J Mol Evol (1999) 49:769–779

JOURNAL OF MOLECULAR EVOLUTION

Gene Conversion and Evolution of Daphniid Hemoglobins (Crustacea, Cladocera)

Paul D.N. Hebert,[1] Young M. Um,[1] Cheryl D. Prokopowich,[1] Derek J. Taylor[2]

[1] Department of Zoology, University of Guelph, Guelph, Ontario N1G 2W1, Canada
[2] Department of Biological Sciences, State University of New York at Buffalo, Buffalo, NY 14260, USA

Received: 8 March 1999 / Accepted: 14 July 1999

Abstract. The extracellular hemoglobins of cladocerans derive from the aggregation of 12 two-domain globin subunits that are apparently encoded by four genes. This study establishes that at least some of these genes occur as a tandem array in both *Daphnia magna* and *Daphnia exilis*. The genes share a uniform structure; a bridge intron separates two globin domains which each include three exons and two introns. Introns are small, averaging just 77 bp, but a longer sequence (2.2–3.2 kb) separates adjacent globin genes. A survey of structural diversity in globin genes from other daphniids revealed three independent cases of intron loss, but exon lengths were identical, excepting a 3-bp insertion in exon 5 of *Simocephalus*. Heterogeneity in the extent of nucleotide divergence was marked among exons, largely as a result of the pronounced diversification of the terminal exon. This variation reflected, in part, varying exposure to concerted evolution. Conversion events were frequent in exons 1–4 but were absent from exons 5 and 6. Because of this difference, the results of phylogenetic analyses were strongly affected by the sequences employed in this construction. Phylogenies based on total nucleotide divergence in exons 1–4 revealed affinities among all genes isolated from a single species, reflecting the impact of gene conversion events. In contrast, phylogenies based on total nucleotide divergence in exons 5 and 6 revealed affinities among orthologous genes from different taxa.

Key Words: Branchiopods — Hemoglobin — Concerted evolution — Crustaceans — Introns

Correspondence to: Dr. Paul D.N. Hebert; *e-mail:* phebert@uoguelph.ca

Introduction

No protein-coding gene family has played a more important role than hemoglobin (Hb) in advancing the understanding of molecular evolution. Most of these insights derive from the study of vertebrates, but there is increasing interest in the globins from invertebrates because of their higher structural diversity (Bolognesi et al. 1997). All vertebrates produce intracellular Hbs (M_r = 64,000) derived from the aggregation of four similarly sized, single-domain globins. The genes encoding these subunits invariably consist of three exons interrupted by two introns which are constantly positioned in the sequences encoding the B and G helices of the globin molecule (Jeffries 1982). In contrast, invertebrate Hbs possess sufficiently varied structures to have forced the development of a nomenclatural system which classifies molecules on the basis of their variation in both the number of oxygen binding domains (single, two, multiple) in each globin subunit and the number of these subunits (single, multiple) which aggregate to form functional Hb (Vinogradov et al. 1993). As a result of this structural diversity, these Hbs vary substantially in size, with molecular weights ranging from 2.0×10^5 to 1.2×10^7 (Peeters et al. 1990). Some invertebrate globin genes share the intron:exon structure of vertebrates, but others possess a central intron interrupting the sequence encoding the E helix at varying positions (Moens et al. 1992; Hardison 1996). In some nematodes and plants, this central intron occurs together with the usual introns, apparently reflecting the structure of the ancestral globin gene (Dixon et al. 1992; Trevaskis et al. 1997). However, other invertebrates (Mansell et al. 1993; Kao et al. 1994;

Gruhl et al. 1997) have only a single intron in the sequence encoding the E helix, or none at all.

The architecture of globin genes in many invertebrates is further complicated in two ways. In species which produce extracellular Hbs, there is typically a 50 to 70-bp coding sequence which is upstream of the 5′ terminus of each globin gene and is separated from it by an intron. This sequence codes for an hydrophobic signal peptide which is attached to the N terminus of the pre-A helix of the globin molecule and aids its transfer from the cell into the hemolymph. A second complication occurs in species with globin subunits consisting of two or more domains; the genes encoding these subunits consist of juxtaposed single-domain units that are separated from one another by a bridge intron. In the mollusk *Barbata* (Suzuki et al. 1996), which produces a two-domain globin subunit, a single bridge intron is present, while anostracan crustaceans produce a nine-domain globin subunit which derives from a series of single-domain units, each separated by a bridge intron (Manning et al. 1990; Trotman et al. 1994; Jellie et al. 1996). Members of three other crustacean lineages (cladocerans, conchostracans, notostracans) also produce two-domain subunits (Peeters et al. 1990), but the intron:exon architecture of their genes is unknown.

The cladoceran crustaceans are a particularly attractive target for comparative studies of Hb evolution. Not only is the group taxonomically diverse (>1000 species, 11 families), but all of its members appear to synthesize Hb, in contrast to many other invertebrate lineages, where only a few taxa share the trait. In addition, their occupancy of marine and freshwater habitats from polar to tropical regions enables studies examining rates of molecular evolution under differing environmental regimes. Finally, in contrast to their constitutive expression in vertebrates, cladoceran Hbs show striking inducibility (Fox and Phear 1953). Although Hbs of the cladoceran *Daphnia* were some of the first proteins to undergo structural analysis (Svedberg and Eriksson-Quensel 1934), study of their evolutionary divergence has been constrained by the limited knowledge of their genetic control. Recent studies have shown that *Daphnia* Hb is composed of 12 globin chains ($M_r = 420,000$) consisting of four subunits (α, β1, β2, and γ), presumably encoded by different genes (Peeters et al. 1990). The isolation of a cDNA clone of one Hb gene from *Daphnia magna* has shown that it encodes an 18-amino acid (aa) signal peptide followed by two globin domains of 176 and 154 aa, respectively, with the first domain possessing a longer pre-A segment (Tokishita et al. 1997).

This study aimed to provide further details concerning the architecture of globin genes in the cladoceran family Daphniidae. It also sought to confirm the presence of a family of globin genes and to ascertain the extent of sequence divergence among them. Finally, because concerted evolution plays such an important role in the evolution of multigene families, this study examined the impacts of gene conversion events on exon evolution.

Materials and Methods

Cladoceran Samples

At least one representative was examined from five of the six genera which comprise the family Daphniidae: *Ceriodaphnia, Daphnia, Daphniopsis, Scapholeberis*, and *Simocephalus.* The genus excluded from the study, *Megafenestra*, is closely allied to *Scapholeberis.* Analysis focused on the genus *Daphnia*, which is partitioned into three deeply divergent subgenera (Colbourne and Hebert 1996). Three members of the subgenus *Ctenodaphnia* were analyzed—*D. exilis* (Texas), *D. magna* (Nebraska), and *D. longicephala* (South Australia); one species of *Hyalodaphnia—D. mendotae* (Ontario); and three species of the nominate subgenus—*D. ambigua* (Florida), *D. obtusa* (New Jersey), and *D. pulicaria* (Ontario). *Ceriodaphnia, Scapholeberis, Simocephalus,* and *Daphniopsis ephemeralis* were also collected from Ontario. A single isolate of each taxon was established in culture, so that all DNA extractions derived from a clonal lineage.

