Comparison of a Vitellogenin Gene Between Two Distantly Related Rhabditid Nematode Species

Carlos E. Winter,* Cristiane Penha,* and Thomas Blumenthal†

*Departamento de Parasitologia, Instituto de Ciências Biomédicas–USP, Brazil; and †Department of Biology, Indiana University, Bloomington

Three vitellogenin genes from the free-living nematode Caenorhabditis elegans have previously been characterized at the molecular level. In order to study evolutionary relationships within this poorly understood taxon, we have cloned a vitellogenin gene, CEW1-vit-6, from a distantly related species belonging to the same family as C. elegans. Screening of a genomic library with a probe to total poly(A⁺) RNA yielded three clones that hybridized more intensely than all others, and all three corresponded to a single gene homologous to C. elegans vit-6. Comparison of CEW1-vit-6 with Ce-vit-6 reveals both strong similarities and surprising differences. Like Ce-vit-6, the gene is about 5 kb long and contains four unusually small introns (38–41 nt), but only one interrupts the gene at the same location as a Ce-vit-6 intron. The promoter region contains five matches to Vitellogenin Promoter Element 1 (VPE1) and no matches to VPE2, both previously shown to be required for vit gene transcription in C. elegans. Codon usage is in general similar to that of the Ce-vit genes, but a few codon biases are quite different. Alignment of the CEW1-vit-6 protein with the Ce-vit-6 and Ce-vit-2 products suggests the existence of two domains which have evolved at different rates. Sequence comparison shows that nematode vitellogenins are much more closely related to vertebrate than to insect vitellogenins.

Introduction

Animal eggs are closed systems that must contain all the macromolecules and cell structures necessary for the beginning and maintenance of animal development. The major protein constituents of the eggs are yolk proteins or vitellins. The vitellins originate as a precursor glyco-lipoprotein complex, called vitellogenin, made outside the ovary and transported to the growing oocyte, where it is taken up selectively and deposited in the yolk granules.

In most animals a small gene family codes for the vitellogenins. Some of those genes, or their corresponding mRNAs, have already been isolated and sequenced in nematodes (Spieith et al. 1991), insects (Trewitt et al. 1992; Chen, Cho, and Raikhel 1994; Yano et al. 1994), jawless fishes (Sharrock et al. 1992), bone fishes (LaFleur et al. 1995), amphibians (Gerber-Huber et al. 1987), and birds (van het Schip et al. 1987), and all appear to have originated from a common ancestor. Only Drosophila melanogaster and other Cyclorrhaphan Diptera seem to have vitellogenins originating from a different ancestral protein which is related to lipases (Bownes 1992). The main function of the vitellogenin protein moiety has been proposed as simply storage of amino acids to be used at some point during early development (Byrne, Gruber, and Ab 1989). If so, the sequences of different vitellogenins would be expected to be highly divergent, and they are. Nevertheless, the sequence of Caenorhabditis elegans vitellogenin can still be aligned to vertebrate-related proteins (Spieith and Blumenthal 1985; Nardelli et al. 1987), even though these phyla diverged more than 1.200 MYA (Vanfleteren et al. 1994). Thus, evolutionary constraints acting on the vitellogenin molecule must also be reflected in its primary structure. To further investigate vitellogenin evolution, we report here the cloning of a nematode vitellogenin from a distantly related free-living Rhabditid nematode belonging to the same family as Caenorhabditis elegans.

Free-living nematodes, especially C. elegans, have been thoroughly explored as models for basic biological studies (Wood 1988) and can also be used to understand the adaptation of parasitic nematodes to their hosts (Politz and Philipp 1992; Roussel et al. 1995). The phylogenetic relationships among nematodes must be established in order to estimate how relevant studies of C. elegans are to parasitic species. Recently, molecular analysis of globin and myoglobin genes from parasitic and free-living nematodes has been done with the aim of establishing phylogenetic relationships among all nematodes (Blaxter 1993; Blaxter et al. 1994; Vanfleteren et al. 1994). Because the vitellogenins are comparatively fast-evolving, they can be particularly useful for establishing phylogenetic relationships among relatively closely related species. Furthermore, because some regions are conserved even among distantly related animals, comparison of vitellogenin genes can also be useful for understanding distant relationships. In addition, studies of 5′ flanking DNA from different nematodes can help define elements that control expression of the genes (e.g., Zucker-Aprison and Blumenthal 1989).

