Evolution of discrete Notch-like receptors from a distant gene duplication in Caenorhabditis

David Rudel a,1 and Judith Kimble a,b,c,d,*

Departments of aBiochemistry and bMedical Genetics and cLaboratory of Molecular Biology, University of Wisconsin–Madison, Madison, WI 53706, USA
dHoward Hughes Medical Institute, 433 Babcock Drive, University of Wisconsin–Madison, Madison, WI 53706-1544, USA

*Author for correspondence (email: jekimble@facstaff.wisc.edu)
1Present address: Max-Planck Institut für Entwicklungsbiologie, Abt. Evolutionsbiologie, Spemannstrasse 35, 72076 Tübingen, Germany

SUMMARY Caenorhabditis elegans possesses two Notch-like receptors, LIN-12 and GLP-1, which have both overlapping and individual biological functions. We examined the lin-12 and glp-1 genes in closely related nematodes to learn about their evolution. Here we report molecular and functional analyses of lin-12 orthologs from two related nematodes, C. briggsae (Cb) and C. remanei (Cr). In addition, we compare these lin-12 findings with similar studies of Cb-glp-1 and Cr-glp-1 orthologs. Cb-LIN-12 and Cr-LIN-12 retain the same number and order of motifs as Ce-LIN-12. Intriguingly, we find that LIN-12 conservation differs from that of GLP-1 in two respects. First, individual motifs are conserved to a different degree for the two receptors. For example, the transmembrane domain is 16–32% identical among LIN-12 orthologs but 65–70% identical among GLP-1 orthologs. Second, certain amino acids are conserved in a receptor-specific manner, a phenomenon most prevalent in the CC-linker. We suggest that LIN-12 and GLP-1 have been molded by selective constraints that are receptor specific and that the two proteins may not be entirely interchangeable. To analyze the functions of the lin-12 orthologs, we used RNA-mediated interference (RNAi). Cb-lin-12(RNAi) or Cr-lin-12(RNAi) progeny are nearly 100% Lag, a larval lethality typical of C. elegans lin-12 glp-1 double mutants, but not the primary defect observed in Ce-lin-12 null mutants or Ce-lin-12(RNAi). Therefore, LIN-12 functions are similar, but not identical, among the Caenorhabditis species. We suggest that ancestral functions may have been divided between LIN-12 and GLP-1 receptors in a process contributing to the retention of both genes after gene duplication (i.e., subfunctionalization).

INTRODUCTION

Gene duplication is a major mechanism by which regulatory pathways can diverge and adopt new biological functions during evolution. Estimates of gene duplication in eukaryotic genomes give an average rate of 0.01 duplications per gene per million years (Lynch and Conery 2000), a rate of the same magnitude as the rate of mutation per nucleotide site (Li 1999). In addition, approximately 49% of Caenorhabditis elegans coding sequences and 38% of human coding sequences may have arisen by gene duplication (Rubin et al. 2000; Li et al. 2001). By the simplest scenario, a gene encoding one component of a regulatory pathway is duplicated to create two genes that are initially redundant. After duplication, the two genes can acquire distinct functions as a result of mutation and selection. Such alterations of genes encoding developmentally critical regulatory proteins are thought to contribute to the generation of new species.

From the avian feather to the nematode vulva, virtually all animals use the Notch-signaling pathway to pattern and specify a variety of tissues and structures. In addition to such distinct roles in diverse organisms, multiple Notch-like receptors with disparate functions are often present within a single genome; for example, mouse has at least four Notch-like receptors (del Amo et al. 1993; Lardelli and Lendahl 1993; Lardelli et al. 1994; Uyttendaele et al. 1996) and nematodes two (Yochem et al. 1988; Austin and Kimble 1989; Yochem and Greenwald 1989; The C. elegans Sequencing Consortium 1998). Although comparisons of receptors from distantly related organisms (e.g., fly and worm) have revealed essential core features of the receptors and widely divergent biological functions, they have failed to explain the origin of distinct biological functions. However, comparison of closely related receptors (e.g., orthologs from closely related organisms such as C. elegans vs. C. briggsae) has revealed small changes that provide a window into recent evolutionary events. We have begun to analyze the evolution of Notch-like receptors in Caenorhabditis nematodes (Rudel and Kimble 2001; this work). To our knowledge, a similar study has not been reported for close Drosophila or vertebrate species.
The most commonly studied nematode, *C. elegans*, possesses two Notch-related receptors, LIN-12 and GLP-1 (for review see Kimble and Simpson 1997). These two nematode proteins have the same overall architecture as all Notch receptors. They are transmembrane proteins with extracellular epidermal growth factor (EGF) repeats and three familial LIN-12/notch repeats (LNR) motifs followed by a linker region, termed the CC linker, that lies N-terminal to a single transmembrane domain (TMD); intracellularly, they possess a RAM domain, six ankyrin (ANK) repeats capped with degenerate ANK repeats, and a PEST sequence. Most Notch-related receptors have 36 EGF repeats; LIN-12 and GLP-1 differ in their number of EGF repeats, with 13 and 10, respectively. We found previously that the number and order of motifs are conserved among GLP-1 orthologs from *C. briggsae* and *C. remanei* (Rudel and Kimble 2001). Notch-like receptors are thought to exist as heterodimers composed of products generated by cleavage in the CC-linker (Blaumueller et al. 1997); additionally, individual Notch-like receptors may interact with other Notch-like receptors in the cell membrane (Kidd et al. 1989; Greenwald and Seydoux 1990).

The LIN-12 and GLP-1 receptors signal by a conserved pathway (for review see Kimble and Simpson 1997). Their ligands are members of the DSL (Delta, Serrate, and LAG-2) family. Best characterized is LAG-2, which signals to both receptors (Lambie and Kimble 1991; Henderson et al. 1994; Tax et al. 1994; Wilkinson et al. 1994). Upon signaling, the receptor’s intracellular domain is thought to be cleaved (Kopan et al. 1996; Schroeter et al. 1998; Struhl and Adachi 1998) and to bind, via their RAM and ANK domains, to the LAG-1 DNA binding protein (a CSL transcription factor, for CBF1, Su(H), and LAG-1), and the LAG-3 transcriptional coactivator (also known as SEL-8) (Christensen et al. 1996; Roehl et al. 1996; Doyle et al. 2000; Petcherski and Kimble 2000). The interaction of LIN-12 and GLP-1 with common pathway components (e.g., LAG-1, LAG-2, LAG-3) indicates that these two receptors transduce signals by a common biochemical mechanism.