Standard PCR

Total DNA was extracted using CTAB protocols (Doyle and Doyle 1987). Polymerase chain reaction (PCR) amplifications consisted of 50-μl reactions containing buffer (Boehringer Mannheim), 1.5 m*M* $MgCl_2$, a 200 μ*M* concentration of each dNTP, a 0.3 μ*M* concentration of each primer, 1 U of Taq polymerase, and 1 μl of DNA template. PCR amplification involved 36 cycles of 30 s at 94°C, 45 s at 55°C, and 1 min at 72°C and 1 cycle of 6 min at 72°C.

Three primers (see www.cladocera.uoguelph.ca) were designed based upon sequence information for the globin gene in *D. magna* (Tokishita et al. 1997) and tested to ensure that they generated PCR products of the expected size. The products were gel purified using Qiaex II (Qiagen) and sequenced using the ABI prism Big Dye Terminator Cycle sequencing kit on an ABI 377 Automated Sequencer (Perkin–Elmer). PCR amplification and sequencing for *D. exilis, D. magna, D. longicephala, D. mendotae, Daphniopsis,* and *Simocephalus* used the H1F/H1R primers (Fig. 1A), while the H2F/H1R primer combination was used for taxa where HIF failed to bind to template DNA.

Long PCR

Long PCR protocols were used to ascertain if the H1F/H1R primers would generate products longer than the standard 1.3-kb fragment, as expected if there were tandemly duplicated genes which conserved these sequences. This analysis focused on *D. magna* and *D. exilis* because of the higher likelihood of sequence conservation within the subgenus. Long PCR reactions utilized Buffer System 3 and its recommended components in the Expand Long Template System (Boehringer Mannheim). PCR involved 2 min of initial denaturing at 92°C; 10 cycles of 10 s at 92°C, 30 s at 65°C, and 8 min at 68°C; followed by 20 cycles of 10 s at 92°C, 30 s at 65°C, and 8 min at 68°C, with a 20-s increment in each subsequent cycle; and one final cycle of 6 min at 68°C. Identical protocols were employed for reamplifications except for a reduction from 20 to 15 cycles.

In both species, long PCR protocols generated the standard 1.3-kb product and a longer product. The 1.3-kb fragments from *D. exilis* and *D. magna* were directly sequenced, but efforts to sequence the long products failed due to inadequate template. Because of internal priming, reamplifications with H1F/H1R did not lead to an increased con-

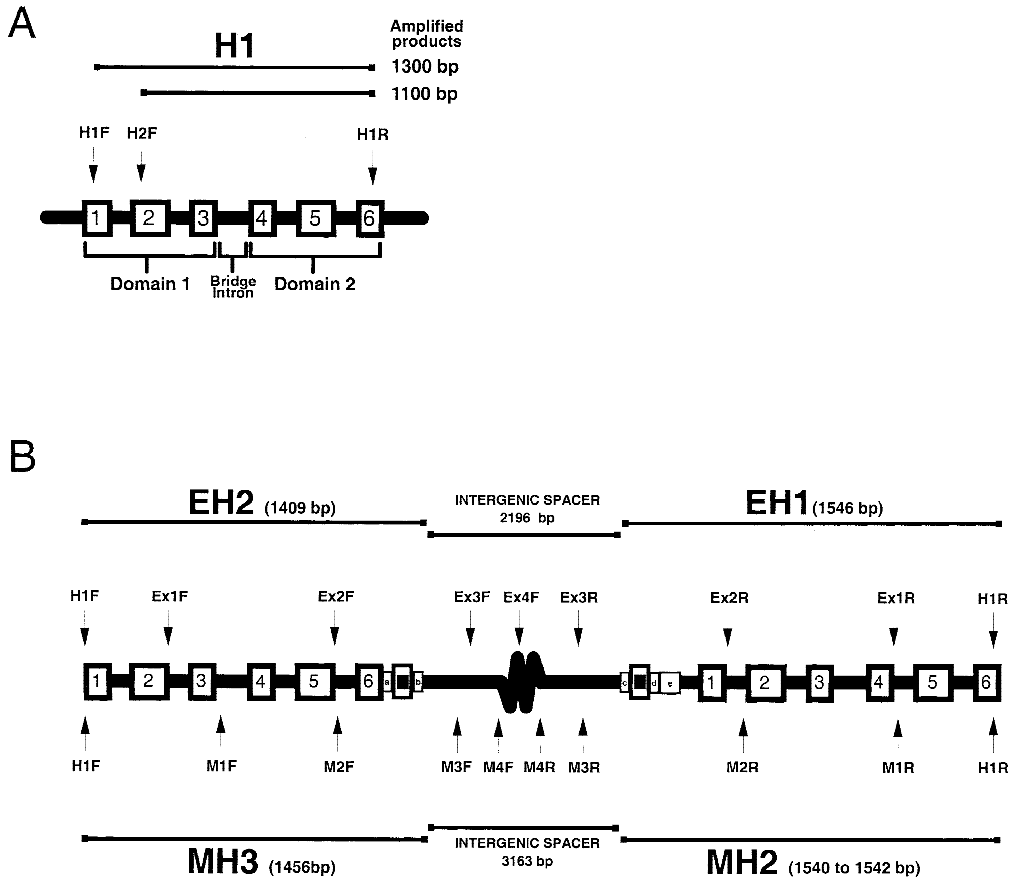

Fig. 1. Diagrammatic representation of Hb genes in cladocerans. **A** Gene structure of the presumptive H1 gene from standard PCR among all daphniids (excepting three cases of intron loss noted elsewhere): exon (□); intron (——). **B** Cloned tandem array of two globin genes from *D. exilis* (EH; 5.2 kb) and *D. magna* (MH; 6.2 kb). Stop codon, a; polyadenylation signal, b; flanking region, c; start codon, d; signal peptide, e; C- and N-terminal extensions, □. The positions of primers used for amplification or sequencing primers are shown by *arrows.*

centration of the large product. As a result, the 5′ end of the H1F primer was extended to include 17 bp of the universal M13(−20) forward primer and this new primer M13H1F was paired with H1R (see www.cladocera.uoguelph.ca). Subsequent reamplification, which employed the M13.20 primer along with H1R to escape internal priming, produced enough of the long product to enable cloning. Following gel purification, the long PCR fragments from *D. exilis* were cloned with the Zero Background Cloning Kit (Invitrogen) and those from *D. magna* were cloned with the TOPO-XL PCR Cloning Kit (Invitrogen). Plasmids were extracted using a standard alkaline lysis prep and directly purified with PEG8000 (13%) precipitation. DNA from five clones of both *D. exilis* and *D. magna* was sequenced, initially using primers encoded in the cloning vector and subsequently using primers designed from intron sequences (see www.cladocera.uoguelph.ca). To confirm the distinctiveness of the globin gene which was dominant under standard PCR protocols, it was also cloned and sequenced.