We characterize in this paper the first gene isolated from a recently identified free-living Rhabditid species. This species has already been used in other studies (see Winter 1992) under the designation of Dolichorhabditis sp. Due to the present dynamic state of the Rhabditid taxonomy, in this paper we will designate this species by its isolate number (CEW1). The genera Oscheius (Andrassy 1976) and Dolichorhabditis (Andrassy 1983), the two most closely related genera to CEW1, are equiv-
alent to *Rhabditis (Oscheius)* according to Sudhaus (1994). The particular organism we study here is most closely related to *Rhabditis (Oscheius) tipulae* Sudhaus 1993, a member of the Dolichura group. This group is equivalent to the genus *Dolichorhabditis* Andrässy, 1983 (Carta, personal communication). Those genera are currently under review by Carta et al. (in preparation).

Although superficially similar in the morphological level, *C. elegans* and CEW1 differ by 48/241 bp of 28S rRNA. These species belong to highly divergent taxonomic groupings comparable to different vertebrate orders (Thomas, personal communication).

The vitellogenin gene we study here has been termed CEW1-vit-6. The comparison of CEW1-vit-6 with its homologue in *C. elegans* (Ce-vit-6) (Spieth et al. 1985) shows that although the proteins they encode are similar in primary structure, the codon usage for some amino acids is very different, as are the intron positions. Using an alignment of all known vitellogenin sequences we show here that, although very similar, these proteins should be considered paralogous rather than orthologous molecules.

Materials and Methods

*C. elegans*, strain N2 (Brenner 1974), and the recently identified nematode CEW1 (deposited at the Cae-norhabditis Genetic Center, CGC) were maintained on *E. coli* NA22 lawns, as described by Winter (1992). Classical procedures were done as described in Sambrook, Fritsch, and Maniatis (1989) and Ausubel (1994).

Genomic Library Construction

A Sau3AI partially digested CEW1 DNA preparation was fractionated in a sucrose gradient as described by Sambrook, Fritsch, and Maniatis (1989). Fractions containing fragments ranging from 15 to 20 kb were pooled and the DNA was precipitated. The random fragments obtained were ligated to XEMBL2 as described by Stratagene. The resulting library was titrated on *E. coli* P2392 giving a titer of 1.5 X 10^11 pfu/mL and stored at 4°C

For intron detection, RT-PCR was performed according to Zorio et al. (1994) with total RNA and primers flanking the four CEW1-vit-6 introns: ObF01 (CCTTCTAGGCCCTTCGGG), ObR01 (TCCTCA-CGGGAATAGGAAAG), ObF02 (GATGACCAAGA-CGGCTTGGCGCA), ObR02 (GGGGGTAGAGA-ACGGATGGGG), ObF03 (AAGGGAATGTCCGA-GAAGACG), ObR03 (TCGGACTTCTGGCAT-TTGG), ObF04 (CTTCCCAGCCGCGGAGATC), and ObR04 (TCACGGGCAAACGCCGGAGA).

Library Screening

The library screenings for vit genes were made using a cDNA probe made from poly(A+)RNA extracted from a mixed culture of CEW1. The cDNA was made using oligo(dT)_{12-18} as a primer and ^{32}P-dATP as the labeled precursor (Sambrook, Fritsch, and Maniatis 1989). The more intense hybridizing plaques were isolated as described above. From the twelve clones isolated, three corresponded to the *CEW1-vit-6* gene: lambdaEMBH3-1, lambdaEMBH1-1, and lambdaEMBG3-1. Only clones lambdaEMBI3-1 and lambdaEMBH1-1 were used in this work.

Subcloning of *CEW1-vit-6* Gene Fragments and Sequencing Strategy

DNA fragments from the recombinant phages were subcloned in pTZ19U (Mead, Szczesna-Skorupa, and Kemper 1986) and sequenced by the dideoxy method (Sanger, Nicklen, and Coulson 1977) using 35S-dATP or ^{32}P-dCTP and the modified T4 DNA polymerase (Sequenase, USB) according to the protocols given by the manufacturer. The sequence of *CEW1-vit-6* is deposited at GenBank under the accession number U35449.