The LIN-12 and GLP-1 receptors are thought to be functionally interchangeable. The first evidence supporting this idea was genetic. Most *lin-12* and *glp-1* single mutants survive to adulthood, whereas *lin-12 glp-1* double mutants die as first-stage larvae (L1) with a characteristic Lag phenotype (Lambie and Kimble 1991). The simplest interpretation was that in single mutants, GLP-1 substitutes for LIN-12 and vice versa, but in the double mutant, which has no receptor activity, the animal dies. The second evidence was a swapping experiment. The glp-1 coding region was placed under control of *lin-12* regulatory sequences and found to rescue a *lin-12* null mutant (Fitzgerald et al. 1993). Therefore, GLP-1 can replace LIN-12 in *lin-12*–specific functions. These results, when coupled to the knowledge that LIN-12 and GLP-1 interact with the same molecular machinery to transduce signals (e.g., LAG-1, LAG-2, LAG-3), strongly supported the idea that the LIN-12 and GLP-1 receptors are functionally equivalent.

The LIN-12 and GLP-1 receptors control numerous events throughout development. Some are receptor specific, and as described above, others are common to the two receptors. Shared functions were inferred from *lin-12 glp-1* double mutants and individual functions from *lin-12* and *glp-1* single mutants. Common functions include proper development of the excretory pore, rectum, and snout (Lambie and Kimble 1991). LIN-12–specific functions include control of left–right asymmetry of the embryonic gut (Hermann et al. 2000) and vulval, uterine, and proximal germline fates in larvae (Greenwald et al. 1983; Seydoux et al. 1990; Newman et al. 1995). GLP-1–specific functions include control of germline proliferation and early embryonic fates (Austin and Kimble 1987; Priess and Thomson 1987). Given the functional interchangeability of these receptors (see above), the receptor-specific functions have been thought to rely on regulatory sequences that control the two receptors differentially in particular tissues.

We focus on the Notch-like receptors in *C. briggsae* and *C. remanei*, the closest known relatives of *C. elegans*. We chose these species because they are phylogenetically distant enough to have acquired differences yet are close enough that these differences may be interpretable. Previously, we characterized the GLP-1 receptors from the same two species (Rudel and Kimble 2001). Here we report our characterization of *lin-12* genes from *C. briggsae* and *C. remanei* and compare their conservation with that among GLP-1 orthologs. Surprisingly, we find that conservation of LIN-12 and GLP-1 proteins differs, at least in certain respects. As a result, we propose that these genes have experienced different selective constraints during the course of evolution and may not be as functionally equivalent as previously thought. We also find that LIN-12 is essential in *C. briggsae* and *C. remanei* for functions that can be executed by either LIN-12 or GLP-1 in *C. elegans*. We hypothesize that the ancestral common roles of LIN-12 and GLP-1 have been subdivided, a process known as subfunctionalization, and that this division has played a role in the retention and subsequent evolution of these two genes from an ancestral Notch-like receptor.

**MATERIALS AND METHODS**

**Strains**

*C. elegans* Bristol N2, *C. briggsae* AF16, and *C. remanei* SB146 were used for this study. Animals were manipulated using traditional *C. elegans* techniques (Brenner 1974) and were maintained at 20°C.

**Cloning Cb-lin-12**

A portion of the *Cb-lin-12* sequence was previously reported (Wilkinson et al. 1994). A fragment was synthesized using *C. briggsae* genomic DNA and primers Br-lin-12(F) (5′-ATGCGGAATTT-
TAGTTGTTTGGCTCTTTT-3') and Br-lin-12(RI) (5'-AAATTACA-
TAACCTGCACAAAAAGTACGAGAGCAG-3') in a 100-μl polymerase chain reaction (PCR) reaction (500 nM Br-lin-12(F), 500 nM Br-lin-12(RI), 250 nM dNTPs, 1× Taq buffer, 0.5 μg DNA, 2.5 units Roche (Roche Diagnostics GmbH, Mannheim, Germany) Taq DNA poly-
merase). The fragment was used to screen a C. briggsae fosmid pool (Incyte Genomics). Ten hybridizing fosmids (27c10, 47b08, 8d9, 1612, 2032, 13o10, 18d18, 42i14, 47L13, and 13L22) were ordered. (Incyte Genomics). Ten hybridizing fosmids (27c10, 47b08, 8d9, 1612, 43a2, 13o10, 18d18, 42i14, 47L13, and 13L22) were ordered.

RACE using oligo(dT) primed cDNA. disparate was found to be likely to contain the entire 16i2, 43a2, 13o10, 18d18, 42i14, 47L13, and 13L22) were ordered. (Incyte Genomics). Ten hybridizing fosmids (27c10, 47b08, 8d9,

Library screening

A C. remanei cDNA library (Rudel and Kimble 2001) and a C. re-
manei genomic library (Haag and Kimble 2000) were screened using standard lift and hybridization techniques (Sambrook et al. 1989).

