosenfeld, M. G. and Swanson, L. W. (1995). Model of forebrain regionalization based on spatiotemporal patterns of POU-III homeobox gene expression, birthdates, and morphological features. *J. Comp. Neurol.* **355**, 237-295.

Anderson, M. G., Perkins, G. L., Chittick, P., Shrigley, R. J. and Johnson, W. A. (1995). *drifter*, a *Drosophila* POU-domain transcription factor, is required for correct differentiation and migration of tracheal cells and midline glia. *Gen. Dev.* **9**, 123-137.

Andersen, B., Weinberg, W. C., Rennekampff, O., McEvilly, R. J., Bermingham Jr., J. R., Hooshmand, F., Vasilyev, V., Hansbrough, J. F., Pittelkow, M. R., Yuspa, S. H. et al. (1997). Functions of the POU domain genes Skn-1a/i and Tst-1/Oct-6/SCIP in epidermal differentiation. *Genes Dev.* **11**, 1873-1884.

Arendt, D. and Nübler-Jung, K. (1997). Dorsal or ventral: similarities in fate maps and gastrulation patterns in annelids, arthropods and chordates. *Mech. Dev.* **61**, 7-21.

Arendt, D. and Nübler-Jung, K. (1999). Comparison of early nerve cord development in insects and vertebrates. *Development* **126**, 2309-2325.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1987). *Current Protocols in Molecular Biology.* New York: Green Publishing Associates and Wiley-Interscience.

Baumeister, R., Liu, Y. and Ruvkun, G. (1996). Lineage-specific regulators couple cell lineage asymmetry to the transcription of the *Caenorhabditis elegans* POU gene *unc-86* during neurogenesis. *Genes Dev.* **10**, 1395-1410.

Bodner, M., Castrillo, J.-L., Theill, L. E., Deerinck, T., Ellisman, M. and Karin, M. (1988). The pituitary-specific transcription factor GHF-1 is a homeobox-containing protein. *Cell* **55**, 505-518.

Bürglin, T. R. (1994). A comprehensive classification of homeobox genes. In *Guidebook to the Homeobox Genes*, (ed. D. Duboule), pp. 25-71. Oxford: Oxford University Press.

Bürglin, T. R. (1995). The evolution of homeobox genes. In *Biodiversity and Evolution* (ed. R. Arai, M. Kato and Y. Doi), pp. 291-336. Tokyo: The National Science Museum Foundation.

Bürglin, T. R. (2000). A two-channel four-dimensional image recording and viewing system with automatic drift correction. *J. Microsc.* **200**, 75-80.

Bürglin, T. R. and De Robertis, E. M. (1987). The nuclear migration signal of *Xenopus laevis* nucleoplasmin. *EMBO J.* **6**, 2617-2625.

Bürglin, T. R., Mattaj, I. W., Newmeyer, D. D., Zeller, R. and De Robertis, E. M. (1987). Cloning of nucleoplasmin from Xenopus laevis oocytes and analysis of its developmental expression. *Genes Dev.* **1**, 97-107.

Bürglin, T. R., Finney, M., Coulson, A. and Ruvkun, G. (1989). *Caenorhabditis elegans* has scores of homoeobox-containing genes. *Nature* **341**, 239-243.

Callaerts, P., Halder, G. and Gehring, W. J. (1997). *Pax-6* in development and evolution. *Annu. Rev. Neurosci.* **20**, 483-532.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. and Prasher, D. C. (1994). Green Fluorescent Protein as a marker for gene expression. *Science* **263**, 802-805.

Chavez, M., Landry, C., Loret, S., Muller, M., Figueroa, J., Peers, B., Rentier-Delrue, F., Rousseau, G. G., Krauskopf, M. and Martial, J. A. (1999). APH-1, a POU homeobox gene expressed in the salt gland of the crustacean *Artemia franciscana*. *Mech. Dev.* **87**, 207-212.

Chisholm, A. (1991). Control of cell fate in the tail region of *C. elegans* by the gene *egl-5*. *Development* **111**, 921-932.