The subclones were first sequenced from both ends using the plasmid primers (universal and reverse). Using these sequences new primers were designed and used to close the gaps. The sequencing gels were read by hand or using DNA-sis® programs (Hitachi) attached to a digitizer table. Not all the regions were read on both strands. Sequences that were read in only one strand came from unambiguous regions of the sequencing gels. The few ambiguous sequences due to G-C compressions were confirmed on both strands. This resolved all GC compressions found without the need to use other methods.

Reverse Transcriptase and PCR Analysis

For intron detection, RT-PCR was performed according to Zorio et al. (1994) with total RNA and primers flanking the four CEW1-vit-6 introns: ObF01 (CCTTCTAGGCCCTTCGGG), ObR01 (TCCTCA-CGGGAATAGGAAAG), ObF02 (GATGACCAAGA-CGGCTTGGCGCA), ObR02 (GGGGGTAGAGA-ACGGATGGGG), ObF03 (AAGGGAATGTCCGA-GAAGACG), ObR03 (TCGGACTTCTGGCAT-TTGG), ObF04 (CTTCCCAGCCGCGGAGATC), and ObR04 (TCACGGGCAAACGCCGGAGA).

Computer Analysis of Amino Acid Sequences

Analyses of the vitellogenin protein sequences were made using the 1993 Genetic Data Environment package, version 2.2 (GDE2.2), running in a Sun Sparstation (with SunOS Release 4.1.3 and Open Windows, version 3). The alignment of different vitellogenins was made using the program CLUSTAL V (Higgins, Bleas-
D13160 (Bombyx mori vitellogenin; Yano et al. 1994); M18061 (Xenopus laevis gene A1; Nardelli et al. 1987); M72980 (Anthonomus grandis vitellogenin gene; Trewitt et al. 1992); M88749 (Ileomyzon unicuspis vitellogenin gene; Sharrock et al. 1992); U02548 (Aedes aegypti vitellogenin gene; Chen, Cho, and Raikhel 1994); U70055 (Fundulus heteroclitus vitellogenin I mRNA; LaFleur et al. 1995); U35449 (CEWI-vit-6; this paper); X03044 (Ce-vit-5; Spieth et al. 1985); X13607 (Gallus gallus vitellogenin II gene; van het Schip et al. 1987); X56212 (Ce-vit-2; Spieth et al. 1991); X56213 (Ce-vit-6; Spieth et al. 1991).

Results and Discussion

Isolation and Characterization of a Vitellogenin Gene

Attempts made to isolate vitellogenin genes from CEW1 using C. elegans probes were unfruitful (results not shown). Thus, contrary to expectations, the vitellogenin genes from the two species were too disparate for cloning by heterologous hybridization. To obtain candidate vit gene clones, we took advantage of the previously isolated vitellogenin genes from the two species were too disparate for cloning by heterologous hybridization. To obtain candidate vit gene clones, we took advantage of the previously isolated vitellogenin genes from the two species. When a CEW1 genomic library was screened with a cDNA probe from total poly(A)+RNA, many clones hybridized, and 12 of the clones which showed the most intense signals were subcloned and sequenced. Figure 1 shows the restriction map, along with the subcloning and sequencing strategy. The 5' end of the mRNA was determined by primer extension (data not shown). When a fragment from one of them (λEMBH3-1) was used to probe a Northern blot containing total RNA from C. elegans and CEW1, a band of around 5 kb was observed in the CEW1 lane only (data not shown). Both the size of this RNA and the fact that the probe failed to hybridize to C. elegans RNA are consistent with a gene coding for a vitellogenin.