Southern analysis

Age I, Apa I, EcoR I, EcoR V, Hind III, Kpn I, Pvu II, Spe I, Xba I, and Xho I were each used to restriction digest 12 μg of C. briggsae genomic DNA. Reactions were allowed to proceed 36 h and run out on a 1% agarose gel. The gel was blotted using standard techniques (Sambrook et al. 1989). The membrane was probed with purified Cb-lin-12.1 RT-PCR product generated using primers BL22 (5'-TCTCAATGGAATCGTCA-3') and BL24 (see above) and subse-
doctoral probes. The 5' end of Cb-lin-12 was cloned by reverse transcriptase (RT)-PCR using a primer corresponding to the SL1 transplice leader as an anchor primer and primers RL32 (5'-TGAAGACATCGTGAGGAGGGA-3') and RL30 (5'-TGGGGACAGCCAGGCT-3') in seminested PCR reactions using oligo(dT)-primed cDNA. Additional sequence was obtained from Cb-lin-12 genomic phage and coding sequences were inferred based on conservation with Ce-lin-12. The Cb-lin-12 cDNA (accession number AF499439) has been deposited in Gen-

Cloning Ce-lin-12

Degenerate primers were designed based on the alignment of the C. elegans glp-1 and lin-12 genes. Forward primer TF1 (5'-GART-
GYAAAAYGARGARTG-3') anneals within the sequence encoding the second LNR of lin-12 and reverse primer DR3 (5'-GTIWYTTCAITGGCCATCCA-3') anneals in the opposite orientation within the sequence encoding the RAM domain of lin-12 (degenerate bases are shown using IUPAC code, I represents inosine). Primers TF1 and DR3 were used in a 100-μl PCR reaction (500 nM TF1, 500 nM DR3, 250 nM dNTPs, 1× Taq extender buffer, 2 μg C. remanei genomic DNA, 2.5 units Roche Taq DNA polymerase, and 2.5 units Stratagene (La Jolla, CA) Taq extender). The reaction was cycled using a step-down program: 10 times (94°C, 1 min; 49°C, 1 min, annealing temperature was lowered 1°C each cycle; 72°C, 2

RNA interference

Double-stranded RNA (dsRNA) made to Ce-lin-12, Cb-lin-12, or Cr-lin-12 was injected at a concentration of 1 μg/μl into C. elegans hermaphrodites, C. briggsae hermaphrodites, or gravid C. remanei females, respectively. For Ce-lin-12, the dsRNA was 1036 bp from the region encoding the 3rd EGF repeat through the 11th EGF repeat. For Cb-lin-12, the dsRNA was 263 bp and corresponded to the 5'UTR and coding sequence of the first exon. For Cr-lin-12, the dsRNA was 916 bp from the region encoding the last two LNRs, CC-linker, TMD, and RAM domain. Injected animals were allowed to rest on a bacterial lawn for 8 h and then passed to new seeded plates every 12 h. Additional RNA-mediated interference (RNAi) experiments were performed in which the dsRNA concentration was varied between 0.2 and 5 μg/μl to try to unmask other phenotypes in C. briggsae and C. remanei; the results of these experiments were similar to those with dsRNA concentrations of 1 μg/μl.

Immunostaining

Wild-type embryos obtained by bleaching were washed twice with M9, placed on unseeded plates, and allowed to hatch for 24 h. L1 larvae were harvested in M9 and concentrated in a microcentrifuge. Approximately 100 L1 larvae in 12 μl of 2% paraformaldehyde were placed on a polylysine subbed slide, and a second subbed slide placed face down on the first, and placed on dry ice for 10 min for
freeze cracking. The bottom slide was placed in −20°C methanol (5 min), transferred to −20°C acetone (5 min), air dried, and then blocked in 0.5% bovine serum albumin in phosphate-buffered saline for 30 min. Larvae were immunostained with either mouse monoclonal antibody 9H27 (Waterston 1988) or M5-6 (Miller et al. 1983) for 6–12 h. Fixed larvae were washed twice with 0.5% bovine serum albumin in phosphate-buffered saline and then incubated with Cy-3 conjugated donkey anti-mouse antibodies (Jackson ImmunoResearch Laboratories, West Grove, PA, USA) and 4′,6-diamidino-2-phenylindole for 1 h. For Cb-lin-12(RNAi) and Cr-lin-12(RNAi) progeny, 30 gravid C. briggsae hermaphrodites or C. remanei females, respectively, were injected with dsRNA as described; animals recovered on a bacterial lawn for 8 h and then were transferred to unseeded plates. Thirty C. remanei male adults were added to the C. remanei plate. Egg laying was allowed for 2 days at 20°C, adults were removed, and remaining eggs were given 1 day to hatch. Larvae were washed off plate in M9 and immunostained as described above.

**Sequence and phylogenetic analyses**

DNA sequences were determined by standard methods, with sequence obtained at least once from each strand. Sequence analyses relied on EDITSEQ, SEQMAN, MAPDRAW, and MEGALIGN (DNASTAR Inc., Madison, WI, USA). Amino acid alignments, nucleotide alignments, and uncorrected pair-wise distances were computed using SEQUED, PILEUP, LINEUP, and DISTANCES (Wisconsin Package, version 10, Genetics Computer Group, Madison, WI, USA). Sequences were trimmed on their 5′ ends so that only directly alignable sequences were used. The sequences were entered using SEQUED and aligned with PILEUP using a decreased penalty value for gap creation. The alignment was edited by eye in LINEUP to ensure the proper alignment of motif-specific conserved amino acids and to take into account knowledge of specific domain function. Phylogenetic analysis of the alignments was done using PAUP, version 4.0b8, for the PPC (Swofford 1998) and over the web using PROTOPARS, which is part of the PHYLIP software package (http://bioweb.pasteur.fr/seqanal/phylogeny/phylip-uk.html). Gaps were treated as additional characters in parsimony analysis. All neighbor-joining (NJ) analyses used standard distances.

**RESULTS**

**Cloning Cb-lin-12 and Cr-lin-12**

We cloned lin-12 orthologs from C. briggsae and C. remanei. We identified these genes as homologs of lin-12 rather than glp-1 by their genomic organization (Fig. 1A), the presence of 13 EGF repeats (Fig. 2), and a higher sequence identity with Ce-lin-12 at the amino acid and nucleotide levels (Fig. 3; not shown). In addition, the position of lin-12 with respect to the 3′ flanking gene, glutathione S-transferase, is conserved between C. briggsae and C. elegans (see Materials and Methods).

