Paralogy and Orthology of Tyrosine Kinases that Can Extend the Life Span of Caenorhabditis elegans

Brad A. Rikke, Shin Murakami, and Thomas E. Johnson
Institute for Behavioral Genetics, University of Colorado

Modification of any one of three transmembrane protein tyrosine kinase (PTK) genes, old-1, old-2 (formerly tkr-1 and tkr-2, respectively), and daf-2 can extend the mean and maximum life span of the nematode Caenorhabditis elegans. To identify paralogs and orthologs, we delineated relationships between these three PTKs and all known transmembrane PTKs and all known mammalian nontransmembrane PTKs using molecular phylogenetics. The tree includes a number of invertebrate receptor PTKs and a novel mammalian receptor PTK (inferred from the expressed-sequence tag database) that have not previously been analyzed. old-1 and old-2 were found to be members of a surprisingly large C. elegans PTK family having 16 members. Interestingly, only four members of this transmembrane family appeared to have receptor domains (immunoglobulin-like in each case). The C-terminal domain of this family was found to have a unique sequence motif that could be important for downstream signaling. Among mammalian PTKs, the old-1/old-2 family appeared to be most closely related to the Pdgfr, Fgfr, Ret, and Tie/Tek families. However, these families appeared to have split too early from the old-1/old-2 family to be orthologs, suggesting that a mammalian ortholog could yet be discovered. An extensive search of the expressed-sequence tag database suggested no additional candidate orthologs. In contrast to old-1 and old-2, daf-2 had no C. elegans paralogs. Although daf-2 was most closely related to the mammalian insulin receptor family, a hydra insulin receptor–like sequence suggested that daf-2 might not be an ortholog of the insulin receptor family. Among PTKs, the old-1/old-2 family and daf-2 were not particularly closely related, raising the possibility that other PTK families might extend life span. On a more general note, our survey of the expressed-sequence tag database suggested that few, if any, additional mammalian PTK families are likely to be discovered. The one novel family that was discovered could represent a novel oncogene family, given the prevalence of oncogenes among PTKs. Finally, the PTK tree was consistent with nematodes and fruit flies being as divergent as nematodes and mammals, suggesting that life extension mechanisms shared by nematodes and fruit flies would be reasonable candidates for extending mammalian life spans.

Introduction

Protein tyrosine kinases (PTKs) orchestrate many of the cell-cell interactions necessary for growth, development, and repair (reviewed by van der Geer, Hunter, and Lindberg 1994). PTKs may even have played a crucial role in establishing the first metazoans, as PTKs are well represented in sponges and hydra (Ottlie et al. 1992; Schacke et al. 1994; Cetkovic et al. 1998; Muller 1998), but no conventional PTKs are found in yeast or plants (Walker 1994; Hunter and Plowman 1997; Suga et al. 1997; Chervitz et al. 1998; Satterlee and Sussman 1998). In mammals, multicellular organization is still highly dependent on the proper functioning of PTKs, as many PTKs become oncogenic when their activity is altered. Because many PTKs are proto-oncogenes, their identification and characterization has been intensely pursued (Wilks 1991), with approximately 100 PTKs having now been identified in humans (Hunter 1998).

Variants of three transmembrane PTK genes have been found to extend the life span of Caenorhabditis elegans. daf-2 partial-loss-of-function mutations of the receptor and catalytic (kinase) domains extend mean and maximum life spans by about 100% (Kenyon et al. 1993; Kimura et al. 1997). daf-2 is most similar in sequence to the mammalian genes for the insulin receptor family (Kimura et al. 1997). old-1 and old-2 (overexpression longevity determinant), formerly tkr-1 and tkr-2, extend mean and maximum life spans by about 65% and 20%, respectively, when their wild-type alleles are overexpressed (Murakami and Johnson 1998). old-1 and old-2 are about 90% identical and most similar in sequence to mammalian genes for mast/stem cell growth factor receptors (Kit) and fibroblast growth factor receptors (Fgfrs) (Murakami and Johnson 1998).

Although old-1 and daf-2 do not share sequence similarity outside their catalytic domains, they appear to extend life span by a shared signaling pathway. Life extension by both old-1 and daf-2 is dependent on daf-16 (Kenyon et al. 1993; unpublished data), encoding a transcription factor homologous to the forkhead family of fruit flies and mammals (Lin et al. 1997; Ogg et al. 1997). In addition, old-1, daf-2, and most other variants that extend C. elegans and Drosophila life spans confer resistance to multiple kinds of stress, such as oxidation, heat, and ultraviolet radiation (reviewed by Lithgow 1996; Murakami and Johnson 1996, 1998; Gems et al. 1998; Lin, Seroude, and Benzer 1998). This mechanistic conservation raises the possibility that additional PTKs, particularly paralogs and orthologs, might be capable of extending life span.

This study specifically addresses a number of questions regarding the molecular evolution of old-1, old-2, and daf-2: (1) Are there paralogs of old-1, old-2, and daf-2 and if so, what are they? (2) Are the mammalian genes that are most similar in sequence to old-1, old-2, and daf-2 likely to be orthologs? (3) Are there additional candidate orthologs? (4) How closely related are old-1,
old-2, and daf-2? The term “paralogs” in this study refers to other members of a gene family within a species (arising by gene duplication). “Orthologs” refers to homologous genes created by speciation; thus, their divergence should be congruent with that of the species being compared.

Because the number of PTK families is quite large and many invertebrate genes and mammalian expressed-sequence tag (EST) sequences have only recently been added to databases, this study secondarily touches on several more general questions: (1) How congruent is the PTK tree with previous PTK trees? (2) Is it likely that many more mammalian PTK families will be discovered? (3) Is the divergence between nematodes and mammals greater than the divergence between nematodes and fruit flies, as previously concluded?

