

# Selection on the Structural Stability of a Ribosomal RNA Expansion Segment in *Daphnia obtusa*

Seanna J. McTaggart and Teresa J. Crease

Department of Zoology, University of Guelph, Guelph, Ontario, Canada

The high rate of sequence divergence in nuclear ribosomal RNA (rRNA) expansion segments offers a unique opportunity to study the importance of natural selection in their evolution. To this end, we polymerase chain reaction amplified and cloned a 589-nt fragment of the 18S rRNA gene containing expansion segments 43/e1 and 43/e4 from six individual *Daphnia obtusa* from four populations. We screened 2,588 clones using single-stranded conformation polymorphism analysis and identified 103 unique haplotype sequences. We detected two pairs of indel sites in segment 43/e4 that complement each other when the secondary structure of the linear sequence is formed. Seven of the 12 observed combinations of length variants at these four sites (haplotypes) are shared between individuals from different populations, which may suggest that some of the length variation was present in their common ancestor. Haplotypes with uncompensated indels were only observed at low frequencies, while compensated indel haplotypes were found at a wide range of frequencies, supporting the hypothesis that the energetic stability of expansion segments is a trait under natural selection. In addition, there was strong linkage disequilibrium between the four complementary indel sites, particularly those that pair with one another in the secondary structure. Despite selection against unpaired bulges at these four indel sites, some nucleotides that form unpaired bulges are highly conserved in segment 43/e4, indicating that they are under a different selective constraint, possibly due to their role in higher level structural interactions.

## Introduction

Eukaryotic nuclear ribosomal DNA (rDNA) consists of one or more tandem arrays containing three genes (18S, 5.8S, 28S) encoding ribosomal RNA (rRNA) separated by spacers. Each rRNA molecule folds into a secondary structure that consists of highly conserved core regions interspersed with rapidly evolving regions called expansion segments, divergent domains, or variable regions (Wuyts, Van de Peer, and De Wachter 2001). The conservation of the core structure, which essentially extends across all three domains of life (Wuyts, Van de Peer, and De Wachter 2001), is presumably the result of selection maintaining the structurally dependent activity of the ribosome. The critical role of rRNA in protein synthesis has been proposed as a reason for both the high multiplicity of rRNA genes and the absence of divergence among the copies.

The intraspecific sequence homogeneity of the core regions of rRNA gene copies is thought to be the result of unequal crossing-over and gene conversion, which together drive the process of concerted evolution (Dover 1982). In contrast to this homogeneity, the divergence of the expansion segments may be so rapid that length variation can be observed among members of the same population and even among the copies present within an individual (Gonzalez et al. 1985; Leffers and Anderson 1993; Crease and Taylor 1998; Gonzalez and Sylvester 2001).

The extent of intraspecific length variation in expansion segments is well documented and varies greatly across different taxa. For example, in the subclass Pterygota (Class Insecta), expansion segments account for 32% of the total length of the 18S rRNA gene in *Carausius morosus*, while they account for 58% of the gene length in *Xenos vesparum* (Crease and Colbourne 1998). Based on the high degree of repetitive nucleotide motifs often found in length variable

regions, Ruiz Linares, Hancock, and Dover (1991) hypothesized that replication slippage is likely responsible for the elongation of expansion segments in many species. However, length expansion is not always correlated with significant levels of sequence simplicity (e.g., *Phreato-moeba balamuthi*, Hinkle et al. 1994; *Daphnia pulex*, Crease and Colbourne 1998), perhaps due to the ancient origin of slippage events that have subsequently been masked by point mutations (Crease and Taylor 1998). Alternatively, indels may be caused by unequal crossing-over in repetitive regions. For example, McLain (2001) observed a 50-bp deletion in degenerate subrepeats of an expansion segment in the 28S rRNA gene of the tick, *Ixodes persulcatus*.