The C. briggsae genome appears to encode two lin-12-like genes. We dub the Cb-lin-12 gene reported in detail here as Cb-lin-12.1 but refer to it throughout as Cb-lin-12 for simplicity. Its nearly identical paralog we dub Cb-lin-12.2; a more thorough analysis of Cb-lin-12.2 is beyond the scope of this article. These two genes were initially found using RT-PCR to identify the Cb-lin-12 5′ end; these reactions reproducibly resulted in two products that were approximately 96% identical at the nucleotide level (Fig. 1B); one product (lin-12.2) lacked exon 3. Because exon 2 can splice in frame to exon 4, the gene lacking exon 3 should make a protein missing a portion of the T+Y motif. Consistent with the idea of two lin-12 genes, a Southern blot of C. briggsae genomic DNA revealed more bands than predicted for Cb-lin-12.1 (Fig. 1C). Because RNAi depletes any mRNAs with sequences > 75% identical to the RNA injected and because the dsRNA used for Cb-lin-12.1(RNAi) was 96% identical to Cb-lin-12.2, this experiment is expected to reduce both Cb-lin-12.1 and Cb-lin-12.2 mRNAs. A similar recent duplication of Cr-lin-12 has not been detected. Although we favor the idea of a duplicated Cb-lin-12, an alternative that cannot be ruled out is that the two transcripts represent allelic variation. This is not likely because C. briggsae is a hermaphroditic laboratory strain and most loci should be in a homozygous state.

**Conservation of lin-12 genes and mRNAs**

The complete Cb-lin-12 genomic sequence was obtained and its exon/intron structure deduced. All Cb-lin-12 exon boundaries have been conserved with those of C. elegans except for the boundary between exons 6 and 7, which are fused in C. briggsae (this work, Fig. 1A) (Yochem et al. 1988). Genomic sequence from C. remanei shows that exons 6 and 7 are not fused but instead retain the same splicing sites as Ce-lin-12 (data not shown). Thus, C. briggsae appears to have recently lost an intron. Like Ce-lin-12 (Christensen et al. 1996), Cb-lin-12 contains numerous consensus LAG-1 binding sites (rtggga): 14 sites within 7.2 kb (Fig. 1A, arrowheads), whereas sites are predicted once every 4.1 kb at random. The Cb-lin-12 LAG-1 binding sites are concentrated in the 5′ flanking region and the large 5′-most introns, consistent with the idea that transcriptional feedback regulation has been conserved. Based on genomic and cDNA sequences, the Cb-lin-12 and Cr-lin-12 mRNAs are predicted to be 4521 and 4647 nt, respectively, including an SL1 transspliced leader, as confirmed by RT-PCR (see Materials and Methods).

**glp-1/lin-12 duplication occurred before radiation of C. elegans, C. briggsae, and C. remanei**

A phylogram of LIN-12 and GLP-1 orthologs was constructed using Drosophila Notch as an outgroup representative in both an exhaustive maximum parsimony (MP) analysis of 945 total trees (Fig. 3A) and an NJ analysis. A single topology was obtained from both analyses. The LIN-12 orthologs grouped together and the GLP-1 orthologs grouped together, confirming that the lin-12/glp-1 duplication occurred before the divergence of these three species. The
branch lengths in the LIN-12 clade are smaller and reflect a higher degree of conservation among LIN-12 proteins than those in the GLP-1 clade. This grouping is strongly supported by high bootstrap values: for MP analysis, 1000 replicates yielded values of 100% for GLP-1 and 92% for LIN-12 (Fig. 3A). A tree length of 2388 character changes was given by MP analysis. Topologies that force a *C. briggsae*/*C. remanei* clade give trees with longer lengths (2436 and 2416 character changes, respectively). A weighted maximum parsimony analysis using PROTOPARS supported the topology from the PAUP MP and NJ analysis with similar bootstrap values. Additionally, MP and NJ analyses of a corresponding nucleotide analysis gave a single tree with the same topology and similar bootstrap values. The topology of the LIN-12 clade confirms our earlier hypothesis, which was based on analyses of GLP-1 homologs, about the evolutionary relationships among these species: *C. briggsae* and *C. remanei* form a clade with *C. elegans* as its sister taxon (Rudel and Kimble 2001). This topology is further supported by other studies (Fitch et al. 1995; Haag and Kimble 2000; Chen et al.
Fig. 2. Amino acid alignment of Cb-lin-12, Cr-lin-12, and Ce-lin-12. Conservation between two or more proteins is represented by solid boxes. Identity among all three proteins is shown in the consensus. Individual motif types are color coded: purple, EGF repeats; pink, T+Y box; dark blue, LNR repeats; light blue, conserved cysteins in the CC-linker; dark green, TMD; light green, RAM domain; gold, putative ANK capping motifs (Rudel and Kimble 2001); orange, ANK repeats; red, PEST sequence. Downward pointing black arrowheads represent divisions between repeats. Downward pointing red arrowhead represents the predicted site of intracellular receptor cleavage (Schroeter et al. 1998).
Fig. 2. continued. The end of an EGF repeat is defined by the last conserved cysteine [C], usually followed by a glutamic acid [E] or aspartic acid [D]. Divisions between LNRs are defined by the first conserved cysteine in the repeat. Divisions between ANK repeats are defined by a semiconserved glycine and alanine in an alignment between ANK repeats. The LIN-12 RAM domain is defined by identity and analogy with the GLP-1 RAM domain. Sites of *lin-12* mutations are shown above the alignment as follows: nonsense mutations, red dot; missense mutations, letters corresponding to the amino acid predicted by the lesion in the genomic sequence (Greenwald and Seydoux 1990; Wen and Greenwald 1999; I. Greenwald, personal communication); black letters are loss-of-function and green letters are gain-of-function alleles. Allele names are given in the order the mutations appear (N-terminal to C-terminal) and mutations in the same codon but not necessarily the same nucleotide change are separated with a slash: *n930* (C138T), *ar170* (G270R), *n941* (W400stop), *oz48* (G449R), *n952* (R672W), *n302* (G747D), *n137/n1005* (S872F), *n950* (A873T), *n177* (L875F), *n676* (G884D), *n379* (E889K), *n653* (A1171V).