Materials and Methods

Databases searched to identify PTK families are shown in table 1. BLAST (Altschul et al. 1990) searches were conducted using the inferred amino acid sequences of old-1, old-2, the old-1 kinase domain, and an old-1/old-2 family consensus sequence. Final searches were current as of November 1998. Representatives from mammalian PTK families with no apparent Caenorhabditis elegans and Drosophila orthologs were used to BLAST search the European Molecular Biology Laboratory (EMBL) invertebrate database, C. elegans databases, and the Berkeley Drosophila Genome Project (BDGP) to help insure that potential orthologs were not missed. Likewise, invertebrate PTKs with no apparent mammalian orthologs were used to BLAST search the SwissProt-translated EMBL database (SPTR).

The amino acids of unfinished sequences (DS02309, DS03465, C30F8, Y50D4), Caenorhabditis briggsae sequences (G40L08a, G40L08b, G40L08c), and ESTs (R61934, A1386314) were inferred using BLAST alignments and GCG FRAMEALIGN. The inferred amino acid sequences of several old-1/old-2 family members were also modified based on BLAST alignments and analysis of intron/exon junctions. Additional alterations not shown in the old-1/old-2 family sequence figure are as follows: (1) VRFFWGBKSGANFQFLKKS-EFPK was deleted from M01B2.1, (2) W04G5.6a and W04G5.6b were derived from W04G5.6, and (3) T17A3.1/F40G9.13 was constructed from the end of cosmid T17A3 and the beginning of cosmid F40G9 with the overlapping sequence GDRQRKSSS removed. Outside the old-1/old-2 family, the only change made was that LVKDHANQLEFNDVVEIKLMKGIGYHKNIVX was added to the beginning of Q25196 (hydra).

Amino acid sequences were multiply aligned using the default settings of CLUSTAL W (Thompson, Higgins, and Gibson 1994). No improvement was achieved using secondary structure information (obtained from SwissProt). The only regions conserved among all PTK sequences were the kinase 1 and 2 domains, as defined by Hubbard et al. (1994). The alignment began with the ATP-binding site of the kinase 1 domain (LGXGXGF motif) and ended at the C-terminal domain. The kinase insert domain was removed. The final alignment covered a region of about 250 amino acids (excluding gaps).

Some alignment gaps were adjusted manually. Private gaps of more than three amino acids were deleted. Privacy was confirmed by BLAST searches of the SPTR database. For regions in which closely spaced gaps were separated by a short region of poor alignment, the entire region was deleted and ambiguous characters were added back (Q15516_human: deleted TAWNCT ... IPKTG and added 21 ambiguous characters; Q21477_m03a1: deleted NTTMGRGGGGG; Q24315_drome: deleted DAERM ... GKRHA and added 18 ambiguous characters; Q25177_hydat: deleted LALCP ... DHDTF and added 10 ambiguous characters; Q26566_fluke: deleted YNHWRGAQ and ILND ... SYSE).

The final PTK tree was constructed using the pairwise distance matrix program Fitch (PHYLIP package; Felsenstein 1993). The invertebrate nonreceptor PTKs P18106 (FPS), P08630 (SRC2), P00528 (SRC1), P03949 (ABL1), C30F8 from WUGSC, Q24592 (HOP), Q24316 (PR2), and P53356 (TK16) were included in the tree construction to help subdivide long branches;

<table>
<thead>
<tr>
<th>Database</th>
<th>Web Site/Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Berkeley Drosophila Genome Project (BDGP)</td>
<td><a href="http://www.fruitfly.org/">www.fruitfly.org/</a></td>
</tr>
<tr>
<td>2. BLOCKS</td>
<td><a href="http://www.hgmp.mrc.ac.uk/">www.hgmp.mrc.ac.uk/</a>; Henikoff and Henikoff (1994)</td>
</tr>
<tr>
<td>3. EMBL invertebrate subsection</td>
<td><a href="http://www.hgmp.mrc.ac.uk/">www.hgmp.mrc.ac.uk/</a></td>
</tr>
<tr>
<td>8. PRODOM</td>
<td><a href="http://www.hgmp.mrc.ac.uk/">www.hgmp.mrc.ac.uk/</a></td>
</tr>
<tr>
<td>10. Protein Kinase Resource</td>
<td><a href="http://www.sdsc.edu/Kinases/pk_home.html">www.sdsc.edu/Kinases/pk_home.html</a>; Smith et al. (1997)</td>
</tr>
<tr>
<td>11. Sanger Centre C. elegans database</td>
<td><a href="http://www.sanger.ac.uk/Projects/Celegans">www.sanger.ac.uk/Projects/Celegans</a></td>
</tr>
<tr>
<td>13. Washington University Genome Sequence Center</td>
<td>genome.wustl.edu/gsc/gschmpg.html</td>
</tr>
</tbody>
</table>

*Each address preceded by http://.
however, they were excluded from the final figure to conserve space. The algorithm assumes an additive-tree model, uses the weighted least-squares criteria of Fitch and Margoliash (1967), and does not assume a molecular clock. The input order of sequences was jumbled 10 times. The distance matrix was constructed using PROTDIST (PHYLIP) and the Dayhoff PAM substitution matrix. No negative branch lengths were obtained. The tree was bootstrapped 100 times (input order jumbled once each time) using SEQBOOT (PHYLIP), PROTDIST, FITCH, and CONSENSE (PHYLIP). The C. elegans sequences T22B11.3, F59F5.3, F54F7.5, B0198.3, and Y38H6C.20 were not included in this bootstrapping because these sequences were added later (see Acknowledgments). These sequences were bootstrapped in the absence of the nonreceptor PTK sequences to reduce computing time, and the effect on the previous bootstrap results was negligible. The tree was rooted using two human mixed-lineage kinases (SPTR accessions P80192 and Q16584), two human LIM domain kinases (P53667 and P53671), and two RIP serine/threonine kinases (human Q13546 and murine Q60855). Trees were displayed and the final tree figure was drawn using the help of TREEVIEW (Page 1996).