Clyne, P. J., Certel, S., de Bruyne, M., Zaslavsky, L., Johnson, W. and Carlson, J. R. (1999). The odor specificities of a subset of olfactory receptor neurons are governed by Acj6, a POU-domain transcription factor. *Neuron* **22**, 339-347.

de Bono, M. and Hodgkin, J. (1996). Evolution of sex determination in *Caenorhabditis*: Unusually high divergence of *tra-1* and its functional consequences. *Genetics* **144**, 587-595.

de Celis, J. F., Llimargas, M. and Casanova, J. (1995). *ventral veinless*, the gene encoding the Cf1a transcription factor, links positional information and cell differentiation during embryonic and imaginal development in *Drosophila melanogaster*. *Development* **121**, 3405-3416.

De Robertis, E. M. (1997). The ancestry of segmentation. *Nature* **387**, 25-26.

Devereux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* **12**, 387-395.

Dreyer, S. D., Zhou, G., Baldini, A., Winterpacht, A., Zabel, B., Cole, W., Johnson, R. L. and Lee, B. (1998). Mutations in LMX1B cause abnormal skeletal patterning and renal dysplasia in nail patella syndrome. *Nature Genet.* **19**, 47-50.

Durbin, R. and Thierry Mieg, J. (1991-). A *C. elegans* database. Code and data available from anonymous FTP servers lirmm.lirmm.fr, ftp.sanger.ac.uk and ncbi.nlm.nih.gov.

Faus, I., Hsu, H.-J. and Fuchs, E. (1994). Oct-6: a regulator of keratinocyte gene expression in stratified squamous epithelia. *Mol. Cell. Biol.* **14**, 3263-3275.

Ferreira, H. B., Zhang, Y., Zhao, C. and Emmons, S. W. (1999). Patterning of *Caenorhabditis elegans* posterior structures by the *Abdominal-B* homolog, *egl-5*. *Dev. Biol.* **207**, 215-228.

Finney, M., Ruvkun, G. and Horvitz, H. R. (1988). The *C. elegans* cell lineage and differentiation gene *unc-86* encodes a protein with a homeodomain and extended similarity to transcription factors. *Cell* **55**, 757-769.

Finney, M. and Ruvkun, G. (1990). The *unc-86* gene product couples cell lineage and cell identity in *C. elegans*. *Cell* **63**, 895-905.

Fire, A., Harrison, S. W. and Dixon, D. (1990). A modular set of lacZ fusion vectors for studying gene expression in *Caenorhabditis elegans*. *Gene* **93**, 189-198.

Fukuta, M., Matsuno, K., Hui, C.-C., Nagata, T., Takiya, S., Xu, P.-X., Ueno, K. and Suzuki, Y. (1993). Molecular cloning of a POU domain-containing factor involved in the regulation of the *Bombyx* sericin-1 gene. *J. Biol. Chem.* **268**, 9471-19475.

Gehring, W. J., Affolter, M. and Bürglin, T. R. (1994). Homeodomain Proteins. *Annu. Rev. Biochem.* **63**, 487-526.

Greenstein, D., Hird, S., Plasterk, R. H. A., Andachi, Y., Kohara, Y., Wang, B., Finney, M. and Ruvkun, G. (1994). Targeted mutations in the *Caenorhabditis elegans* POU homeo box gene *ceh-18* cause defects in oocyte cell cycle arrest, gonad migration, and epidermal differentiation. *Genes Dev.* **8**, 1935-1948.

Hara, Y., Rovescalli, A. C., Kim, Y. and Nirenberg, M. (1992). Structure and evolution of four POU domain genes expressed in mouse brain. *Proc. Natl. Acad. Sci. USA* **89**, 3280-3284.

Hauptmann, G. and Gerster, T. (2000). Combinatorial expression of zebrafish *Brn-1*- and *Brn-2*-related POU genes in the embryonic brain, pronephric primordium, and pharyngeal arches. *Dev. Dyn.* **218**, 345-358.

He, X., Treacy, M. N., Simmons, D. M., Ingraham, H. A., Swanson, L. W. and Rosenfeld, M. G. (1989). Expression of a large family of POU-domain regulatory genes in mammalian brain development. *Nature* **340**, 35-42.