Conservation of LIN-12 and GLP-1 receptor sequences

LIN-12 and GLP-1, like all Notch-related receptors, have a variety of motifs with known functions. The EGF motifs are thought to interact with the DSL ligands (for review see Artavanis-Tsakonas et al. 1995), the LNR motifs mediate receptor activation upon signaling (Greenwald and Seydoux 1990; Lyman and Young 1993; Berry et al. 1997; Rand et al. 2000), the CC-linker is cleaved to produce mature receptor (Blaumueller et al. 1997; Logeat et al. 1998) and may mediate receptor dimerization (Kidd et al. 1989; Greenwald and Seydoux 1990), the PEST domain down-regulates the receptor (Mango et al. 1991), and the RAM domain and ANK repeats are required for assembly of a ternary complex that activates transcription of target genes (Roehl and Kimble 1993; Roehl et al. 1996; Petcherski and Kimble 2000). This motif architecture is conserved among the three LIN-12 orthologs (Fig. 2). In particular, all three LIN-12 orthologs have 13 EGF repeats, with each repeat exhibiting a characteristic sequence. Conservation of EGF repeat number is also conserved among the Ce-, Cb-, and Cr-GLP-1 orthologs, which all have 10 repeats with characteristic sequences (Rudel and Kimble 2001) (Fig. 2). The remarkable conservation of most of the LIN-12 and GLP-1 domains is likely to reflect highly specific protein–protein interactions that are critical for receptor functions.

Over most of the protein, LIN-12 is better conserved than GLP-1. The overall amino acid identity among the LIN-12 proteins in *C. elegans*, *C. briggsae*, and *C. remanei* is 68–73% (Fig. 3B). This compares with an overall amino acid identity among GLP-1 proteins from the same three species of 61–68% (Fig. 3B). Comparing individual motifs (Fig. 4), the EGF, LNR, CC-linker, and PEST domains are most conserved among LIN-12 receptors and the LNR, CC-linker, TMD, and PEST domains are most conserved among GLP-1 receptors. The ANK repeats are also well conserved in both LIN-12 and GLP-1, with the RAM domain being the least conserved domain. LIN-12 and GLP-1 have different patterns of conservation. Three observations stand out in particular. First, the extracellular portion of LIN-12 is much more conserved than that of GLP-1, that is, the EGF repeats, with each repeat exhibiting a characteristic sequence. Conservation of EGF repeat number is also conserved among the Ce-, Cb-, and Cr-GLP-1 orthologs, which all have 10 repeats with characteristic sequences (Rudel and Kimble 2001) (Fig. 2). The remarkable conservation of most of the LIN-12 and GLP-1 domains is likely to reflect highly specific protein–protein interactions that are critical for receptor functions.
identity vs. 56–65% identity) and the LNR motifs (85–86% identity vs. 67–74% identity). Second, the TMD of LIN-12 is poorly conserved (16–32% identity) compared with that of GLP-1 (60–70% identity). Third, within the CC-linkers of LIN-12 and GLP-1 are amino acids that are absolutely conserved among orthologs (whether LIN-12 or GLP-1) but not between paralogs. The number of receptor-specific amino acids is greater in the CC-linker than any other region: approximately 33% of total receptor-specific amino acids occur in the CC-linker (data not shown), which comprises only approximately 14% of the aligned sequence (Fig. 4, see legend). These striking differences in conservation suggest that LIN-12 and GLP-1 have been molded during evolution, at least in part, by receptor-specific constraints (see Discussion).

C. briggsae and C. remanei animals with reduced lin-12 activity have a Lag-like phenotype
To examine the biological functions of Cb-lin-12 and Cr-lin-12, we used RNAi. As a control, we first examined Ce-lin-12(RNAi) animals (n > 500). The most frequent Ce-lin-12(RNAi) defect was proximal proliferation of the germ line (approximately 50%, Fig. 5B, compare with A). A less frequent defect was a protruding vulva (approximately 12%, Fig. 5C, compare with A), though they were less pronounced than those observed in lin-12(0) mutants. Rarely, animals arrested as L1 larvae with Lag defects (Fig. 5D): a bump on the nose and/or tail, absence of an anus, and immobility (<1%). One Lag defect, a twisted snout, was not observed. Ce-lin-12(0) mutants exhibit proximal proliferation and a protruding vulva in virtually all adults, and 9% arrest as dead L1 larvae with a LAG phenotype (Greenwald et al. 1983; Lambie and Kimble 1991). Therefore, Ce-lin-12(RNAi) defects are similar but less penetrant than those observed in lin-12(0) mutants.

The Cb-lin-12(RNAi) and Cr-lin-12(RNAi) animals had a strikingly different phenotype from that observed in Ce-lin-12(RNAi) animals and in lin-12(0) mutants. C. briggsae hermaphrodites and gravid C. remanei females were injected with Cb-lin-12 dsRNA and Cr-lin-12 dsRNA, respectively. After a recovery window of 8 h, all RNAi progeny arrested as dead L1 larvae (Fig. 6, A and B). In C. elegans lin-12 glp-1 double mutants and lag-1 and lag-2 single mutants result in 100% larval lethality, but lin-12 or glp-1 single mutants do not (Lambie and Kimble 1991). Therefore, in contrast to Ce-lin-12, Cb-lin-12 and Cr-lin-12 are essential for larval viability.