Deletion analysis of the PTK tree to investigate bootstrap support for grouping the old-1/old-2 family with mammalian Pdgfr, Fgfr, and Ret families was carried out by excluding the root sequences, C08H9.8, T01G5.1, T22B11.3, F59F5.3, F54F7.5, B0198.3, Y38H6C.20, Tie1, Tie2, R61934 (sea urchin EST), O76167 (planarian), P11475 (Drosophila torso), O76809 (hydra), Q25196m (hydra), Q25199 (hydra), A1386314 (murine EST), B0252.1, C24G6.2A, F09A5.2, F08F1.1, F09G2.1, Q24315 (Drosophila), T14E8.1, C25F6.4, C16D9.2, Q24592 (Drosophila jak), and ZK1067.1 (let-23). The tree was bootstrapped 100 times.

A maximum-likelihood tree of the old-1/old-2 family was constructed using PUZZLE 4.0 (Strimmer, Goldman, and von Haeseler 1997), with eight classes of substitution rate heterogeneity and the BLOSUM 62 substitution matrix (Henikoff and Henikoff 1992). The sequence alignment began with the first residue that is reverse highlighted (P) in the sequence figure and ended with the last residue that is reverse highlighted (W). The tree thus included the two kinase domains, the kinase insert domain, and the C-terminal domain. The tree was rooted using the Fgfr family. Because PUZZLE uses a heuristic approach (quartet puzzling and neighbor joining) to identify the best tree, an exhaustive search was also conducted using the maximum-likelihood program PROTML (Adachi and Hasegawa 1996). There was no appreciable difference between the best trees found using PUZZLE and those found using PROTML. There was also no appreciable difference between trees constructed using maximum-likelihood and those constructed using GCG PAUP (version 4.0.4d55 for UNIX; Swofford 1993) and PROTPARS (PHYLIP).

The genes flanking old-1/old-2 family members were identified from the GeneFinder sequence annotations of the Sanger Centre and WUGSC databases. A BLAST search of the SPTR and GenBank databases was also conducted using the first 5 kb of sequence upstream and downstream of each old-1/old-2 family member.

Nonkinase intra- and extracellular PTK domains were identified using SwissProt and Pfam (www.sanger.ac.uk/Software/Pfam/search.shtml). Some domains were inferred based on sequence similarity to a known domain of a related sequence (e.g., the IG domain of O76167 was based on similarity to P18460).

We searched for sequence motifs using MetaMeme, version 2.0.1 (Grundy et al. 1997; http://metameme.sdsc.edu/mhmmlinks.html). Meta-Meme was also used to conduct homology searches of the National Center for Biotechnology Information’s nonredundant peptide sequence database using the old-1/old-2 family C-terminal domain motif.


The EMBL/GenBank EST database was searched using the amino acid sequences of old-1, the last 25 residues of old-1, an old-1/old-2 family consensus, the last 31 residues of the old-1/old-2 family kinase domain, and an old-1/old-2 family consensus of the C-terminal domain. When last searched, the EST database contained 2.1 million entries (January 15, 1999). The only known mammalian PTK families not detected were Ros/Sevenless and AATK, whose ESTs did not overlap significantly with the PTK catalytic domain. Hits from the full-length consensus sequence query (all kinase domain hits) having expected probabilities of <5.0 (approximately 1,000 ESTs) were used as BLAST queries of the EMBL nucleotide database. ESTs that were less than 85% identical to a known sequence or whose hits covered less than 190 nt were analyzed further by using all six reading frames to BLAST query the SPTR or Entrez protein databases. ESTs whose best protein hits were PTKs but whose best sequence identities with a PTK kinase domain was less than 80% were included in the tree as possible representatives of novel PTK families. (The 80% cutoff was a conservative estimate designed to include false positives, given that the identity between all known transmembrane PTK families is <70%.) Only three ESTs (one murine, two invertebrate) appeared to represent potentially novel families. The murine EST was compared with the EST consensus sequences compiled by the Institute for Genome Research (TIGR, http://www.tigr.org/tigr_home/tdb/), but no contiguous sequence data were available.

The GenBank EST database was also searched using GCG FRAMESEARCH and GCG PROFILESEARCH (weighted matrix) with the old-1/old-2 family C-terminal domains as queries. The significance of the two best hits was determined by comparison against a C-terminal domain consensus sequence using GCG FRAMEALIGN. Hits with Z scores of >5 were also examined using Entrez and BLAST.

Access to most of the databases, programs, and program packages (BLAST, BLOCKS, CLUSTAL W, GCG, MPSEARCH, PHYLIP, ProDom, PROTML, and PUZZLE) was obtained using the U.K. Human Genome
Results
Paralog and Ortholog Searches Using BLAST Queries

A BLAST search of the SwissProt database using full-length old-1 and old-2 amino acid sequences resulted in two outstanding hits, kin-15 and kin-16, both from C. elegans. BLAST searches of the TrEMBL database and Sanger Centre/WUGSC C. elegans database, covering virtually the entire C. elegans genome (C. elegans Sequencing Consortium 1998), resulted in eight additional outstanding hits from cosmids R09D1, F59F3, W04G5, M01B2, T17A3/f40G9, and yeast artificial chromosome Y50D4 (expected probabilities \(<10^{-50}\) vs. \(>10^{-30}\) for all other hits). To help ensure that potential paralogs were not overlooked, the top 60 hits from the C. elegans database were used as BLAST queries of the SPTR and GenBank databases. Those queries whose best hits were transmembrane PTKs were then included in the molecular phylogenetic tree.