Genome Structure

Fragments from two of the clones (λEMBH1-1 and λEMBH3-1) containing the CEW1-vit-6 gene were subcloned and sequenced. Figure 1 shows the restriction map, along with the subcloning and sequencing strategy. The 5' end of the mRNAs was determined by primer extension (data not shown) as described in Materials and Methods and in the legend of figure 2. The positions of the introns in the gene were initially deduced from the A+T composition.
### FIG. 2.—Alignment of the 5' promoter region of CEW1-vit-6, Ce-vit-2, Ce-vit-5, and Ce-vit-6. All sequences were aligned from the TATA box region and were numbered from the start of transcription. The first ATG, the VPEs, the first transcribed nucleotide, and the TATA box are shown inside square boxes. The start of transcription was determined by primer extension using a primer (oligo 28 = AACCCGAAGAGGGCTAGG) complementary to positions 34-48 of the CEW1-VIT-6 mRNA. VPE1 (TGTCACAT) and VPE2 (CTGCTAAC) elements are shown as square boxes over the sequences.

The intron positions were subsequently confirmed by RT-PCR showing that the predicted introns are missing from the mature transcript (results not shown). Four introns were found in CEW1-vit-6. Placement of introns at these four positions allows translation of the remainder of the genomic sequence from the AUG at position +20 to the stop codon at position +5159. After removal of the four introns, the open reading frame is 4983 bp long. Thus, the mRNA is predicted to be approximately 5 kb plus the poly(A) tail, which is approximately the size of the mRNA detected in Northern blots of CEW1 total RNA probed with CEW1-vit-6 probes (data not shown). This mRNA encodes a conceptual protein of 192 kDa (1,660 amino acids). CEW1-VIT-6 is the precursor to VT2 and VT3 vitellins (Winter 1992; unpublished data), just as Ce-VIT-6 is the precursor to YP33 and YP115.
FIG. 3.—Alignment of the proteins coded by CEW1-vit-6 (CEW1-VIT-6), Ce-vit-6 (Ce-VIT-6), and Ce-vit-2 (Ce-VIT-2). In the figure CEW1-V6 corresponds to CEW1-VIT-6. The alignment was done by CLUSTAL V, using PAM 250 as described in Materials and Methods. Matching positions are shadowed, cysteine residues (conserved or not) are shown inside boxes, and intron positions are shown as vertical bars in the sequences and as asterisks (*) under the alignment. Each line in the alignment consists of 53 positions which are occupied by either amino acid residues or spaces inserted by the program (shown as dashes). When an intron interrupts a codon it is shown as a vertical bar which crosses the corresponding residue.

The Promoter Region

In the case of Caenorhabditis (elegans and briggssae) vitellogenin genes, comparison of the 5' flanking DNA revealed the presence of two sequence elements (called VPE1 and VPE2 for Vitellogenin Promoter Element) present in multiple copies in each of the 11 promoter regions analyzed (Zucker-Aprison and Blumenthal 1989). Both VPE1 and VPE2 have subsequently been shown to be required for expression of the vit-2 promoter in vivo (Spieth et al. 1988; MacMorris et al. 1992, 1994). In order to determine how evolutionarily widespread these transcriptional activator binding sites might be, we analyzed the DNA flanking the CEW1-vit-6 gene for the presence of six out of seven matches to the VPE1 and VPE2 consensus sequences. The alignment of 5' noncoding regions of several nematode vitellogenin genes is shown in figure 2. There are five VPE elements, including two which match the consen-
sus at all seven positions in CEW1-vit-6. Although there are no VPE2 in CEW1-vit-6, there are some imperfect matches to VPE2. These VPEs do not precisely match the positions (relative to the TATA box, which begins with a C instead of a T in CEW1-vit-6) of those found in C. elegans. The VPEs do cluster approximately 150 bp upstream of the transcription start site. Functional analysis will have to be performed to determine the role of the VPEs in the expression of CEW1-vit-6 as has been reported for the C. elegans vitellogenin genes (Spieth et al. 1988; MacMorris et al. 1992, 1994).

Intron Analysis

The introns found in CEW1-vit-6 are unusually short, even for nematodes. They are also remarkably uniform in length, ranging from 38 to 41 nt. This is shorter than the typical short C. elegans intron, which is about 50 nt long. The splice sites for the four introns of CEW1 also differ from the consensus determined for C. elegans introns (table 1). In particular, they are less closely matched to a consensus sequence than are C. elegans introns. The latter are characterized by a strong TTTCA/G/A or G at the 3' splice site, while the CEW1-vit-6 introns do not appear to have this extended consensus.