Data Analysis

All nucleotide and amino acid sequences were aligned against the known *D. magna* globin gene (GenBank accession No. U67067) using the Seqapp 1.9a sequence editor (Gilbert 1992). All distance matrices were constructed in MEGA 1.02 (Kumar et al. 1993) using the K2P model (Kimura 1980) for nucleotide sequences and p-distances for amino acid sequences. Synonymous and nonsynonymous nucleotide substitutions in coding regions were determined in MEGA using the Jukes–Cantor (1969) correction. All potential gene sequences were scanned for regulatory signals and intron placement using GENSCAN (Burge 1998).

Phylogenetic relationships among globin genes were assessed using neighbor-joining (NJ) methods. Cladistic analyses were not employed because of the interpretational problems which arise when there is varying exposure to concerted evolution (Sanderson and Doyle 1992).

Results

Characterization of Hbs: D. exilis *and* D. magna

Sequencing of the 1.3-kb products from standard PCR confirmed that they contained a globin gene, which was termed EH1 for *D. exilis* and MH1 for *D. magna.* Both genes shared the same structure; they contained two globin domains separated by a bridge intron, with each domain consisting of three exons and two introns (Fig. 1A). The MH1 sequence derived in this fashion was identical to that obtained from sequencing a cloned copy of the gene. The long PCR protocol generated a 1.3-kb product and another less abundant product in both *D. exilis* and *D. magna* (5.2 and 6.2 kb, respectively).

No sequence heterogeneity was detected in the long

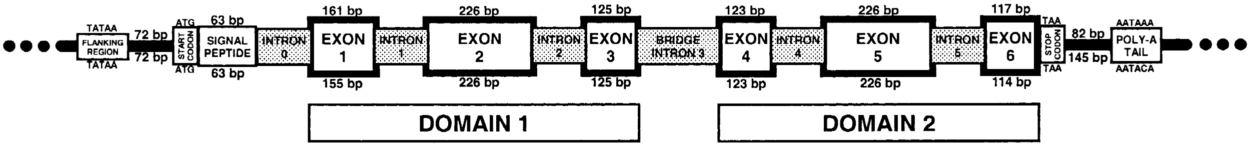

Fig. 2. Fine structure of the globin gene, with sequence lengths marked for *D. exilis* (*above*) and *D. magna* (*below*). Structural information for the 5′ end of the gene (flanking region to intron 1) was derived from EH1 and MH2, while the 3′ end [exon 6 to poly(A)tail] was derived from EH2 and MH3. Central regions were similar in all four genes, excepting intron lengths reported in Table 3.

PCR products from *D. exilis*; each clone consisted of two globin genes separated by a large intergenic spacer. One of these genes was identical to EH1, while the second gene, which was located upstream, was termed EH2 (Fig. 1B). The large PCR product from *D. magna* also contained two Hb genes separated by a large intergenic region, but as both showed substantial sequence divergence from MH1, they were called MH2 and MH3. In contrast to *D. exilis*, the clones from *D. magna* showed sequence heterogeneity. This sequence variation included a single nucleotide insertion and/or deletion in an intron of one gene (MH2), as well as nucleotide substitutions in both exon and intron regions. However, sequence divergence among isolates was small, never exceeding 0.1%.

All five Hb genes (EH1–2, MH1–3) shared the same structure and intron placement (Fig. 1). The first intron in each domain (introns 1 and 4) followed the B12-2 position (i.e., the second nucleotide in the 12th codon of the B helix), the second intron in each domain (introns 2 and 5) followed the G6 codon, and the bridge intron (intron 3) followed the H29-2 position. The sequence from the 5′ terminus of EH1 and MH2 revealed an additional intron (intron 0) separating the pre-A segment from the sequence coding for the signal peptide (22 aa). Variation in exon lengths was detected only in the 5′ and 3′ terminal regions of the global genes (Fig. 2). As a result of this variation, the pre-A region of domain 1 was two aa longer in EH1 than in MH2, while the terminal domain was one aa longer in EH2 than in MH3.

Examination of exon:intron junctions revealed the GT/AG splice signals that ordinarily mark the 5′ and 3′ boundaries of introns (Zhang 1998), except for MH2 clone 4, which had an AT/AG splice signal for intron 2. GENSCAN analysis of all Hb genes resulted in the recognition of expected regulatory signals and exon:intron junctions (Fig. 2). *D. exilis* used a standard polyadenylation signal (AATAAA), while *D. magna* employed a different signal (AATACA) as reported by Tokishita et al. (1997). All introns (introns 0 to 5) were small, averaging just 77 bp in *D. exilis* and 76 bp in *D. magna* (Table 1). When the large intervening sequences separating EH1/EH2 and MH2/MH3 were analyzed using GENSCAN, no coding sequences were detected. The A/T content of these intergenic regions was 69%, a value similar to the A/T content of the introns (63%) but higher than the A/T content of exons (approximately 44–47%). Exons 1 through 5 showed similar nucleotide compositions, with each nucleotide representing approximately 20–30%, but exon 6 showed a significant increase in thymine (as high as 35%) and a decrease in adenine (as low as 13%).

Sequence Divergence: D. exilis *and* D. magna

Because the length variation of introns complicated their alignment, sequence comparisons were restricted to the exons. Exons for all four genes from *D. magna*, including that isolated by Tokishita et al. (1997), were compared, as well as both genes from *D. exilis*. Nucleotide and amino acid divergences among the genes from *D. magna* were substantial (3.0–8.7 and 2.4–11.0%). This trend was sustained in *D. exilis*—its two genes showed 8.4% nucleotide divergence and 7.1% divergence in the amino acid composition of their products (Table 2a).

As expected, levels of sequence divergence for synonymous substitutions were higher (>5×) than for nonsynonymous substitutions (Table 2b). Levels of both synonymous and nonsynonymous substitutions were relatively uniform across all exons, excepting the last (exon 6), which showed a nearly 10-fold higher level of sequence divergence.

Structural Analysis of the Presumptive H1 Gene: All Daphniids

The H1F/H1R primers generated a PCR product for all members of two subgenera (*Ctenodaphnia*, *Hyalodaphnia*) in the genus *Daphnia* but failed for the nominate subgenus. However, these primers did generate product for species of both *Daphniopsis* and *Simocephalus*. The combination of the H2F/H1R primers generated products for all of the remaining taxa, excepting *Scapholeberis*, which failed to amplify with either pair. Most PCR reactions generated a single-sized product, but, as two bands were evident in *Ceriodaphnia*, each was sequenced. Sequences were obtained for all taxa, but due to the admixture of two or more templates, only a few exons were unambiguously aligned for *D. mendotae* and *D. pulicaria*. As a result, these taxa were excluded from the distance and phylogenetic analyses.