The top seven old-1 nonnematode hits from an un-gapped BLAST search of the SPTR database were all Kits (human, horse, pig, bovine, guinea pig, cat, and rat), with expected probabilities \(<10^{-37}\). However, a gapped BLAST search (Altschul et al. 1997) resulted in 22 top nonnematode hits that were all Fgfrs (human, mouse, rat, chicken, frog, newt, and fruit fly). The Kit and Fgfr hits were then used to query the EMBL invertebrate database. Similar to previous strategies for finding orthologs (Tatusov, Koonin, and Lipman 1997; Chervitz et al. 1998), the expectation was that if Kits and/or Fgfrs were orthologs of old-1 and old-2, they would reciprocate by finding old-1 and old-2 as their best C. elegans hits. However, the best C. elegans hit for each Kit and Fgfr was egl-15, not old-1 or old-2 (or any other member of the old-1old-2 family). Therefore, no PTKs stood out as probable orthologs of old-1 and old-2.

A more extensive survey (see Materials and Methods) was then conducted to (1) identify all of the potential orthologs of old-1 and old-2, (2) identify all of the potential orthologs of daf-2, (3) identify all of the PTK families that may have evolved from the last common ancestor of old-1, old-2, and daf-2, (4) identify as many PTK families as possible having putative mammalian and nematode orthologs, and (5) include as many PTK families as necessary to subdivide long branches of the PTK tree. The end result was that approximately 250 different (i.e., \(>5\%\) divergent) vertebrate and invertebrate PTKs were identified, representing all known PTK families.

In aggregate, approximately 200 different PTK catalytic domains (noncatalytic domains were not widely conserved and thus were excluded) were examined by molecular phylogenetic analysis. Most trees were constructed using the Fitch-Margoliash distance matrix program FITCH, which does not assume a molecular clock. Parsimony methods (PAUP and PHYLIP PROTPARS) were also used to examine approximately 100 PTKs from families shown in figure 1 as being most closely related to old-1, old-2, and daf-2. The best parsimony trees (not shown) were not appreciably different from those produced using FITCH, but parsimony trees required much more computing time. These exploratory analyses suggested that old-1, old-2, and daf-2 were more closely related to transmembrane PTKs than to nontransmembrane PTKs, and that nontransmembrane PTK families tended to cluster together. Therefore, most of the invertebrate nontransmembrane PTKs were excluded to reduce computing time (some were retained to subdivide long branches). In addition, most vertebrate PTKs whose catalytic domains were more than 70% identical to that of a human PTK were excluded because these did not contribute significantly to subdividing long branches or distinguishing families. The final tree of 144 PTKs still contained representatives from all known vertebrate nontransmembrane and transmembrane PTK families (fig. 1).

An old-1old-2 Family

As shown in figure 1, old-1 and old-2 were found to be members of a large gene family in C. elegans which will be referred to as the old-1old-2 family and coincides with the Kin15/Kin16 family (Hanks and Hunter 1995). This family included 14 other C. elegans genes, 4 more than suggested by BLAST searches and 5 more than previously suggested for the Kin15/Kin16 family (Ruvkun and Hobert 1998). Ignoring F54F7.5 and B0198.3 (which occasionally clustered within the old-1old-2 family), bootstrap support clustering the 16 genes together was 100%. Only two family members, T01G5.1 and F59F3.1, were represented in the C. elegans EST database. Three genes from the nematode C. briggsae also clustered with the old-1old-2 family and will be discussed below.

Within the old-1old-2 family, only kin-15 and kin-16 have been functionally characterized. Expression patterns suggested that both genes are involved in post-embryonic development of the hypodermis (Morgan and Greenwald 1993). In spite of their apparent functional similarity, however, a kin-15 + kin-16 clade was not strongly supported. Interestingly, kin-16 had a greater tendency (51 bootstrap trees vs. 7) to group with clade T17A3.1/f40G9.13 + T17A3.8 + F59F3.1 + F59F3.5 than with kin-15.

It was also interesting that T17A3.1/f40G9.13, T17A3.8, F59F3.1, and F59F3.5 were the only old-1old-2 family members to have recognizable receptor domain sequences. These domains were similar to the immunoglobulin-type domains (IG domains) that are found in many different proteins, including many receptor PTKs (fig. 1). These IG domains thus suggested a receptor function for this branch of the old-1old-2 family rather than the adapter function proposed by Morgan and Greenwald (1993) for kin-15 and kin-16. F59F3.1 and F59F3.5 appeared to have seven IG-like domains, which among PTKs is the same number found only in Vegfrs.
The tree also suggested several vertebrate gene families as being most closely related to the old-1/old-2 family. One of these families includes the macrophage colony stimulating factor receptor (Fms), Kit, FMS-like tyrosine kinase 3 (Flt3), platelet-derived growth factor receptors (Pdgfrs), and Vegfrs and will be referred to hereafter as the Pdgfr family according to the classification of Hanks and Hunter (1995). The other gene families were Fgfr, the ret proto-oncogene (Ret), and the “tyrosine kinase with IG and EGF homology/tunica interna endothelial cell tyrosine kinase” receptors (Tie/Tek). The bootstrap support grouping the old-1/old-2 family with these mammalian families was low, largely due to variability in the positioning of other invertebrate sequences and due to 28% of bootstrap trees placing the old-1/old-2 family as an outgroup of all the other PTK families. Ignoring these, 48 of 72 trees (67%) grouped the old-1/old-2 family with the Pdgfr, Fgfr, Ret, and Tie/Tek families; 12 trees grouped the old-1/old-2 family with the Egfr family, and only 1–3 trees supported other groupings. Similarly, deleting some of the more divergent sequences from the tree (see Materials and Methods) increased the bootstrap support grouping the old-1/old-2 family with the Pdgfr, Fgfr, and Ret families to 69%. Although the bootstrap support was still not high, the Pdgfr, Fgfr, Ret, and Tie/Tek families stood out as the best candidates for orthology with the old-1/old-2 family. As described below, shared intron sites and shared C-terminal domain sequences provided additional support. The second best candidate for orthology with the old-1/old-2 family appeared to be the Egfr family.