Despite the magnitude of length variation between species, a common feature of expansion segments is that they generally fold into one or more helices, which are the most energetically stable conformation that a linear RNA molecule can assume (Juan and Wilson 1999). Several authors have suggested that the stability of the secondary structure of a sequence may be a trait under natural selection (Hancock, Tautz, and Dover 1988; Ruiz Linares, Hancock, and Dover 1991). Hancock and Dover (1990) proposed a two-step process called compensatory slippage to explain how the rRNA molecule can persist as a stable helix while potentially accumulating indels in the linear DNA molecule. The first step is a replication slippage event that causes a destabilizing bulge in a base-pairing region. Owing to the high multiplicity of rRNA genes, this energetically unfavorable copy could elude elimination by selection and remain at low frequency in the same way that a deleterious recessive allele of a single-copy locus can persist at a low frequency in a population (Hartl and Clark 1989). Next, a second slippage event inserts nucleotides that complement those added in the first step, thus compensating the bulged nucleotides and restoring the double-stranded helix. If the new longer variant is selectively neutral relative to the original sequence, its frequency will change as a result of random genetic drift (Hancock and

Key words: *Daphnia obtusa*, ribosomal RNA, secondary structure, expansion segment.

E-mail: tcrease@uoguelph.ca.

*Mol. Biol. Evol.* 22(5):1309–1319. 2005

doi:10.1093/molbev/msi119

Advance Access publication March 2, 2005

Dover 1990). Under the primary assumption of this model (that selection acts to maximize structural stability of expansion segments), two qualitative predictions can be made: uncompensated length variants will only occur at very low frequencies within individuals due to the action of selection, but compensated length variants will occur at a wide range of frequencies, as is expected among neutral alleles whose frequencies are influenced primarily by drift. These predictions have yet to be tested, although several cases of intraindividual variation have been reported.

Gonzalez et al. (1985) were the first to document intraindividual length variation in an rRNA gene. They sequenced six recombinant plasmids containing the D6 expansion segment from the human 28S rRNA gene and found four length variants that differed in the copy number of a trinucleotide repeat (CGG). In contrast to the first prediction of the compensatory slippage model, the secondary structures of two of the variants contained unpaired regions or "bulges" that were not present in the other two (Gonzalez et al. 1985). The variant frequencies were not calculated in this study; however, given that there are approximately 300 copies of rRNA genes per haploid genome in humans (Gonzalez et al. 1985) and that only six plasmids were sequenced, it is unlikely that both of the uncompensated variants were present at low frequency. Similarly, Holzmann, Piller, and Pawlowski (1996) described the high levels of intraindividual length variation in the 28S rRNA gene of *Ammonia* sp. (Foraminifera, Protozoa) and noted that no compensatory mutations were observed. A recent study of the plastid 23S rRNA gene from the holoparasite *Cynomorium corrineum* (Cynomoriaceae) also identified sequence variants containing indels that appeared to destabilize the secondary structure of the molecule (García, Nicholson, and Nickrent 2004). However, as the studies that have documented intraindividual variation only analyzed a small number of sequences and variant frequencies were not calculated, it is not possible to infer how the variation is partitioned among all the gene copies.

*Daphnia* are freshwater crustaceans in the order Cladocera that are known to have four expansion segments in the 18S rRNA gene (Crease and Colbourne 1998). Cladocerans reproduce by cyclical parthenogenesis, and concerted evolution is known to occur even during the parthenogenetic phase of reproduction (Crease and Lynch 1991). Previous work has shown that intraindividual length variation occurs in the expansion segments of some species of *Daphnia* (Crease and Taylor 1998). In this study, we amplified and cloned a fragment containing two expansion segments and the flanking core region of the 18S rRNA gene from six individual *Daphnia obtusa* from four populations and assayed several hundred clones from each individual in the first comprehensive survey of intraindividual variation of expansion segments in rRNA genes. Based on the hypothesis that selection favors more energetically stable expansion segments over less stable structures, we expect that length variants that maintain base pairing in the expansion segments (compensated indels) will be selectively neutral relative to each other and therefore will be found at a wide range of frequencies, while length variants that create bulges in the expansion segments (uncompensated indels) will only be present at low frequencies.