The structure of these globin genes was generally similar to those of *D. exilis* and *D. magna*. The positioning of introns was conserved, and the introns were small, with an average length of just 73 bp (Table 1). However,

Table 1. Exon and intron lengths for daphniid Hb genes determined through sequence analysis of cloned or PCR-amplified (∗) products[a]

	Exon (bp)						
	Domain 1			Domain 2			
	1	2	3	4	5	6	Total nt amplified
Daphniid spp.	114	226	125	123	226	78	892
Simocephalus	114	226	125	123	229	78	895

			Intron (bp)					
				Domain 1			Domain 2	
Taxon	Gene	Clone	Pre-A, 0	1	2	Bridge, 3	4	5
D. exilis	1	1,∗	64	61	74	113	77	75
	2	1	—	62	72	117	66	70
D. magna	1	∗	—	61	77	93	73	75
	2	1, 2, 5	67	60	76	117	74	71
		3	67	60	76	117	73	71
		4	67	60	76[b]	117	75	71
	3	1–5	—	61	80	93	69	71
D. longicephala	1	∗	—	60	80	91	71	71
D. ephemeralis	1	∗	—	64	70	87	58	71
D. obtusa[c]	1	∗	—	—	67	69	71	57
D. ambigua[c]	1	∗	—	—	71	69	72	63
Ceriodaphnia[c]	1	∗	—	—	74	141	56	79
	2	∗	—	—	68	72	0	69
Simocephalus	1	∗	—	64	0	63	69	73
D. mendotae	1	∗	—	0	72	—	—	—
D. pulicaria[c]	1	∗	—	—	77	—	—	—

[a] Standard PCR amplifications truncated the 5′ end of exon 1 and the 3′ end of exon 6. Where complete sequence information was obtained through long PCR, the extent of truncation ranged from 41 to 47 and from 36 to 39 bp, respectively. The initiation of exon 1 was marked by the 3′ terminus of intron 0, while exon 6 terminated before the stop codon. Dashes indicate unavailable data. nt, nucleotides.

[b] AT/AG slice site; all other intron boundaries conformed to the usual GT/AG signal.

[c] Denotes taxa amplified with H2F/H1R primers (total exon length of 738 bp); the H1F/H1R primers were used for all others.

several cases of intron loss were detected. Intron 1 was absent from *D. mendotae*, intron 2 was absent from *Simocephalus*, and intron 4 was absent from one of the *Ceriodaphnia* genes.

Phylogenetic Analyses: All Daphniids

Sequence alignments were simplified because of the identity in size of the exon regions amplified by H1F/H2R, excepting the insertion of a single codon in exon 5 of *Simocephalus*. The initial phenetic analysis of nucleotide divergence, which was conducted without an outgroup, established that the two genes from *Ceriodaphnia* were most divergent. As a result, these genes were employed as an outgroup in subsequent phylogenetic analyses.

When synonymous substitutions were examined via NJ analysis (Fig. 3A), all four genes from *D. magna* formed a single clade, as did the pair of genes from *D. exilis*. As a group the globin genes showed patterns of relationship mirroring accepted taxonomic affinities. Globin genes from the three species of *Ctenodaphnia* were most closely allied, and those from the other two subgenera of *Daphnia* showed the next closest relationship. The genus *Daphniopsis*, which is closely allied to *Daphnia*, showed the next closest similarity in globin sequences, while the globins of *Simocephalus* and *Ceriodaphnia* were much more distinctive.

The NJ tree based on nonsynonymous substitutions revealed a different pattern, as genes no longer grouped according to their source taxon (Fig. 3B). Instead the H1 gene from *D. exilis* showed a close affinity to the H1 gene from *D. magna* and to the gene (MTOK) isolated by Tokishita et al. (1997). Similarly the H2 gene from *D. exilis* grouped with the globin gene from *D. longicephala*. The phylogenetic affinities of the remaining genes generally approximated those expected based on taxonomic affinities, but the sole globin gene isolated from *D. ambigua* was unexpectedly divergent.

Phylogenetic analysis (Fig. 3), based on consideration of both synonymous and nonsynonymous substitutions, indicated that the MH1 gene was very similar to the gene isolated by Tokishita et al. (1997). This similarity suggests that the latter gene represents an allelic variant at

Table 2a. Sequence distance matrix for nucleotide and amino acid data based on Kimura's two-parameter (lower-left matrix) and p-distances (upper-right matrix), respectively, with pairwise deletions of gaps and missing data[a]

	EH1	EH2	MH1	MH2	MH3	MTOK
EH1	—	0.071	0.071	0.104	0.098	0.054
EH2	0.084 (0.038)	—	0.122	0.100	0.097	0.104
MH1	0.079 (0.036)	0.111 (0.065)	—	0.096	0.091	0.024
MH2	0.106 (0.056)	0.082 (0.048)	0.087 (0.050)	—	0.110	0.094
MH3	0.112 (0.058)	0.092 (0.054)	0.080 (0.049)	0.082 (0.054)	—	0.088
MTOK	0.069 (0.027)	0.103 (0.056)	0.030 (0.012)	0.087 (0.048)	0.084 (0.047)	—

[a] Values in parentheses represent sequence distance analyses without third codon positions. For the five MH2 and MH3 clones, the average distances among all pairwise comparisons are shown. MTOK represents *D. magna* from Tokishita et al. (1997).

the MH1 locus and it is hereafter termed MH1*. This conclusion was reinforced by consideration of the signal peptides from MH1* and the orthologous gene from *D. exilis* (EH1). Both synonymous and nonsynonymous substitutions supported a much closer relationship between the EH1 and MH1* signal peptides than with the MH2 signal peptide (Fig. 4).

Patterns of Sequence Divergence: D. magna *and* D. exilis

A comparison of nucleotide sequences in the four genes from *D. magna* and two genes from *D. exilis* revealed divergent patterns of substitutions among exons (Fig. 5, Table 3). Concerted evolution was apparent at 16 of 196 codons in exons 1–4 and, in all but one case (15; two nucleotides), involved a single nucleotide position. Interestingly the MH1* gene shared the nucleotides characteristic of MH1 at 13 of these 16 codons. In contrast, none of the nucleotide positions in the 102 codons in exons 5 and 6 showed evidence of gene conversion. Codons showing the effects of concerted evolution were within 95 bp of the 5′ terminus of exons 1, 3, and 4, while in exon 2, which was twice as long as the other exons, conversion tracts occurred within 95 bp of its 5′ and 3′ termini, while its central region lacked these sites.

Twenty-three gene-specific codons were present but only five of these sites were in exons 1–4 (Table 3). With a single exception, the gene-specific codons in these exons were in regions unimpacted by gene conversion, occurring near the 3′ termini of exons 1, 3, and 4 or in the central region of exon 2 (Fig. 5). In contrast, gene-specific substitutions were more common in exon 5 and especially so in exon 6, occurring near the 5′ end of both exons.