Like the old-1/old-2 family, the Pdgfr, Fgfr, Ret, and Tie/Tek families encode transmembrane PTKs. With the exception of Ret, the receptor domains of these families also contain IG domains. However, the PTK tree of figure 1 suggested that each of these mammalian families was more closely related to egl-15 than to the old-1/old-2 family, which was consistent with the BLAST results. The relatedness between these mammalian PTK families and egl-15 also agrees with previous molecular phylogenetic analyses (Rousset et al. 1995; Coulier et al. 1997).

The PTK tree was then used to investigate whether the Pdgfr, Fgfr, Ret, and Tie/Tek families might be considered probable orthologs of the old-1/old-2 family by comparing their split with the range of splits observed between other vertebrate PTKs and their probable orthologs. As shown in figure 1 (yellow column), the range of orthologous C. elegans–mammal splits also corresponded closely with other orthologous invertebrate-mammal splits. However, the split between the old-1/old-2 family and the Pdgfr, Fgfr, Ret, and Tie/Tek families preceded the probable range of orthologous invertebrate-mammal splits (likely to include some non-orthologous splits as well). Thus, the Pdgfr, Fgfr, Ret, and Tie/Tek families appear to have diverged too early from the old-1/old-2 family to be orthologs. Grouping the old-1/old-2 family with the Egfr family also resulted in a split that preceded the range of probable orthologous invertebrate-mammal splits.

It was curious that the old-1/old-2 family did not group more closely with the Pdgfr family given sequence similarities in both the kinase and the IG domains and given the absence of a Pdgfr ortholog in C. elegans. Because most members of the old-1/old-2 family did not have an IG domain and appeared to have evolved faster than the IG branch, we asked whether the IG branch might be more representative of the old-1/old-2 family progenitor. However, when the non-IG branch (with or without kin-16) was excluded from the tree, the genetic distance between the IG branch and the Pdgfr family increased rather than decreased, and the IG branch became an outgroup of all of the other PTK families. (When the IG branch, the C. briggsae branches, and the four longest C. elegans branches were excluded, the non-IG branch gave the same topology as the old-1/old-2 family as a whole.) Therefore, the kinase domain sequences of just the IG branch of the old-1/old-2 family also did not suggest an orthologous relationship with the Pdgfr family (or any other family).

We also noticed that in BLAST searches using an old-1/old-2 IG branch receptor domain consensus sequence, the only significant PTK hits (several non-PTKs gave similar alignment scores) in the SwissProt database were from the Pdgfr family (probability by chance <10^-4 vs. <2 for Fgfrs). However, the alignment scores were quite low (<51 bits), and significance was due to the detection of extended regions of low identity (about 16%) with Vegfrs. In the region of about 180 residues where the F59F3.1 receptor domain overlapped both the human Vegfr1 and the human Fgfr4, the number of identities with Vegfr1 and Fgfr4 was actually the same (50). The IG-like receptor domains of the old-1/old-2 family and Pdgfrs also did not reciprocate as each other’s best BLAST hits in species comparisons. In contrast, in the region of about 200 residues in which the egl-15 receptor domain overlapped both the rat Vegfr1 (the best Pdgfr hit) and the human Fgfr4, the number of identities with Vegfr1 and Fgfr4 was actually the same (50). The IG-like receptor domains of the old-1/old-2 family and Pdgfrs also did not reciprocate as each other’s best BLAST hits in species comparisons. Hence, the gene pairs old-1 and old-2, W04G5.6a and W04G5.6b, and F59F3.5 and F59F3.1 arose as a result of cis duplications.

old-1/old-2 Family Genomic Organization

We also examined the genomic organization of the old-1/old-2 family and found that most members were arranged as tandem gene pairs, and many members were flanked by chitinase genes (fig. 2A). This organization suggested that some old-1/old-2 subfamilies might have arisen as trans duplications, perhaps similar to those of the Pdgfr family (Rousset et al. 1995). None of the subfamily arrangements, however, appeared to correspond to the subfamily arrangements of the Pdgfr family. Furthermore, Pdgfr family genes are not known to be flanked by chitinase-like genes, and the known mammalian chitinase-like genes (Jin et al. 1998) do not colocalize with any known PTK. The PTK tree clearly indicated that the gene pairs old-1 and old-2, W04G5.6a and W04G5.6b, and F59F3.5 and F59F3.1 arose as a result of cis duplications.
Given that old-1 and old-2 were two of the old-1 old-2 family members flanked by chitinase genes, we also reasoned that the paralogs most closely related to old-1old-2 might be flanked by chitinase genes. It was found that three of the five other genes flanked by chitinase genes (R09D1.12, M01B2.1, and C08H9.8) were also suggested by the PTK tree to be most closely related to old-1 and old-2.

old-1old-2 Family Intron Sites

To further elucidate relationships within the old-1 old-2 family, we examined intron/exon junctions (fig. 3). Preceding the kinase domain, the intron/exon junctions were not well conserved. Nevertheless, the N-terminal domains of old-1 and old-2 shared an intron site with R09D1.12 and kin-16. Given the close relationship suggested by the tree between R09D1.12 and M01B2.1, it was curious that M01B2.1 did not share this site. Moreover, it was curious that the 5’ end of M01B2.1 seemed to terminate prematurely. Therefore, we searched and found that M01B2.1 could indeed have an intron at exactly the same shared site; however, an initiating methionine codon for the putative upstream exon was not found.