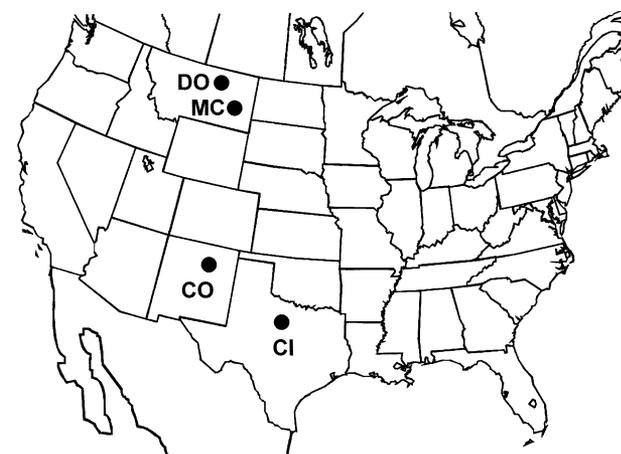


FIG. 1.—Map of the United States of America showing the location of the four *Daphnia obtusa* populations sampled in this study.

Furthermore, we expect to see linkage disequilibrium between those variants at complementary indel sites that result in compensated structures.

## Methods

### DNA Extraction and Polymerase Chain Reaction Amplification

Using the GenElute mammalian Genomic DNA Mini-prep Kit (Sigma, St. Louis, Mo.), we isolated total genomic DNA from six individual *D. obtusa*: two from a population near Cisco, TX (denoted CIa and CIb), two from a population near Miles City, MT (denoted MCa and MCb), one from a population near Dodson, MT (denoted DO), and one from a population in Coconino County, AZ (denoted CO) (fig. 1). After ethanol precipitation, the DNA was resuspended in 25  $\mu$ l of sterile distilled water. We used primers 1413F and 2004R, described in Crease and Colbourne (1998), to amplify a 589-nt fragment of the 18S rRNA gene containing helices 37 through 46 according to the helix nomenclature of Wuyts, Van de Peer, and De Wachter (2001). This fragment contains the eukaryotic expansion segments 43/e1 and 43/e4, which are adjacent to one another and are the focus of the subsequent analyses. We used 5  $\mu$ l of the genomic DNA extraction in a 50- $\mu$ l polymerase chain reaction (PCR) containing 1.5 mM MgCl<sub>2</sub>, 0.2 mM of each deoxynucleoside triphosphate (dNTP), 1 $\times$  reaction buffer (Invitrogen, Carlsbad, Calif.), 1 mM of each primer, and 1 U of Platinum High Fidelity *Taq* polymerase (Invitrogen, Carlsbad, Calif.). The reactions were carried out in 500- $\mu$ l thin-walled tubes in an MJ-PTC 100 thermocycler (MJ Research, Waltham, Mass.). Amplification conditions were 94°C for 2 min, 30 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 1 min, followed by a final extension step of 72°C for 5 min. The amplified product was visualized with ethidium bromide on a 0.8% agarose gel in TBE buffer. We purified the PCR product from the agarose gel using a freeze-thaw method (Penton, Hebert, and Crease 2004) and resuspended the resulting DNA in 20  $\mu$ l of sterile distilled water.

## Plasmid Cloning

We cloned the amplified fragments using the Topo TA Cloning Kit (Invitrogen, Carlsbad, Calif.) (CIa, CIb, and CO) and pGEM T Cloning Kit (Promega, Madison, Wisc.) (MCA, MCB, and DO) according to the manufacturer's specifications. Three hundred and eighty-four colonies from CO and 480 colonies from each of CIa, CIb, MCA, MCB, and DO were streaked onto Luria Broth (Difco, Franklin Lakes, N.J.) agar plates with ampicillin (200 mg liter<sup>-1</sup>) and grown overnight at 37°C. Some of the resulting growth was transferred to a single well of a 96-well culture plate containing 20% glycerol in Luria Broth (Difco, Franklin Lakes, N.J.) and ampicillin (200 mg liter<sup>-1</sup>). The cultures were grown overnight at 37°C and stored at -80°C.