Table 2b. Sequence distance matrix for synonymous (lower-left matrix) and nonsynonymous (upper-right matrix) nucleotide substitutions using the Jukes–Cantor-corrected proportion of differences with pairwise deletions for gaps and missing data[a]

	EH1	EH2	MH1	MH2	MH3	MTOK
EH1	—	0.037	0.032	0.055	0.053	0.024
EH2	0.263	—	0.059	0.047	0.046	0.050
MH1	0.252	0.305	—	0.047	0.048	0.011
MH2	0.305	0.214	0.233	—	0.052	0.046
MH3	0.354	0.268	0.193	0.193	—	0.047
MTOK	0.242	0.309	0.097	0.244	0.223	—

[a] The mean distances among all pairwise comparisons are shown for the five MH2 and MH3 clones. MTOK represents *D. magna* from Tokishita et al. (1997).

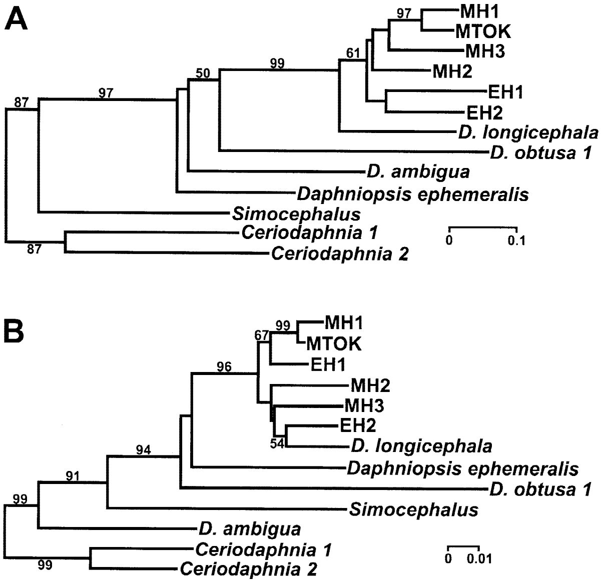

Fig. 3. Phylogenetic trees based on synonymous **(A)** and nonsynonymous **(B)** nucleotide substitution rates produced using the neighbor-joining (NJ) method with the Jukes–Cantor-corrected proportion of differences. The branch numbers represent bootstrap values that were 50% or greater (1000 replicates, with distances based on pairwise deletion of gaps and missing sites) derived from MEGA (Kumar et al. 1993). The NJ trees were rooted with *Ceriodaphnia* genes 1 and 2. Consensus sequences were employed for MH2 and MH3. MTOK represents *D. magna* from Tokishita et al. (1997).

The differential distribution across exons of sites impacted by concerted evolution and of sites showing gene-specific substitutions had strong impacts on phylogenetic analysis. The cladogram for exons 1–4 was markedly different from that for exons 5 and 6, with the terminal pair of exons recovering a gene tree, while analysis of the other exons revealed a species tree (Fig. 6).

Discussion

The earlier sequence characterization of a single globin cDNA clone from *Daphnia magna* indicated its two-domain structure (Tokishita et al. 1997). The present

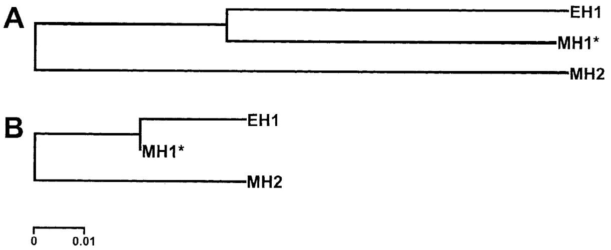

Fig. 4. NJ trees of synonymous **(A)** and nonsynonymous **(B)** nucleotide substitutions for signal peptide sequences for three globin genes from *D. magna* and *D. exilis* based on the Jukes–Cantor corrected proportion of differences.

study has established that these domains are separated by a bridge intron reminiscent of that in the mollusk *Barbata* (Naito et al. 1991; Suzuki et al. 1996). The bridge intron in *Daphnia* interrupts the codon for aa 29 in the H helix, indicating that the two domains are now structurally integrated. Each domain typically consists of three exons and two introns, although several cases of intron deletion were identified in other daphniids. Introns ordinarily interrupted the sequences coding for the B and G helices of the globin molecule and their placement (following B12-2, between G6 and G7) coincided with that in many other taxa, including the only other crustacean which has been examined, *Artemia* (Jellie et al. 1996). However, these intron placements differ from those in globin genes of the dipteran *Chironomus*, which either lack introns or possess a single intron in the sequence coding the E helix (Kao et al. 1994). This divergence suggests that the basal arthropod lineage possessed all three introns and that the central intron was excised in crustaceans, while the B and G introns were lost in insects.

Jellie et al. (1996) proposed that the loss of the central intron (E-helix) antedated the duplication events which led to the multidomain globin gene now present in *Artemia.* The cladocerans support this hypothesis, as they possess one of the presumed transitional stages, a two-domain globin lacking a central intron. Since all four orders comprising the branchiopods produce globins with two or more domains, the first round of globin duplication likely occurred prior to their diversification (Fig. 7). As all of these lineages, barring the cladocerans, which derive from the conchostracans (Taylor et al. 1999), are represented in the fossil record some 400 million years ago (Gray 1988), two-domain globins originated in the early Paleozoic, explaining the low sequence similarity between the two globin domains in *Daphnia* (Tokishita et al. 1997). This model of globin evolution also predicts that a central E intron will be absent from the other two branchiopod groups, conchostracans and notostracans, since its excision antedated domain duplication.

Aside from architectural details, this study has provided information on globin gene diversity in cladocerans. Standard PCR protocols typically led to the amplification of an array of similarly sized products. In some species, one amplified fragment was dominant enough so that its sequence characterization was possible. In other cases, sequencing was successful because PCR products varied in size as a consequence of the deletion of an intron in one of the genes which was amplified. However, the most detailed information on gene diversity resulted from coupling long PCR with cloning. This work showed that some, if not all, globin genes in *Daphnia* occur as tandem arrays separated by relatively long intergenic spacers. Two globin genes were isolated from *D. exilis*, and three from *D. magna.* The occurrence of multiple globin genes is not surprising since peptide analysis has suggested the presence of at least four genes (Peeters et al. 1990). However, the discovery of sequence divergence among all five tandem arrays from *D. magna* suggests that each array must be repeated at least three times in the genome, even if half of the variants represent allelic alternatives.

The sequence analysis of globin genes isolated through long PCR indicated that an additional intron separated the 5′ end of the globin gene from the sequence encoding a signal peptide that not only was the same length (22 aa) in both *D. magna* and *D. exilis*, but showed less sequence divergence than the globin genes themselves. These signal peptides are four amino acids longer than that reported by Tokishita et al. (1997), but they lacked the information on intron placement needed to identify precisely the boundary between the peptide and the pre-A helix of the globin gene. The cloning studies also revealed the separation of daphniid globin genes by a large intergenic region. This A/T-rich tract, which lacks coding sequences, is 2.2 kb long in *D. exilis* and nearly 1 kb longer in *D. magna.* Although these regions are not remarkably large in a broader taxonomic context, they are 30–40 times longer than the introns, which averaged only 77 bp. These introns are far smaller than homologous introns in the globin genes of *Artemia*, which average 2165 bp in length (Jellie et al. 1996), and are among the smallest yet reported for arthropods, being matched only by those in drosophilids (Mount et al. 1992; Moriyama et al. 1998). These two groups of organisms also possess among the smallest genome sizes (2C < 0.4 pg) known for arthropods (Beaton and Hebert 1989). By way of comparison, the genome size of *Artemia* is more than seven times as large (Rheinsmith et al. 1974). Since the maintenance of small genome sizes appears to require an effective mechanism for scouring the genome of nonfunctional DNA (Petrov and Hartl 1997), the large intergenic regions in daphniids likely function in regulating globin expression.