In the juxtamembrane domain, kin-16 shared an intron site with the IG branch genes, consistent with kin-16 being more closely related to the IG branch than to kin-15 as suggested by the tree. Also in the juxtamembrane domain, old-1 and old-2 again shared an intron site with R09D1.12 and M01B2.1.

Within the kinase and C-terminal domains, intron sites were highly conserved. All members of the C. elegans old-1old-2 family shared six of six intron sites, often with the intron/exon junctions occurring within or between the same codons. Finding that these intron sites were highly conserved, we predicted and found six additional exons (one partial) that were not identified in the C. elegans database (fig. 3, residues in lowercase letters). These included the C-terminal exons of R09D1.12, C08H9.8, W04G5.6a, and W04G5.6b. We also found that the old-1old-2 family shared four intron sites with Pdgfr subfamilies, three with human Fgfrs, one with egl-15, and one with the C. elegans Ys3j family.

old-1old-2 Family Sequence Features Outside the Kinase Domain

Shared sequence features among old-1old-2 family members in regions outside the kinase domain were also examined. The N-terminal domain (the receptor domain of transmembrane PTKs) has previously been noted to be unusually short for kin-15 and kin-16 and thus more likely to function as an adapter rather than as a receptor (Morgan and Greenwald 1993). As shown in figure 3, such short domains (20–80 residues) appeared to be characteristic of the non-IG members of the old-1old-2 family. No other transmembrane PTK family, including Ryk (Hovens et al. 1992), had N-terminal domains that were nearly as short. Although short, the domain lengths appeared to be quite variable. No similarities to known domains or sequence motifs were found using Meta-Meme, BLOCKs, Pfam, or BLAST. However, a sequence motif (R-[VI]x-R-[AS]x-E-x-D-D) was found among the non-IG family members that appeared to be homologous to a motif within the IG domains of all four IG members (fig. 3). Among non-IG members (excluding kin-16), this sequence motif was shared by old-1, old-2, R09D1.12, and M01B2.1, again supporting R09D1.12 and M01B2.1 as being most closely related to old-1 and old-2.

The N-terminal domains of the IG members of the old-1old-2 family tended to be in the size range typically associated with PTK receptor domains. T17A3.8, however, had only one IG domain and thus appeared to have a short N-terminal domain as well.

The transmembrane domain of 15–20 hydrophobic residues was well conserved among the old-1old-2 family members. Although this domain appeared to be missing in W04G5.6a, Y50D4, and T01G5.1, their 5’ ends were not well defined. The juxtamembrane domain sequence was not well conserved except for its C-terminal end, which might actually be part of the kinase 1 domain. Similarly, the kinase insert domain was only well conserved at its N-terminal end, which could be part of the kinase 1 domain.

Interestingly, almost all of the C-terminal domain sequence was conserved among the old-1old-2 family members. The first half of this domain was similar to Pdgfrs and Fgfrs. The last half of this domain, however, was unique, with a conserved 16-residue sequence motif of [KR]-[LI]-x-E-x-[NE]-[EN]-Q-x-x-L-x-[ND]-W-I (fig. 3). Because the C-terminal domain of PTKs is typically involved in downstream signaling, this motif may have important functional implications unique to the old-1old-2 family as well. A Meta-Meme search suggested a potential hit (11:1 odds) for this motif within the C-terminal domain of insect mitochondrial cyto-
Fig. 2.—old-1/old-2 family genomic organization. Arrows point in the direction of transcription. Large arrows indicate old-1/old-2 family members. Small arrows indicate the genes that most closely flank the old-1/old-2 family (inferred by GeneFinder and/or BLAST). chi = chitinase gene.

EST Search

Because none of the known mammalian PTK families appeared to be probable orthologs of the old-1/old-2 family, additional candidate orthologs were searched for in the EMBL/GenBank EST database (see Methods and Materials). Approximately 360 mammalian PTK BLAST hits were obtained (almost entirely from mice and humans), representing all but two of the known PTK families. Only one EST appeared to represent a novel mammalian PTK family. This murine EST (accession number AI386314) covered 139 residues of the PTK kinase domain and was 100% identical to another, slightly smaller, EST (accession number AA0980224). On the tree, this EST grouped most closely with the mammalian Pdgfr, Fgfr, Ret, and Tie/Tek families (fig. 1). The most closely related C. elegans genes were B0252.1 and the Y3j family. Although B0252.1 had a long branch length, the EST still grouped best with the C. elegans Y3j family when B0252.1 was removed. No bootstrap trees grouped the EST with the old-1/old-2 family, making it seem unlikely that this EST is an old-1/old-2 ortholog. As a representative of a novel PTK family, this EST could be useful for discovering additional proto-oncogenes.

There were also two invertebrate ESTs that appeared to represent novel PTKs. One, from the sea urchin (accession number R61934), belonged to the Tie/Tek family (fig. 1). The other, from the nematode Onchocerca (accession number AI087777), appeared to be a nontransmembrane PTK and thus was not included in the tree.

daf-2
daf-2 did not appear to have any C. elegans paralogs (fig. 1). With regard to orthology, daf-2 received high bootstrap support for being most closely related to the mammalian Insr family, which is consistent with previous sequence comparisons (Kimura et al. 1997). Based on the range of probable nematode-mammal splits, the divergence between daf-2 and the Insr family was also congruent with orthology. It was curious, however, that a hydra gene (TK7) was more closely related to the Insr family than daf-2 was with a very high bootstrap score of 99. This hydra gene thus suggests that daf-2 and the mammalian Insr family might not be orthologous.