He, X., Gerrero, R., Simmons, D. M., Park, R., Lin, C. R., Swanson, L. W. and Rosenfeld, M. G. (1991). Tst-1, a member of the POU Domain gene family, binds the promoter of the gene encoding the cell surface adhesion molecule Po. *Mol. Cell. Biol.* **11**, 1739-1744.

Herr, W., Sturm, R. A., Clerc, R. G., Corcoran, L. M., Baltimore, D.,

Sharp, P. A., Ingraham, H. A., Rosenfeld, M. G., Finney, M., Ruvkun, G. et al. (1988). The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and *Caenorhabditis elegans unc-86* gene products. *Genes Dev.* **2**, 1513-1516.

Hobert, O., Tessmar, K. and Ruvkun, G. (1999). The *Caenorhabditis elegans lim-6* LIM homeobox gene regulates neurite outgrowth and function of particular GABAergic neurons. *Development* **126**, 1547-1562.

Ingraham, H. A., Chen, R., Mangalam, H. J., Elsholtz, H. P., Flynn, S. E., Lin, C. R., Simmons, D. M., Swanson, L. and Rosenfeld, M. G. (1988). A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. *Cell* **55**, 519-529.

Johnson, W. A. and Hirsh, J. (1990). Binding of a *Drosophila* POU-domain protein to a sequence element regulating gene expression in specific dopaminergic neurons. *Nature* **343**, 467-470.

Mathis, J. M., Simmons, D. M., He, X., Swanson, L. W. and Rosenfeld, M. G. (1992). Brain 4: a novel mammalian POU domain transcription factor exhibiting restricted brain-specific expression. *EMBO J.* **11**, 2551-2561.

Monuki, E. S., Weinmaster, G., Kuhn, R. and Lemke, G. (1989). SCIP: a glial POU domain gene regulated by cyclic AMP. *Neuron* **3**, 783-793.

Monuki, E. S., Kuhn, R., Weinmaster, G., Trapp, B. D. and Lemke, G. (1990). Expression and activity of the POU transcription factor SCIP. *Science* **249**, 1300-1303.

Nakachi, Y., T., H., Oota, H., Sumiyama, K., Wang, L. and Ueda, S. (1997). Nucleotide compositional constraints on genomes generate alanine-, glycine-, and proline-rich structures in transcription factors. *Mol. Biol. Evol.* **14**, 1042-1049.

Nakai, S., Kawano, H., Yudate, T., Nishi, M., Kuno, J., Nagata, A., Jishage, K., Hamada, H., Fujii, H., Kawamura, K. et al. (1995). The POU domain transcription factor Brn-2 is required for the determination of specific neuronal lineages in the hypothalamus of the mouse. *Genes Dev.* **9**, 3109-3121.

Nelson, F. K. and Riddle, D. L. (1984). Functional study of the *Caenorhabditis elegans* secretory-excretory system using laser microsurgery. *J. Exp. Zool.* **231**, 45-56.

Oka, T., Yamamoto, R. and Rutai, M. (1997). Three *vha* genes encode proteolipids of *Caenorhabditis elegans* vacuolar-type ATPase. *J. Biol. Chem.* **272**, 24387-24392.

Ono, M. and Takayama, Y. (1992). Structures of cDNAs encoding chum salmon pituitary-specific transcription factor, Pit-1/GHF-1. *Gene* **226**, 275-279.

Priess, J. R. and Hirsh, D. I. (1986). *Caenorhabditis elegans* morphogenesis: the role of the cytoskeleton in elongation of the embryo. *Dev. Biol.* **117**, 156-173.

Quiring, R., Walldorf, U., Kloter, U. and Gehring, W. J. (1994). Homology of the *eyeless* gene of *Drosophila* to the *Small eye* gene in mice and *Aniridia* in humans. *Science* **265**, 785-789.

Robertson, H. M. (1998). Two large families of chemoreceptor genes in the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae* reveal extensive gene duplication, diversification, movement, and intron loss. *Genome Res.* **8**, 449-463.