The Cb-lin-12(RNAi) and Cr-lin-12(RNAi) dead larvae had defects characteristic of lin-12 glp-1 double mutants and lag-1 and lag-2 single mutants in C. elegans, a constellation of defects known as the Lag phenotype (Lambie and Kimble 1991). Thus, the Cb-lin-12(RNAi) and Cr-lin-12(RNAi) dead
larvae exhibited a bump on the head, consistent with the loss of a functional excretory pore, and a bump on the tail, consistent with the loss of a functional rectum. Like Ce-lin-12(RNAi) Lag larvae, one typical C. elegans Lag defect that was not observed for either Cb-lin-12(RNAi) or Cr-lin-12(RNAi) was a twisted snout. To further characterize the Lag-like phenotypes of Cb-lin-12(RNAi) and Cr-lin-12(RNAi) animals, we immunostained with MH5-6, which recognizes myosin heavy chain A, and MH27, which recognizes adherens junctions (see Materials and Methods) (Lambie and Kimble 1991). C. briggsae and C. remanei Lag animals exhibited characteristic transformations in the head (Fig. 6, C and D; not shown) and lacked an anal depressor muscle and a rectum visible by Nomarski microscopy (Fig. 6, E and F; not shown). Also, like C. elegans Lag animals, rarely a muscle with similar morphology to the anal depressor muscle was observed inappropriately in a position anterior to the sphincter (Fig. 6G). We conclude that the Lag phenotype observed for Cb-lin-12 and Cr-lin-12 is similar, but not identical, to that of Ce-lin-12.

It is unlikely that lin-12(RNAi) abolishes both lin-12 and glp-1 mRNAs in C. briggsae and C. remanei to result in Lag defects. The primary glp-1(RNAi) defect in all three species is embryonic arrest and loss of the anterior pharynx (Rudel

---

**Fig. 4.** Percent identities between individual LIN-12 and GLP-1 motifs. LIN-12 and GLP-1 receptors cartooned at left. With exception of the CC-linker, domains are demarcated by color-coded bars. The CC-linker extends from the last LNR to the TMD. Numbers in identity matrix are color-coded by motif type as in Fig. 2 and receptor cartoon.
and Kimble 2001), a phenotype not observed with lin-12(RNAi). Furthermore, >75% nucleotide identity is required to reduce gene function by RNAi in C. briggsae and C. remanei (M. Montgomery, personal communication), and the regions used for Cb-lin-12(RNAi) and Cr-lin-12(RNAi) both have approximately 50% identities with their glp-1 counterparts. Likewise, because Cb-lin-12.1 and Cb-lin-12.2 are 95% identical over the length of the dsRNA used in the experiments, it is likely both genes are reduced in function. Thus, to dissect the roles of individual Cb-lin-12 paralogs, the generation of mutants is required.

In addition to the larval lethality observed in most Cb-lin-12(RNAi) or Cr-lin-12(RNAi) progeny, we also observed rare vulval defects among Cb-lin-12(RNAi) larval lethal survivors. From 60 injected C. briggsae animals, only 85 progeny survived to adulthood; all survivors were laid within the initial incubation period. Therefore, we isolated 25 L1 larvae that were laid within the initial incubation period and 85 L1 larvae that were laid after the incubation period. Without exception, the progeny of these 85 L1 larvae were larval lethal, whereas 50% of the progeny of the 25 L1 larvae laid within the initial incubation period were larval lethal.

**Fig. 6.** Cr-lin-12(RNAi) and Cb-lin-12(RNAi) Lag phenotypes. (A) Cb-lin-12(RNAi) L1 larvae. (B) Cr-lin-12(RNAi) L1 larvae. In both A and B, L1 larvae arrest with no excretory pore or anus. Arrows, head bump in position of excretory pore and tail bump near anus. (C and D) Wild-type C. remanei and Cr-lin-12(RNAi) animals stained with MH27, an antibody to adherens junctions. (C) Open arrowheads, one- and two-ring structures just anterior to posterior pharyngeal bulb. Insert, alternate focal plane. (D) Open arrowhead, multiring structure just anterior to posterior pharyngeal bulb. (E, F, and G) Wild-type C. remanei and Cr-lin-12(RNAi) animals with M5-6, an antibody to myosin heavy chain. Arrowheads, anal sphincter; arrows, abdominal depressor muscle. Anal depressor muscle was not detected in >99% of Lag animals. A misplaced depressor muscle was seen rarely (<1%). (H) Cb-lin-12(RNAi) adult hermaphrodite. Protruding vulva, open arrowhead.
tional recovery window. Two of the Cb-lin-12(RNAi) survivors had a protruding vulva (Fig. 6H), suggesting a conserved role for lin-12 in vulval development. From 60 injected C. remanei animals, only 33 Cr-lin-12(RNAi) survivors were obtained; no vulval defects were observed. No Cb-lin-12(RNAi) or Cr-lin-12 (RNAi) survivors exhibited proximal proliferation of the germ line, the most common defect in Ce-lin-12(RNAi) animals. Although the survivors may have escaped RNAi altogether, given the two C. briggsae animals observed with vulval defects, we tentatively conclude that the postembryonic functions of lin-12 are similar but not identical. Mutants are required to definitively assess the lin-12 postembryonic functions in these species.

DISCUSSION

The Ce-lin-12 and Ce-glp-1 genes are thought to have evolved by duplication (Yochem and Greenwald 1989) (Fig. 7, A and B, a). Our phylogenetic studies of glp-1 and lin-12 in closely related Caenorhabditis species support this hypothesis and reveal changes that suggest how these duplicated genes may have evolved to their present state. In the following discussion, we draw together our findings for lin-12 (this work) and glp-1 (Rudel and Kimble 2001) and propose a model for the recent evolution of these two duplicated genes. We also speculate about how these two genes may have evolved from a common ancestor.

Duplication of lin-12 and glp-1 from a common ancestor

A comparison of C. elegans lin-12 and glp-1 suggested their origin by gene duplication (Yochem and Greenwald 1989). We extended this analysis by comparing lin-12 and glp-1 from three closely related species: C. elegans, C. briggsae, and C. remanei. The exon/intron structures of these homologs share certain features and reveal others that are lin-12- or glp-1-specific (Yochem and Greenwald 1989; Rudel and Kimble 2001; this work, Fig. 7). Based on similarities and differences among these orthologs, we envision the exon/intron structure of the ancestral gene (Fig. 7A, middle). The first three exons of lin-12 and glp-1 are conserved, suggesting that they may reflect exons of the common ancestor. With the exception of the 5′ splice site for lin-12 intron 5 and the 5′ splice site of glp-1 intron 4, which are homologous, no other exon/intron boundaries have been unambiguously conserved between paralogs. It is likely that ancestral exon 4 and a portion of ancestral exon 5 were lost in glp-1: the EGF repeats encoded by lin-12 exon 4 do not retain any signature that might suggest they were generated by duplication. As previously described (Yochem and Greenwald 1989), this putative loss underlies the major structural difference between LIN-12 and GLP-1: three missing EGF repeats from the extracellular domain of GLP-1. Finally, we envision a large exon that has been broken up in distinct ways in lin-12 and glp-1 because the exon/intron boundaries of this region are distinct in the two orthologs.