Among PTKs, it can be seen that daf-2 and the old-1/old-2 family were not particularly closely related (fig. 1). In fact, no PTKs could be excluded as having evolved from the common ancestor of the daf-2 and the old-1/old-2 family.

Discussion

This study is distinct from previous large-scale PTK studies in that it includes a number of invertebrate
receptor PTKs and a novel mammalian PTK (EST AI386314) that have not previously been analyzed. It also represents a complete survey of the EMBL/GenBank EST database, containing over 2 million entries. The results are largely congruent with previous trees (Hanks and Hunter 1995; Suga et al. 1997), although one difference is noted.

In this study, the Kin15/Kin16 family was found to be more closely related to the Pdgfr, Fgfr, Ret, and Tie/Tek families than to the Hgfr family as previously suggested by Hanks and Hunter (1995). This difference is almost certainly due to our including additional Kin15/Kin16 family genes and the *C. elegans* F11E6.8 gene. The F11E6.8 gene appears to be the most probable ortholog of the Hgfr family and was not available at the time Hanks and Hunter (1995) constructed their tree.

**Nematode–Fruit Fly–Mammal Divergence**

Our tree also differs from the view that nematodes are more closely related to fruit flies than they are to mammals. Although it has been recognized for some time that most protein-based trees place the *C. elegans* split before the *Drosophila*-mammal split, this phylogeny has recently lost favor, based largely on analyses of slow-evolving nematode 18S ribosomal RNA sequences (reviewed by Adoutte et al. 1999). Mushegian et al. (1998) also argued that slow-evolving *C. elegans* proteins support a later split between *C. elegans* and *Drosophila*, but 3 of 6 and 7 of 12 of their slowest evolving proteins actually support a later split between *Drosophila* and mammals. More recently, Wang, Kumar, and Hedges (1999) compiled a set of 18 *C. elegans* genes that also did not exhibit a significant divergence rate bias. By combining these sequences, they obtained a confidence probability greater than 99% that favored a later split between *Drosophila* and mammals. Our results are similar in that all four (five, if the Insr family is included) transmembrane PTK families having mammal, *Drosophila*, and *C. elegans* representatives suggest a later split between *Drosophila* and mammals (Fgfr, Ryk, Ltk/Alk, and Egfr). The well-characterized Fgfr and Egfr families in particular had bootstrap values of 83 and 98, respectively. Most nontransmembrane PTKs, having shorter branch lengths, also suggested a later *Drosophila*-mammal split (not shown). Similar to observations made regarding developmental control genes (reviewed by Ruvkun and Hobert 1998), we also observed that a *Drosophila* PTK gene is more likely to have a mammalian ortholog and no *C. elegans* ortholog (DS03465, Ret, Ror or Nrk, and Trk) than it is to have a *C. elegans* ortholog and no mammalian ortholog (none), again consistent with *Drosophila* and mammals having split later and also independent of substitution rate bias. Therefore, the total molecular evidence would still seem to favor a later split between *Drosophila* and mammals. At the very least, the divergence between *C. elegans* and *Drosophila* is as great as the divergence between *C. elegans* and mammals (Wang, Kumar, and Hedges 1999). This conclusion has particularly important implications for genetic studies of aging, given that such studies are long-term and expensive to conduct using mammalian models. In particular, it suggests that mechanisms, such as stress resistance, that can extend both *C. elegans* and *Drosophila* life spans would be reasonable candidates for extending mammalian life spans.

**old-1 and old-2 Are Members of a Large Gene Family**

Although ours is not the first PTK tree, this study is the first to carefully analyze PTKs known to extend life span. One of our major observations is that *old-1* and *old-2* are members of a surprisingly large multigene family in nematodes. In *C. elegans*, this family represents about one fourth of all PTKs and about half of all transmembrane PTKs. With 16 members, the *old-1/old-2* family surpasses the mammalian Eph/Elk/Eck family (14 members) as the largest known transmembrane PTK family for a single species. Although not shown, *C. elegans* also has a large nontransmembrane family of about 20 members that appears to be most closely related to the Fes family. These two large families contrast with the observation that PTK duplications are unique to the evolution of vertebrates (Suga et al. 1997; Gibson and Spring 1998). On the other hand, large PTK families in *C. elegans* are the exception rather than the rule, as most of the other *C. elegans* PTKs are not members of multigene families.

Because of its large size, the *old-1/old-2* family could provide a unique opportunity to address questions regarding when and how *old-1* acquired its capacity to extend life. Did this capacity develop early or late in the evolution of *C. elegans*? Did it arise little by little or all at once? Was there coevolution with stress resistance? Were particular residue or transcriptional control changes responsible? For example, it can be predicted that if no other members of the *old-1/old-2* family can extend life span, then *old-1*’s ability was acquired mostly after the relatively recent split with *old-2*. The molecular changes likely to be responsible could then be identified using maximum-likelihood/parsimony molecular phylogenetic analyses. In vitro mutagenesis could be used to determine which changes actually extend life span and whether those changes coevolved directly with an ability to increase stress resistance.

As suggested by the tree, the paralogs most closely related to *old-1* and *old-2*, and thus the best candidate genes to also extend the life span of *C. elegans*, appear to be R09D1.12 and M01B2.6. The closeness of this relationship was further supported by a shared N-terminal sequence motif, shared intron/exon junctions, and shared flanking chitinase genes. However, because it is not clear how these features should be weighted relative to shared amino acid changes, we have not attempted to quantify the additional support.