This is now enough information to construct a tentative model of globin gene organization in cladocerans. The sequencing studies on *D. exilis* and *D. magna* suggest that globin genes occur as a tandem array of at least three genes separated by intergenic spacers. As peptide

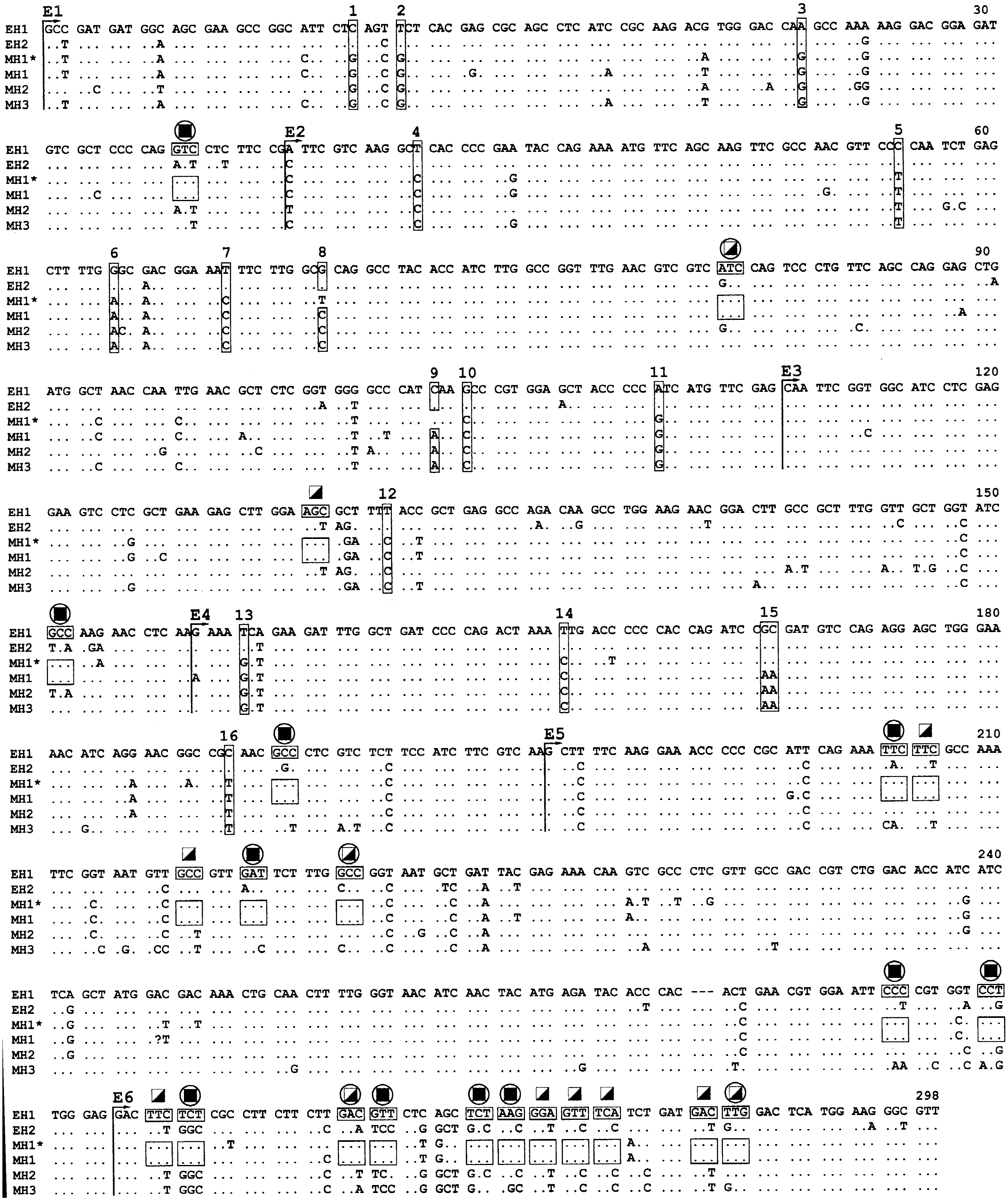

Fig. 5. Nucleotide (nt) sequences of four hemoglobin genes from *D. magna* and two genes from *D. exilis. Bold numerals* indicate 16 codons where nt show concerted evolution, which are those sites where both *D. exilis* genes shared a sequence different from that possessed by the three genes from Nebraskan *D. magna.* Gene-specific codons are those where all H1 genes shared a common sequence which differed from at least two of the other three genes (EH2, MH2, MH3). Gene-specific substitutions are coded; one nt substitution (◪); two or three nt substitutions (■). *Plain symbols* indicate synonymous substitutions, while *circled symbols* are nonsynonymous. Exons (E1 to E6) are shown *above* the sequence, codons are labeled at the *top right* of each sequence row, and identical nts are denoted by *periods.*

Table 3. Number of codons showing evidence of either concerted evolution or gene-specific substitutions for each of six exons in the globin genes of *D. magna* and *D. exilis*

			Codon Changes	
Domain	Exon	No. Codons	Concerted	Gene-Specific
1	1	38	3	1
	2	75	8	1
	3	42	1	2
2	4	41	4	1
	5	76	0	7
	6	26	0	11

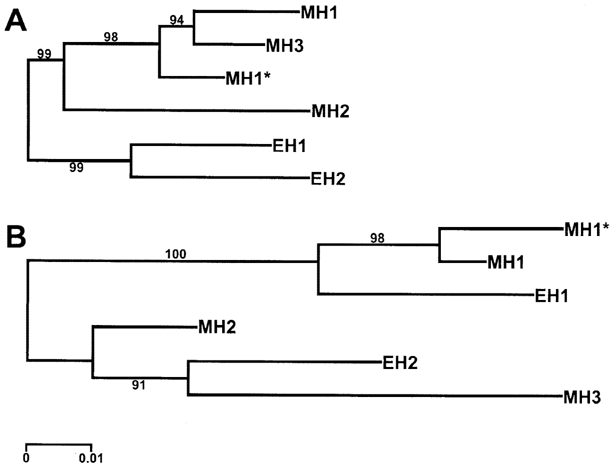

Fig. 6. NJ trees based on analysis of exons 1–4 **(A)** and exons 5 and 6 **(B)** based on Kimura's two-parameter nucleotide distances for four globin genes from *D. magna* and two globin genes from *D. exilis. Branch numbers* represent bootstrap values (1000 replicates, with distances based on pairwise deletion of missing sites) derived from MEGA (Kumar et al. 1993).

analyses have suggested the presence of four globins, each tandem array likely includes a fourth gene. It is possible that the gene from *D. ambigua,* which showed high levels of substitutions at nonsynonymous sites, represents it. The sequence diversity among different isolates of MH2 and MH3 further suggests that there are multiple copies of these tandem arrays in the genome.