In the C. elegans vit gene family the four vit-6 introns occur in different positions from the four introns of the other genes. Furthermore, most of the introns differ from the 26 intron positions of the vertebrate vitellogenin genes. Strikingly, in the alignment shown in figure 3, only intron 3 of CEW1-vit-6 is found in the same position as one of the introns of Ce-vit-6. This same intron also matches the location of one of the introns in the chicken and Xenopus laevis vitellogenin genes (Spieth et al. 1985; Nardelli et al. 1987).

Codon Usage Bias

Comparison of codon usage bias in CEW1-vit-6 and Ce-vit-6 is shown in table 2. There is a marked preference in CEW1-vit-6 for the use of G or C in the third position when compared to Ce-vit-6: 72% of the CEW1-vit-6 codons use G or C in the third position, whereas only 59% of Ce-vit-6 codons do. Hence the codon usage in the two genes can be surprisingly different (e.g., Pro and Gln codons). These differences in codon usage indicate that CEW1 and C. elegans are not especially closely related species in spite of being morphologically very similar. Clearly, in the Rhabditida, as in other animals, evolution at the molecular level is not mirrored by that of the overall body morphology (Dick-
Table 2
Codon Usage in Ce-vit-6 (Ce) and CEWI-vit-6 (CEWI)

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Ce</th>
<th>CEWI</th>
<th>Amino Acid</th>
<th>Ce</th>
<th>CEWI</th>
<th>Amino Acid</th>
<th>Ce</th>
<th>CEWI</th>
<th>Amino Acid</th>
<th>Ce</th>
<th>CEWI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phe</td>
<td>UUU</td>
<td>11</td>
<td>Ser</td>
<td>UCU</td>
<td>28</td>
<td>Tyr</td>
<td>UAU</td>
<td>10</td>
<td>Cys</td>
<td>UGU</td>
<td>13</td>
</tr>
<tr>
<td>Leu</td>
<td>UUA</td>
<td>1</td>
<td>UGA</td>
<td>UAA</td>
<td>11</td>
<td>End</td>
<td>UAG</td>
<td>100</td>
<td>End</td>
<td>UGA</td>
<td>100</td>
</tr>
<tr>
<td>UUG</td>
<td>5</td>
<td></td>
<td>3</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>CUU</td>
<td>51</td>
<td>Pro</td>
<td>CUC</td>
<td>0</td>
<td>His</td>
<td>CAU</td>
<td>21</td>
<td>20 Arg</td>
<td>CGU</td>
<td>37</td>
</tr>
<tr>
<td>CUC</td>
<td>39</td>
<td>55</td>
<td>CCC</td>
<td>1</td>
<td>64</td>
<td>CAC</td>
<td>79</td>
<td>80</td>
<td>CGC</td>
<td>40</td>
<td>31</td>
</tr>
<tr>
<td>CUA</td>
<td>1</td>
<td>2</td>
<td>CCA</td>
<td>99</td>
<td>9</td>
<td>Gln</td>
<td>CAA</td>
<td>85</td>
<td>17 CGA</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>CUG</td>
<td>0</td>
<td></td>
<td>CCG</td>
<td>0</td>
<td>9</td>
<td>CAG</td>
<td>15</td>
<td>83</td>
<td>CGC</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ile</td>
<td>AUU</td>
<td>21</td>
<td>19 Thr</td>
<td>ACU</td>
<td>29</td>
<td>Asn</td>
<td>AAU</td>
<td>18</td>
<td>10 Ser</td>
<td>AGU</td>
<td>1</td>
</tr>
<tr>
<td>AUC</td>
<td>77</td>
<td>81</td>
<td>ACC</td>
<td>61</td>
<td>59</td>
<td>AAC</td>
<td>82</td>
<td>90</td>
<td>ACG</td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>AGA</td>
<td>2</td>
<td>0</td>
<td>ACA</td>
<td>10</td>
<td>2</td>
<td>Lys</td>
<td>AAA</td>
<td>7</td>
<td>8 Arg</td>
<td>AGA</td>
<td>22</td>
</tr>
<tr>
<td>Met</td>
<td>AUG</td>
<td>100</td>
<td>100 ACG</td>
<td>0</td>
<td>2</td>
<td>AAG</td>
<td>93</td>
<td>92</td>
<td>AGG</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Val</td>
<td>GUU</td>
<td>45</td>
<td>25 Ala</td>
<td>GCU</td>
<td>49</td>
<td>Asp</td>
<td>GAU</td>
<td>50</td>
<td>23 Gly</td>
<td>GGU</td>
<td>0</td>
</tr>
<tr>
<td>GUC</td>
<td>47</td>
<td>70</td>
<td>GCC</td>
<td>43</td>
<td>62</td>
<td>GAC</td>
<td>50</td>
<td>77</td>
<td>GGC</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>GUA</td>
<td>3</td>
<td>5</td>
<td>GCA</td>
<td>6</td>
<td>7</td>
<td>Glu</td>
<td>GAA</td>
<td>24</td>
<td>21 GGA</td>
<td>94</td>
<td>81</td>
</tr>
<tr>
<td>GUG</td>
<td>4</td>
<td>0</td>
<td>GCG</td>
<td>2</td>
<td>3</td>
<td>GAG</td>
<td>76</td>
<td>79</td>
<td>GGG</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