When did the common ancestor duplicate? We speculate that the duplication happened well before divergence of the three species studied here, perhaps before the radiation of Caenorhabditis. This idea is based on two lines of evidence. First, all three lin-12 orthologs possess essentially the same gene-specific exon/intron structures (Fig. 1) as do all three glp-1 orthologs (Rudel and Kimble 2001). Thus, most architectural differences between lin-12 and glp-1 gene structures occurred before the radiation of these species. Second, phylogram branch lengths between the LIN-12 clade and the GLP-1 clade (dotted lines, Fig. 3A) are comparable with branch lengths between orthologs (dashed lines, Fig. 3A). Because these branch lengths are a coarse measure of evolutionary distance, the duplication is likely to have occurred well before the species diverged. Additionally, we cloned a nucleotide sequence corresponding to the ANK repeats of a putative Notch-like receptor from PS1010 (data not shown), a Caenorhabditis species distant to those analyzed here (Sudhaus and Kiontke 1996). The PS1010 ANK repeats are more similar to GLP-1 than LIN-12 (D. Rudel and J. Kimble, unpublished data), suggesting the existence of both genes in PS1010. Because RNAi was not successful in PS1010, identification of the biological functions for this putative GLP-1 ortholog could not be determined. We conclude that the split between lin-12 and glp-1 is likely to have occurred long before speciation of the strains used here.

Evolutionary constraints on motifs within nematode Notch-like receptors

In general, the extracellular domains of LIN-12 and GLP-1 are more conserved than their intracellular domains. The extracellular domain is responsible for interactions with ligands, with other receptors, and for receptor activation (Kidd et al. 1989; Greenwald and Seydoux 1990; Lyman and Young 1993; Blaumuehler et al. 1997; Logeat et al. 1998; Rand et al. 2000; for review see Artavanis-Tsakonas et al. 1995). The high conservation within the extracellular domain of LIN-12 and GLP-1 may represent stringent constraints placed on its motifs to prevent a leaky or intermediate response during signaling.

The RAM and ANK domains have the lowest conservation among the principal LIN-12/GLP-1 motifs. Yet these two regions are critical for forming the transcriptional ternary complex with LAG-1 and LAG-3 (Roehl et al. 1996; Schroeter et al. 1998; Doyle et al. 2000; Petcherski and Kimble 2000). It has been proposed that eukaryotic transcriptional regulatory machinery has a degree of built-in disorder (Kirschner and Gerhart 1998). Consequently, interactions are less precise so that the same core transcriptional machin-
ery can be used in differing contexts, that is, with additional players to increase specificity and transcribe various target genes when appropriate. This built-in promiscuity may translate into reduced constraints during evolution and therefore may lead to a relatively low conservation of ANK repeats and the RAM domain in the Notch-related receptors in two ways. Either interactions between the ternary complex and the core eukaryotic transcriptional machinery may have a relatively low affinity and specificity, or low conservation may allow the ternary complex to interact with tissue specific transcription cofactors to effect the pleiotropic results of Notch signaling.

Molding the individual LIN-12 and GLP-1 receptors

The LIN-12 and GLP-1 sequences have receptor-specific features. First, the EGF repeats and the LNR motifs of LIN-12 are more highly conserved than those of GLP-1 (74–78% vs. 56–65% and 85–86% vs. 67–74%, respectively). Perhaps this reflects higher specificity for individual ligands; the C. elegans genome encodes four DSL ligands, only two of which (LAG-2 and APX-1) are involved in known interactions (Lambie and Kimble 1991; Henderson et al. 1994; Mello et al. 1994; Tax et al. 1994; Wilkinson et al. 1994; The C. elegans Sequencing Consortium 1998). Alternatively, high conservation may reflect interactions within or between LIN-12 receptors. Second, conserved receptor-specific amino acids occur in the CC-linker, which has been implicated in interactions between the extracellular moiety and the transmembrane moiety of the receptor and perhaps in dimerization (Kidd et al. 1989; Greenwald and Seydoux 1990; Rand et al. 2000). Distinct CC-linkers may ensure proper segregation of receptor domains or may regulate receptor dimerization. Third, the TMD of GLP-1 is more highly conserved than that of LIN-12 (60–70% vs. 16–32%). This difference suggests that the GLP-1 TMD may have a function not shared with LIN-12. Because one major TMD function is its cleavage upon signaling, perhaps the GLP-1 TMD plays some receptor-specific role in that cleavage. Previous studies suggested that LIN-12 and GLP-1 proteins are functionally equivalent (Lambie and Kimble 1991; Mango et al. 1991; Fitzgerald et al. 1993). However, the receptor-specific differences in conservation observed by these two receptors suggest that they may not be the same. Small differences in LIN-12 and GLP-1 may have subtle roles in species fitness in the wild, and earlier investigations may not have tested such small differences. Additionally, it is not known whether LIN-12 can substitute for GLP-1 in LIN-12–specific functions. We suggest that lin-12 and glp-1 are not entirely equivalent.