Phylogenetic analyses revealed that patterns of relationship among globin genes were influenced by the phenotypic effects of the nucleotide change. Synonymous substitutions were typically shared by all genes of a species, while nonsynonymous substitutions exhibited patterns reflecting the phylogenetic history of the genes. This difference suggests that synonymous substitutions have often swept through the entire array of globin genes in a species, while substitutions provoking amino acid substitutions have not. The evidence for concerted evolution is, in one sense, not surprising, given the broad indications for its occurrence in other globin genes (Zimmer et al. 1980; Jeffries 1982; Ristaldi et al. 1995; Oakenfull and Clegg 1998). However, these cases have involved dramatic cases of sequence convergence apparent across an entire gene or a substantial block of it. These large-scale conversion events result from rare unequal crossing-over events (Ristaldi et al. 1995), while the present results suggest a more diffuse form of concerted evolution based on conversion events involving short tracts of DNA.

Detailed inspection of sequence diversity in the globin exons from *D. magna* and *D. exilis* revealed a complex pattern of sequence change. Codons showing evidence of concerted evolution were localized near the 5′ termini of exons 1–4 or the 3′ terminus of exon 2 and were entirely absent from exons 5 and 6. In contrast, gene-specific substitutions were restricted to regions sheltered from gene conversion. Reflecting this difference, the eight conversion sites in exons 1, 3, and 4 were positioned an average of 48.2 bp downstream of the 5′ termini, while gene-specific substitutions were 89.2 bp downstream. Similarly in exon 2, the eight conversion sites were located an average of 49.9 bp from an intron boundary, while the sole site with a gene-specific substitution was 111.5 bp away from an intron boundary.

The distribution of conversion sites suggests that intron boundaries play a role in their initiation. These impacts, which were restricted to within 95 bp of the intron/exon boundary, were generally asymmetric, affecting sites downstream of the intron. The lengths of conversion tracts in *Daphnia* appear to be similar to those in other animals; Betrán et al. (1997) reported an average track length of 122 bp in *Drosophila*, while Elliot et al. (1998) found that 80% of the conversion tracts in *Mus* were less than 58 bp in length. Although no prior studies have linked the initiation of conversion events to intron boundaries, areas of secondary structure, similar to those likely at introns, are important in initiating conversion. There is also evidence of asymmetries in the directionality of conversion tracts around these points of initiation (Cho et al. 1998), similar to those detected in this study. The present investigation has provided no direct evidence on the frequency of conversion events but the fact that most sites were shared between North American and Japanese *D. magna* indicates that many cases of sequence concordance antedate the isolation of these lineages, which may have occurred more than a million years ago.

These considerations of the patterning of sequence diversity suggest that synonymous substitutions have tended to spread through entire gene families, while nonsynonymous substitutions have been resisted by selection, enabling the maintenance of divergence among paralogous genes. The results also indicate that exons 1–4, especially their 5′ regions, are very susceptible to the erosion of gene-specific differences as a result of conversion, while exons 5 and 6 are protected from such

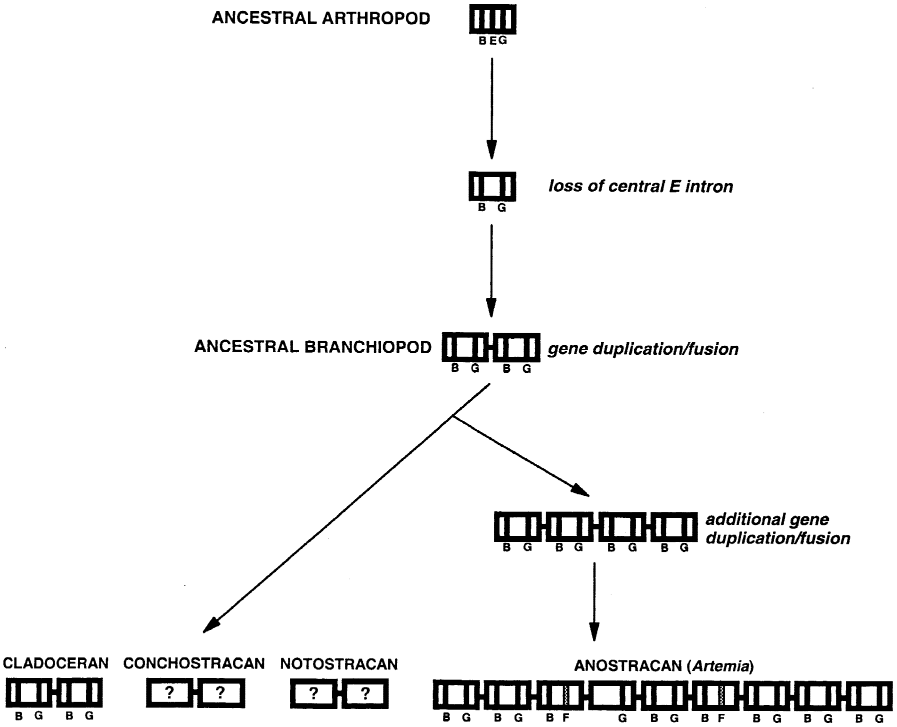

Fig. 7. Tentative reconstruction of major structural shifts in globin arrays of the branchiopod crustaceans.

effects. These underlying patterns of nucleotide change have important consequences for phylogenetic studies. Analyses of synonymous substitutions, especially those which focus on exons 1–4, provide the best approach for studies which seek to clarify phylogenetic affinities among taxa, because the results are relatively insensitive to which member of the gene array is analyzed. In contrast, analyses of nonsynonymous substitutions and of nucleotide divergence in regions protected from gene conversion are the optimal approach for studies which seek either to identify orthologous genes or to reconstruct the evolutionary pathways which led to the modern array of globin genes.

Although this study has indicated that cladoceran hemoglobins are encoded by a complex multigenic family, they represent an attractive target for further molecular characterization because introns are so small that structural blocks can be easily isolated. The use of these genes in phylogenetic contexts is currently compromised by the difficulties in jointly ensuring that analysis is focused on orthologous genes and in assessing the effects of converted evolution on sequence diversity. However, with further knowledge of globin structure, it will be possible both to restrict analysis to orthologous genes and to evaluate the effects of concerted evolution. Such information will set the stage for studies which provide new insights concerning the environmental modulation of globin expression which is so prominent in these organisms. In addition, comparative studies probing sequence diversity in these genes across both environmental gradients and taxonomic assemblages should provide intriguing new insights into globin evolution.

Acknowledgments. This study was supported by a research grant from NSERC to P.D.N.H. Teri Crease provided helpful comments on an early version of the manuscript. We also thank Angela Holliss for key assistance with the sequencing.