NOTE.—C. elegans values are according to Spieth et al. (1991). Values are given in % of each codon for that amino acid. Prominent differences in codon usage between Ce-vit-6 and CEWI-vit-6 are shown in bold.

The conclusion is supported by the fact that the codon usage table of C. elegans vitellogenin genes is almost the same as the table obtained from a broader sample of sequenced genes from the same species (Stenico, Lloyd, and Sharp 1994). Thus the differences between Ce- and CEWI-vit-6 codon usage frequencies presumably reflect global changes within each organism.

Comparison of Vitellogenin Sequences

When CEW1-VIT-6 sequence was aligned with the vitellogenins from other animals common features could be observed. CEW1-VIT-6 has several regions of high similarity with the other described vitellogenins (e.g., ANAGID in the N-terminal half and CGLCG in the C-terminal half). ANAGID was previously pointed out by Spieth et al. (1991). CGLCG includes a pair of cysteines that occurs in other proteins that are related to the vitellogenins (von Willebrand factor and apolipoprotein b [Baker 1988a, 1988b]). This same motif occurs in E. coli thioredoxin (protein disulfide isomerase) and in other proteins (Boniface and Reichert 1990; see Banaszak, Sharrock, and Timmins 1991). The consensus sequence for the thioredoxin active site is CGXC and it is exactly repeated near the C-termini of all sequenced vitellogenins. We also note some CXXC motifs that are conserved in the vertebrate and nematode vitellogenins (table 3). The insect vitellogenins do not show any CXXC motif in their sequence (data not shown). The function of these conserved cysteine pairs in nematodes and vertebrates is at present unknown. The C-terminal-most CGLCG motif could be part of the receptor recognition site of the vitellogenin, as has been suggested by Sharrock et al. (1992) and Wallace, Hoch, and Carnevali (1990), although there is also evidence for the receptor-binding sequence in the N-terminal part of the vertebrate protein.

Table 3
Position of the Motifs CXXC in Nematode and Vertebrate Vitellogenins

<table>
<thead>
<tr>
<th>Species</th>
<th>N-terminal-most</th>
<th>Median</th>
<th>C-terminal-most</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEW1-VIT-6</td>
<td>219 (CNEC)</td>
<td>—b</td>
<td>1486 (CGLC)</td>
</tr>
<tr>
<td>Ce-VIT-6</td>
<td>232 (CKDC)</td>
<td>—</td>
<td>1476 (CGLC)</td>
</tr>
<tr>
<td>Ce-VIT-5</td>
<td>222 (CKEC)</td>
<td>—</td>
<td>1435 (CGLC)</td>
</tr>
<tr>
<td>Ce-VIT-2</td>
<td>225 (CPEC)</td>
<td>—</td>
<td>1437 (CGLC)</td>
</tr>
<tr>
<td>G. gallus</td>
<td>998 (CQLC)</td>
<td>1716 (CGLC)</td>
<td></td>
</tr>
<tr>
<td>X. laevis</td>
<td>204 (CIISC)</td>
<td>1002 (CIQC)</td>
<td></td>
</tr>
<tr>
<td>I. unicuspis</td>
<td>198 (CPTC)</td>
<td>—</td>
<td>1692 (CGLC)</td>
</tr>
<tr>
<td>F. heteroclitus</td>
<td>202 (CDKC)</td>
<td>—</td>
<td>1577 (CGLC)</td>
</tr>
</tbody>
</table>

a The position of the first C and the actual sequence of the motif.
b The dash (—) means the absence of any motif at this position.