**A Mammalian Ortholog of old-1 and old-2 Appears to Be Unlikely**

A second major observation is that the *old-1/old-2* family does not appear to have a mammalian ortholog. Three explanations are possible. The first is that a mammalian ortholog has not yet been cloned. Given that *old-
A second possibility is that the *old-1/old-2* family is really orthologous to the Pdgfr family in spite of what the tree shows. Although the *old-1/old-2* family appears to be Pdgfr-like in terms of IG domain similarities and intron/exon junctions, an orthologous relationship seems unlikely for a number of reasons: (1) no bootstrap trees grouped the *old-1/old-2* family only with the Pdgfr family (or with the Vegfr family), (2) probable *Drosophila* and hydra orthologs of the Pdgfr family were readily recognizable, (3) the IG branch of the *old-1/old-2* family alone did not group more closely to the Pdgfr family, (4) the genomic organization of *old-1/old-2* subfamilies does not correspond to that of Pdgfr subfamilies, and (5) the Pdgfr family does not exhibit sequence similarity to the *old-1/old-2* family C-terminal motif. The third possibility is that the gene lineage leading to the mammalian ortholog is extinct. This possibility has merit for a number of reasons: (1) the identification of human PTK families has been intensively pursued, yet no new mammalian family has been reported during the 2-year course of this study; (2) only one novel mammalian PTK family was detected in the EST database; (3) most ESTs come from normalized libraries, making it less likely that rare transcripts would be missed (Soares et al. 1994); and (4) PTK-specific degenerate primers to amplify genomic DNA did not suggest any new mammalian PTK families (Oates et al. 1998).

Although no mammalian orthologs of the *old-1/old-2* family were suggested, three potential orthologs were suggested in *C. briggsae*, which diverged from *C. elegans* 20–60 MYA (Heschl and Baillie 1990; Lee et al. 1992; Kennedy et al. 1993). These *C. briggsae* genes lacked an IG domain, matched well with the highly conserved intron-exon junctions (fig. 3), exhibited similarity to the C-terminal motif (fig. 3), and were flanked by a chitinase gene (fig. 2B). Nevertheless, these genes are probably not orthologous to the *old-1/old-2* family. Based on a sample of 10 kinases (including two transmembrane PTKs), kinase domain divergence between *C. elegans* and *C. briggsae* orthologs is only 5%–15%. However, the divergence between the *C. briggsae* and *C. elegans old-1/old-2* family kinase domains is about 50% (uncorrected for multiple substitutions). Therefore, it seems likely that better *C. briggsae* ortholog candidates can be found (only about 15% of the *C. briggsae* genome was sequenced at the time of this analysis), per-
haps even a candidate orthologous to just *old-1* and *old-2* or to just *old-1*.

daf-2 Has No Paralogs, Orthology Is Unclear

A third observation is that *daf-2* does not have paralogs in *C. elegans*. This is curious, given that *C. elegans* has 10 paralogs of an insulin-like gene (Duret et al. 1998). It is also curious that the hydra TK7 gene is more closely related to the mammalian insulin receptor family than *daf-2* is. Can *daf-2* be considered an insulin receptor ortholog? Initial evidence in support of orthology was based on the finding of significant similarity in both the kinase and receptor domains of INSR, IRR, and IG1R (Kimura et al. 1997). However, in the absence of a molecular phylogenetic tree, it is difficult to know whether the divergence coincides with that expected for the species split (Abouheif et al. 1997; Tatusov, Koonin, and Lipman 1997). For example, the criteria of finding significant similarity in both the kinase and the receptor domains when applied to *old-1/old-2* family members F59F3.5 and F59F3.1 would have erroneously led to the conclusion that these genes are probably orthologous to the Vegfrs. In this study, we did find that the split between *daf-2* and the mammalian insulin receptor family is congruent with that expected for a fruit fly-mammal split if TK7 is ignored. Recently, it has also been shown that *daf-2* signaling through *age-1* phosphoinositide-3-OH kinase, *daf-18* phosphatase, *pdk-1* 3-phosphoinositide-dependent kinase, *akt-1* and *akt-2* kinases, and *daf-16* transcription factor parallels the insulin receptor pathway in mammals via PI3K kinase, PTEN phosphatase, PDK-1 kinase, Akt/PKB kinase, and HNF-3 transcription factor (Paradis et al. 1999). Therefore, perhaps hydra TK7 has an unusually slow divergence rate. Resolving this will require finding additional insulin receptor-like genes from other early metazoan phyla (e.g., sponge).

Finally, among PTKs the *old-1/old-2* family and *daf-2* do not share a particularly close relationship, which is due in part to most PTK families having arisen at about the same time (Suga et al. 1997). Therefore, given that *old-1* and *daf-2* appear to interact with the same life extension pathway and that the split between the *old-1/old-2* family and *daf-2* occurred prior to the nematode-mammal split, it is tempting to speculate that PTKs from other families might also extend life span. Theoretically, there is perhaps little reason to think that interactions between PTKs and life extension pathways would be widely conserved given the declining force of natural selection with age (reviewed by Rose 1991). Em-
It is intriguing that the Drosophila insulin receptor–like gene appears to extend the life span of Drosophila (M. Tatar, personal communication of preliminary results) and that a chimeric gene carrying the human Fgfr-l kinase domain, substituting for the kinase domain of old-l, can extend the life span of C. elegans (unpublished data). Regardless, an understanding of the paralogy and orthology of life-extending PTKs is particularly important for interpreting and designing experiments along these lines.

Acknowledgments

Sequences F59F5.3, T22B11.3, F54F7.5, B0198.3, and Y38H6C.20 were suggested to us as probable receptor PTKs by Tony Hunter during the review of this manuscript. We are grateful to Steve Hardies, Sam Henderson, and Chris Link for helpful comments and discussions. Funding was provided by National Institutes of Health grants R01-AG08322, R01-AG16219, P01-AG08761, and AA00195, as well as grants from the American Federation for Aging Research and the Ellison Medical Foundation.

LITERATURE CITED


THOMAS H. EICKBUSH, reviewing editor
Accepted January 4, 2000