References

Beaton MJ, Hebert PDN (1989) Miniature genomes and endopolyploidy in cladoceran crustaceans. Genome 32:1048–1053

Betrán E, Rozas J, Navarro A, Barbadilla A (1997) The estimation of the number and length distribution of gene conversion tracts from population DNA sequence data. Genetics 146:89–99

Bolognesi M, Borda D, Rizzi M, Tarricone C, Ascienzi P (1997) Nonvertebrate hemoglobins: Structural bases for reactivity. Prog Biophys Mol Biol 68:29–68

Burge C (1998) GENSCAN. Published electronically on the Internet at gnomic.stanford.edu/GENSCANW.html

Cho JW, Khalsa GJ, Nickoloff JA (1998) Gene-conversion tract directionality is influenced by the chromosome environment. Curr Genet 34:269–279

Colbourne JK, Hebert PDN (1996) The systematics of North American *Daphnia* (Crustacea: Anomopoda): A molecular phylogenetic approach. Trans R Soc Ser B 351:349–360

Dixon B, Walker B, Kimmins W, Pohajduk B (1992) A nematode hemoglobin gene contains an intron previously thought to be unique to plants. J Mol Evol 35:131–136

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Elliott B, Richardson C, Winderbaum J, Nickoloff JA, Jasin M (1998) Gene conversion tracts from double-strand break repair in mammalian cells. Mol Cell Biol 18:93–101

Fox HM, Phear EA (1953) Factors influencing hemoglobin synthesis in *Daphnia.* Proc R Soc Lond B 141:238–242

Gilbert D (1992) SeqApp, Version 1.9a, a multiple-sequence editor for MacIntosh computers. Published electronically on the Internet at fly.bio.Indiana.edu

Gray J (1988) Evolution of the freshwater ecosystem: The fossil record. Palaeogeogr Palaeoclimatol Palaeoecol 62:1–214

Gruhl M, Kao WY, Bergtrom G (1997) Evolution of orthologous intronless and intron-bearing globin genes in two insect species. J Mol Evol 45:499–508

Hardison RC (1996) A brief history of hemoglobins: Plants, animals, protist and bacteria. Proc Natl Acad Sci USA 93:5675–5679

Jeffries AJ (1982) Evolution of globin genes. In: Dover GA, Flavell RB (eds) Genome evolution. Academic Press, New York, pp 157–176

Jellie AM, Tate WP, Trotman CNA (1996) Evolutionary history of introns in a multidomain globin gene. J Mol Evol 42:641–647

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, pp 21–132

Kao WY, Trewitt PM, Bergtrom G (1994) Intron-containing globin genes in the insect *Chironomus thummi.* J Mol Evol 38:241–249

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kumar S, Tamura K, Nei M (1993) MEGA: Molecular evolutionary genetic analysis, Version 1.02. Distributed by the authors. Pennsylvania State University, University Park

Manning AM, Trotman CNA, Tate WP (1990) Evolution of a polymeric globin in the brine shrimp *Artemia.* Nature 348:653–656

Mansell JB, Timms K, Tate WP, Moens L, Trotman CNA (1993) Expression of a globin gene in *Caenorhabditis elegans.* Biochem Mol Biol Int 30:643–647

Moens L, Vanfleteren J, De Baere I, Jellie AM, Tate W, Trotman CNA (1992) Unexpected intron location in non-vertebrate globin genes. FEBS Lett 312:105–109

Moriyama EN, Petrov DA, Hartl DL (1998) Genome size and intron size in *Drosophila.* Mol Biol Evol 15:770–773

Mount SM, Burks C, Hertz G, Stormo GD, White O, Fields C (1992) Splicing signals in *Drosophila:* Intron size, information content and consensus sequences. Nucleic Acids Res 20:4255–4262

Naito Y, Riggs CK, Vandergon TL, Riggs AF (1991) Origin of a “bridge” intron in the gene for a two-domain globin. Proc Natl Acad Sci USA 88:6672–6676

Oakenfull EA, Clegg JB (1998) Phylogenetic relationships within the genus *Equus* and the evolution of α and O globin genes. J Mol Evol 47:772–783

Peeters K, Mertens J, Hebert P, Moens L (1990) The globin composition of *Daphnia pulex* hemoglobin and the comparison of the amino acid composition of invertebrate hemoglobins. Comp Biochem Physiol 97B:369–381

Petrov DA, Hartl DL (1997) Trash DNA is what gets thrown away: High rate of DNA loss in *Drosophila.* Gene 205:279–289

Rheinsmith FL, Hinegardner R, Bachmann K (1974) Nuclear DNA amounts in Crustacea. Comp Biochem Physiol 48B:343–348

Ristaldi MS, Casala S, Rando A, Veshi R (1995) Sheep α-globin gene sequences: Implications for their concerted evolution and for the down-regulation of the 3′ genes. J Mol Evol 40:349–353

Sanderson MJ, Doyle JJ (1992) Reconstruction of organismal and gene phylogenies from data on multigene families: Concerted evolution, homoplasy, and confidence. Syst Biol 41:4–17

Suzuki T, Kawasaki Y, Arita T, Nakamura A (1996) Two-domain hemoglobin of the blood clam *Barbata lima* resulted from the recent gene duplication of the single domain δ chain. Biochem J 313:561–566

Svedberg T, Eriksson-Quensel I (1934) The molecular weight of erythrocryorins—II. J Am Chem Soc 56:1700–1706

Taylor DJ, Crease TJ, Brown W (1999) Phylogenetic evidence for a single long-lived clade of crustacean cyclic parthenogens and its implications for the evolution of sex. Proc R Soc Ser B 266:791–797

Tokishita S, Shiga Y, Kimura S, Ohta T, Kobayashi M, Hanazato T, Yamagata H (1997) Cloning and analysis of a cDNA encoding a two-domain hemoglobin chain from the water flea *Daphnia magna.* Gene 189:73–78

Trevaskis B, Watts RA, Andersson CR, Llewellyn DJ, Hargrove MS, Olson JS, Dennis ES, Peacock WJ (1997) Two hemoglobin genes in *Arabidopsis thaliana*: The evolutionary origins of hemoglobins. Proc Natl Acad Sci USA 94:12230–12234

Trotman CNA, Manning AM, Bray JA, Jellie AM, Moens L, Tate WP (1994) Interdomain linkage in the polymeric hemoglobin molecule of *Artemia.* J Mol Evol 38:628–636

Vinogradov SN, Walz DA, Pohajdak B, Moens L, Kapp OH, Suzuki T, Trotman CNA (1993) Adventitious variability? The amino acid sequences of nonvertebrate globins. Comp Biochem Physiol 106B: 1–26

Zhang SH (1998) Origin and evolution of exon/intron junctions. Spec Sci Technol 21:17–21

Zimmer EA, Martin SL, Beverley SM, Kwon YW, Wilson AC (1980) Rapid duplication and loss of genes coding for the α chains of hemoglobin. Proc Natl Acad Sci USA 77:2158–2162