Fig. 4.—Non-Matching Position Index in the three aligned nematode vitellogenins. The graph shows the different values calculated for the lines corresponding to Ce-VIT-6 and Ce-VIT-2, each containing 53 positions (amino acid residues or spaces) as shown in figure 3. The positions of the probable cleavage site of CEW1-VIT-6 and of the Phosvitin sequence in the vertebrate vitellogenins are shown by labeled vertical arrows.
FIG. 5.—Parsimony and distance trees obtained from aligned vitellogenin sequences. (A) Unrooted parsimony and distance trees constructed from the aligned vitellogenin sequences of different animals. The vitellogenin sequences were aligned using CLUSTAL W and the BLOSUM series of alignment matrices (see Materials and Methods). The aligned sequences were used to generate the unrooted parsimony tree shown above at left using the default parameters of the PHYLIP3.5c program (PROTPARS), with bootstrapping. This is the most common tree of 100 trees generated by the parsimony method. The tree is drawn with branches proportional to the calculated distances. Using the same alignment we have calculated a distance tree using the Fitch-Margoliash algorithm of the PHYLIP3.5c package. This distance tree is shown at right. For the 199 examined trees, the sum of squares \( E = \sum (Obs - Exp)^2/Obs^2 \) is 0.20466 and the average percent standard deviation is 4.35317. (B) Rooted parsimony tree obtained from aligned vertebrate and nematode vitellogenins. The vitellogenins were first aligned with CLUSTAL V and PAM250. Using some of the vertebrate vitellogenin sequences as outgroup, rooted trees were generated using PAUP. This procedure was followed for the N-terminal region (positions 1 to 1049 of CEW1-VIT-6) or the C-terminal region (positions 1001 to the end of CEW1-VIT-6) of the nematode vitellogenins. The region containing the phosvitin in the vertebrate vitellogenins is not included in the C-terminal analysis (position 1100 to 1205 of CEW1-VIT-6, in the original alignment). The italicized values over the branches show the percentage of trees showing this branch after bootstrapping, and the values under the branches represent the absolute distance of those lines (which are not drawn proportional to these figures). In this analysis the insect vitellogenins were not included, because they insert a great number of gaps in the C-terminal region. The original alignments can be obtained at ftp://www.icb2.usp.br/pub/align/.


When three of the nematode vitellogenins are aligned some common structural characteristics can be observed (fig. 3). During evolution indels must have occurred in the genes that code for those proteins both before and after the split of the ancestors of *C. elegans* and CEW1. The insertions found at positions 17-24 in Ce-VIT-6 and at positions 61-69 in CEW1-VIT-6 probably occurred after the separation of the species. On the other hand, the indel at positions 881-886 of CEW1-VIT-6 and 894-899 of Ce-VIT-6 must have occurred before the species separated, but after Ce-VIT-6 and Ce-VIT-2 diverged one from the other.

We have also analyzed the relationships between these three vitellogenin proteins by measuring the number of instances in each three-way comparison where only one of the three failed to match the other two. Thus the data reflect changes that occurred in one sequence, but not in the other two. Figure 4 presents the data as a function of distance from the N-terminus. In this figure each point represents the sum of nonmatching positions in a
port the earlier observation of Trewitt et al. (1992) that the vitellogenins of nematodes are more closely related to the vertebrate vitellogenins than they are to the insect