Although we propose that the varying degree of conservation between motifs both among orthologs and among paralogs may reflect selective constraints (i.e., purifying or stabilizing selection), we cannot rule out two other possibilities. First, differences between receptors may reflect different neutral rates of evolution for the two receptor sequences. An argument against this is that the LIN-12 and GLP-1 molecules interact with many of the same molecular players in some of the same tissues. Consequently, the environments in which selection occurs should be similar for both receptors. Second, differences may be the result of positive selection. An analysis of non-synonymous changes per non-synonymous site versus synonymous changes per synonymous site for the nucleotide sequences corresponding to the alignable
LIN-12 and GLP-1 protein sequences gave K/K \(_\text{s}\) values < 1 for both orthologous pairs and paralogous pairs (data not shown). Values >> 1 are required to make confident statements that positive selection is occurring (Kreitman and Comer 1999). The low ratios obtained do not rule out the possibility that positive selection occurred at some time during Caenorhabditis Notch evolution. Likewise, insights into the importance of amino acid sequence differences among receptors in specific biological functions and in fine tuning of signaling and regulating cell response are not gained by this analysis. Rather, these ratios may indicate that purifying selection, or stabilizing selection, has occurred through a large portion of Caenorhabditis Notch evolution and may suggest that regulatory changes have played large roles in differences in function across phyla.

**Evolution of duplicated genes to adopt distinct functions**

A comparison of the similarities and differences of lin-12 and glp-1 functions among species can provide insight into how those genes may have evolved to achieve their current constellation of biological functions. At least one set of cell fate decisions governed by LIN-12 receptors has remained largely the same among the three species examined here: lin-12 in all three species regulates processes critical for larval viability, for example, rectum formation and excretory pore development (Lambie and Kimble 1991; this work). In C. elegans, GLP-1 has also retained these functions. By contrast, in C. briggsae and C. remanei this role has been delegated to lin-12 alone. Whereas Ce-lin-12(\text{RNAi}) led to postembryonic defects typical of a lin-12(0) mutant (Greenwald et al. 1983; Seydoux et al. 1990), Cb-lin-12(\text{RNAi}) and Cr-lin-12(\text{RNAi}) led to larval lethality with defects commonly seen in C. elegans when both lin-12 and glp-1 functions are removed (Lambie and Kimble 1991). The most probable explanation is that the relative contributions of LIN-12 and GLP-1 in the late embryo have diverged, that is, glp-1 activity is lower in C. briggsae and C. remanei embryos than in C. elegans embryos. We speculate that glp-1 may have either lost a positive regulatory element that enhanced its embryonic activity or acquired a negative regulatory element (Fig. 7B, c). The 5' flanking regions and introns of Ce-lin-12 and Cb-lin-12 contain a number of small (15–40 bp) sequences with high nucleotide identity (>90%) that are potential transcriptional regulatory elements (data not shown); comparisons with the complete genomic sequence for Cr-lin-12, which remains unfinished, would be useful to learn which of these sequences may be important. Ultimately, transgene experiments are required to test the roles of regulatory elements in controlling LIN-12 and GLP-1 to achieve individual developmental decisions.

Traditional models argue that the acquisition of new roles for duplicated genes prevents the accumulation of deleterious loss-of-function mutations in each gene and, thereby, leads to retention of both genes (Ohno 1970). A modified scheme suggests that the distribution of ancient roles between duplicated genes ensures their retention (Hughes 1994; Force et al. 1999). For example, loss of specific regulatory elements from one gene, but not both, might lead to one gene retaining an ancestral function and the other losing it. Similarly, regulatory elements can be acquired to permit new biological functions. After time, each gene is responsible for a discrete constellation of functions and both are retained (Force et al. 1999).

Our phylogenetic comparison of LIN-12 and GLP-1 functions suggests that the subdivision of ancestral roles played a role in retaining duplicated receptors and generating distinct roles. Two observations from earlier studies suggested that Ce-lin-12 and Ce-glp-1 have overlapping functions but are evolving toward more distinct roles. First, in contrast to C. briggsae and C. remanei, both Ce-lin-12 and Ce-glp-1 govern mid to late embryonic functions to make the excretory pore and rectum. Second, control of vulval fates has been largely allocated to lin-12 in C. elegans (Greenwald et al. 1983); lin-12, but not glp-1, null mutants are vulva defective. However, glp-1 can affect vulval cell fates when the receptor has lost its PEST domain (Mango et al. 1991). In this situation, GLP-1 is thought to have acquired a negative regulation via the PEST domain that inhibits its vulval function. We previously reported novel roles for GLP-1 in C. briggsae (Rudel and Kimble 2001) (Fig. 7B, d), but we do not yet know if these are new functions in C. briggsae or masked functions yet to be revealed for GLP-1 and/or LIN-12 in C. elegans and C. remanei. We suggest that the lin-12 and glp-1 genes in C. elegans, C. briggsae, and C. remanei represent an intermediate stage in their evolution towards disparate genes with discrete functions.

**lin-12 and glp-1: a paradigm for analyzing gene duplication of developmentally critical genes**

The duplication of genes and their subsequent evolution is a major mechanism by which genomes and species evolve. Among the many examples of genes that have duplicated throughout evolution, lin-12 and glp-1 stand out as a particularly well-characterized example of this basic evolutionary event. The intensive genetic and molecular analyses of lin-12 and glp-1 already completed in C. elegans (for review see Kimble and Simpson 1997) provide an essential foundation for analyses of these genes in other nematodes. Our phylogenetic analyses of glp-1 (Rudel and Kimble 2001) and lin-12 (this work) from closely related species strongly suggest that the lin-12/glp-1 gene duplication was sufficiently recent to provide insight into the ancestral Notch-like gene and sufficiently distant to explore the evolution of the duplicated gene pair. This gene pair therefore should be a superb model for in-depth analyses of a gene duplication and evolution of its products and biological roles. Challenges for the future include identification...
of Notch-related receptors from more distant nematodes, at least in part to identify a point before the gene duplication, and the characterization of receptor mutants in these other nematodes to define their functions more rigorously.

Acknowledgments

We thank Mary Montgomery and Iva Greenwald for sharing unpublished data. We thank Christian Eckmann for reagents and thoughtful comments on this work. We also thank Betsy Goodwin, Sarah Crittenden, Eric Haag, and many lab members and colleagues who provided support and criticism during the course of this work. We also thank the reviewers for their efforts to make this a better manuscript. D. R. was a National Institute of Health Molecular Sciences trainee; J. K. is an investigator with the Howard Hughes Medical Institute.

REFERENCES