Rose, K. L., Winfrey, V. P., Hoffman, L. H., Hall, D. H., Furuta, T. and Greenstein, D. (1997). The POU gene *ceh-18* promotes gonadal sheath cell differentiation and function required for meiotic maturation and ovulation in *Caenorhabditis elegans*. *Dev. Biol.* **192**, 59-77.

Ryan, A. K. and Rosenfeld, M. G. (1997). POU domain family values: flexibility, partnerships, and developmental codes. *Genes Dev.* **11**, 1207-1225.

Saló, E., Muñoz-Mármol, A. M., Bayascas, J. R., Garcia-Fernàndez, J., Miralles, A., Casali, A., Cocominas, M. and Baguñá, J. (1995). The freshwater planarian Dugesia (G) tigrina contains a great diversity of homeobox genes. *Hydrobiologia* **305**, 269-275.

Salser, S. J. and Kenyon, C. (1996). A *C. elegans* Hox gene switches on, off, on and off again to regulate proliferation, differentiation and morphogenesis. *Development* **122**, 1651-1661.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Schonemann, M. D., Ryan, A. K., McEvilly, R. J., O'Connell, S. M., Arias, C. A., Kalla, K. A., Li, P., Sawchenko, P. E. and Rosenfeld, M. G. (1995). Development and survival of the endocrine hypothalamus and posterior pituitary gland requires the neuronal POU domain factor Brn-2. *Genes Dev.* **9**, 3122-3135.

Seimiya, M., Watanabe, Y. and Kurosawa, Y. (1997). Identifaction of POU-class homeobox genes in a freshwater sponge and the specific expression of these genes during differentiation. *Eur. J. Biochem.* **243**, 27-31.

Sharman, A. C. and Holland, P. W. H. (1996). Conservation, duplication, and divergence of developmental genes during chordate evolution. *Neth. J. Zool.* **46**, 47-67.

Sluder, A. E., Mathews, S. W., Hough, D., Yin, V. P. and Maina, C. V. (1999). The nuclear receptor superfamily has undergone extensive proliferation and diversification in nematodes. *Genome Res.* **9**, 103-120.

Smith, D. B. and Johnson, K. S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione *S*-transferase. *Gene* **67**, 31-40.

Spaniol, P., Bornmann, C., Hauptmann, G. and Gerster, T. (1996). Class III POU genes of zebrafish are predominantly expressed in the central nervous system. *Nucl. Acids Res.* **24**, 4874-4881.

Sturm, R. A., Das, G. and Herr, W. (1988). The ubiquitous octamer-binding protein Oct-1 contains a POU domain with homeo box subdomain. *Genes Dev.* **2**, 1582-1599.

Sulston, J. E., Schierenberg, E., White, J. G. and Thomson, J. N. (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans. Dev. Biol.* **100**, 64-119.

Sumiyama, K., Washio-Watanabe, K., Saitou, N., Hayakawa, T. and Ueda, S. (1996). Class III POU genes: generation of homopolymeric amino acid repeats under GC pressure in mammals. *J. Mol. Evol.* **43**, 170-178.

The *C. elegans* Sequencing Consortium. (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012-2018.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl. Acids Res.* **25**, 4876-4882.

Wang, B. B., Müller-Immergluck, M. M., Austin, J., Robinson, N. T., Chisholm, A. and Kenyon, C. (1993). A homeotic gene cluster patterns the anteroposterior body axis of *C. elegans. Cell* **74**, 29-42.

Yamada, S., Hata, J. and Yamashita, S. (1993). Molecular cloning of fish Pit-1 cDNA and its functional binding to promoter of gene expressed in the pituitary. *J. Biol. Chem.* **268**, 24361-24366.

Zhou, H., Yoshioka, T. and Nathans, J. (1996). Retina-derived POU-domain factor-1: a complex POU-domain gene implicated in the development of retinal ganglion and amacrine cells. *J. Neurosci.* **16**, 2261-2274.

Zwaal, R. R., Broeks, A., Van Meurs, J., Groenen, J. T. M. and Plasterk, R. H. (1993). Target-selected gene inactivation in *Caenorhabditis elegans* by using a frozen transposon insertion mutant bank. *Proc. Natl. Acad. Sci. USA* **90**, 7431-7435.