These data also supported the earlier observation of Trewitt et al. (1992) that the vitellogenins of nematodes are more closely related to the vertebrate vitellogenins than they are to the insect vitellogenins. A similar relationship was observed when globin amino acid sequences were used to draw a tree that included nematodes, insects, and vertebrates (Blaxter 1993). Although both sets of data are consistent, the vitellogenins could be paralogous and not orthologous, because of independent gene duplication events which occurred during the evolution of the different animals so far studied (Morris and Cobabe 1991). This conclusion is supported by the fact that 18S rRNA sequence comparison showed that nematodes are more closely related to insects than to vertebrates (Raff, Marshall, and Turbeville 1994). We can speculate that the similarity between nematode and vertebrate vitellogenin amino acid sequences is related to the lipids associated with them. The lipid compositions of the nematode and vertebrate vitellogenins are very similar (see Banaszak, Sharrock, and Timmins 1991) but different from the insect vitellogenins which contain less lipid (Hagedorn and Kunkel 1979; Mundall and Law 1979; Izume and Tomino 1980). Furthermore, the nematode and vertebrate vitellogenins have triacylglycerol as the main neutral lipid (Banaszak, Sharrock, and Timmins 1991), whereas insect vitellogenins (like other insect lipoproteins; see Soulages and Wells 1994) have diacylglycerol as the main neutral lipid (Kunkel and Nordin 1985).

When separate analyses were done with the two domains defined previously, trees rooted in the vertebrates were generated that were topologically very similar (fig. 5B). But the bootstrap values obtained for the CEW1-VIT-6 and Ce-VIT-6 relationship dropped from 100% in the N-terminal domain to 50% in the C-terminal domain. A distance analysis showed that CeW1-VIT-6 is more closely related to Ce-VIT-6 in the N-terminal domain than it is in the C-terminal domain (table 4). On the other hand, the distances in both domains between CEW1-VIT-6 and Ce-VIT-2 are approximately the same. These results, taken together, suggest that the two domains defined for the nematode, and possibly also for the vertebrate vitellogenins, are evolving at different rates, with the C-terminal domain being less conserved than the N-terminal domain.

### Table 4
Calculated Mean Distances for the Two Domains Defined for Vitellogenins

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CEW1-VIT-6</td>
<td>—</td>
<td>0.552</td>
<td>0.668</td>
<td>0.661</td>
<td>0.814</td>
<td>0.803</td>
</tr>
<tr>
<td>2</td>
<td>Ce-VIT-6</td>
<td>0.689</td>
<td>—</td>
<td>0.678</td>
<td>0.673</td>
<td>0.823</td>
<td>0.816</td>
</tr>
<tr>
<td>3</td>
<td>Ce-VIT-5</td>
<td>0.666</td>
<td>0.717</td>
<td>—</td>
<td>0.446</td>
<td>0.826</td>
<td>0.828</td>
</tr>
<tr>
<td>4</td>
<td>Ce-VIT-2</td>
<td>0.680</td>
<td>0.716</td>
<td>0.415</td>
<td>—</td>
<td>0.821</td>
<td>0.800</td>
</tr>
<tr>
<td>5</td>
<td>Chicken</td>
<td>0.841</td>
<td>0.844</td>
<td>0.839</td>
<td>0.840</td>
<td>—</td>
<td>0.680</td>
</tr>
<tr>
<td>6</td>
<td>Lamprey</td>
<td>0.840</td>
<td>0.840</td>
<td>0.841</td>
<td>0.820</td>
<td>0.672</td>
<td>—</td>
</tr>
<tr>
<td>7</td>
<td>Xenopus</td>
<td>0.828</td>
<td>0.814</td>
<td>0.819</td>
<td>0.813</td>
<td>0.602</td>
<td>0.688</td>
</tr>
</tbody>
</table>

**Note:** Above diagonal: N-terminal domain; below diagonal: C-terminal domain. Made by PAUP parsimony.

*Adjusted for missing data.

The distances between CEW1-VIT-6, Ce-VIT-6, and Ce-VIT-2 at the N-terminal and C-terminal domains are shown in bold.
genetic analysis with PAUP, and Gerson A. Calgaro for technical assistance. C.P. is a graduate fellow of the FA-PESP.

LITERATURE CITED


FELSENSTEIN, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.


DAN GRAUR, reviewing editor

Accepted February 1, 1996