## Identification of Variants by Single-Stranded Conformation Polymorphism Analysis and Sequencing

We amplified expansion segments 43/e1 + e4 directly from the bacterial colonies with primers 1522F (5'-ATTCCGATAACGAACGAG) and 1880R (5'-GAA-GACTGCGTGACGGAC) (numbering corresponds to the *D. pulex* 18S rRNA gene sequence, GenBank accession number AF014011) using the following 10- $\mu$ l reaction: 1 $\times$  buffer, 1.5 mM MgCl<sub>2</sub>, 0.05 mM of each dNTP, 0.75  $\mu$ M of each primer labeled with the fluorescent dye HEX (Lab Services, University of Guelph, Guelph, Ontario, Canada), and 1 U of *Taq* polymerase. DNA was added by touching a sterile pipette tip to one of the bacterial streaks and twirling it briefly in the reaction mix. Amplification conditions were 94°C for 1 min, 35 cycles of 94°C for 20 s, 55°C for 20 s, and 72°C for 1 min, followed by 72°C for 5 min. A molecular size standard was made by amplifying and pooling three PCR fragments of the human mitochondrial *Cytochrome c oxidase I* gene with the following primers: HSCO1F, 5'-CTACAAACCACAAAGACATYGGAAC (labeled with HEX); HSCO1-165R, 5'-CAAATGCA-TGGGCTGTGACGATAAC; HSCO1-177R, 5'-AGAA-GATTATTACAAATGCATGGGC; and HSCO1-241R, 5'-GCACCGATTATTAGGGGAAGTAGTC. The PCR conditions were the same as those for samples amplified with 1522F and 1880R, except that the reaction volume was 100  $\mu$ l. Following amplification of both the expansion segment and the size standard, a fivefold volume of 95% deionized formamide (Unison Biotek, Hamilton, Canada) with 1 mg ml<sup>-1</sup> blue dextran in 500 mM Na<sub>2</sub>EDTA (pH 8.0) was added to each sample. All samples were then denatured at 95°C for 5 min and placed immediately on ice.

We loaded 1  $\mu$ l of each denatured sample into a 6% native polyacrylamide gel (37.5:1 acrylamide:bis-acrylamide, 1 $\times$  TBE, 5% glycerol, 0.6% ammonium persulfate, 0.05% N,N,N',N'-trimethylethylene diamine). We also ran 2  $\mu$ l of size standard in every lane of the gel. Gels were run at 4°C for 5 h at 35 W. All gels were scanned on a Hitachi FMBIO II at 585 nm. We calculated the mobility of each sample relative to the in-lane size standard using the 1D-Gel Analysis tool (FMBIO Analytical software, version 8.0). For each gel, one representative was chosen from each group of samples that had the same mobility. The representative samples from all gels were then run together in order to directly compare mobility groups across gels. Once all

the mobility groups had been identified, we selected one to eight clones from each group for sequencing.

Each selected clone was grown by adding 10  $\mu$ l of the glycerol stock to 1.5 ml of Luria Broth + ampicillin (200 mg liter<sup>-1</sup>) in a 2-ml microcentrifuge tube and shaking it overnight at 37°C. Plasmid DNA was isolated according to the method of Ahn et al. (2000) and resuspended in 20  $\mu$ l of sterile distilled water (Invitrogen, Carlsbad, Calif.). We ran 2  $\mu$ l of this preparation on an agarose gel in TBE to estimate its concentration by eye.

Approximately 50 ng of plasmid DNA was sequenced with the Big Dye Terminator Kit, version 3.1 (Applied Biosystems, Foster City, Calif.), according to the manufacturer's instructions, except that the total reaction volume was 10  $\mu$ l and the volume of Big Dye reaction mix was 2  $\mu$ l. We used the 1413F primer for all sequencing reactions, which were resolved on an ABI 377 automated sequencer (Applied Biosystems, Foster City, Calif.).

## Identification of Length Variants Using Denaturing Polyacrylamide Gels

The assignment of all clones to a mobility class from the single-stranded conformation polymorphism (SSCP) gels was problematic, as many of the mobility differences were very small. Thus, we ran denaturing polyacrylamide gels to determine the frequency of the length variants in each individual. The same PCR products that were run on the SSCP gels were also loaded onto 8% denaturing gels (37.5:1 acrylamide:bis-acrylamide; 1 $\times$  TBE, 7.8 M urea) and electrophoresed for 5 h at 35 W. The resulting gels were scanned on the FMBIO II as before and scored by eye. If the overall length of two distinct variants was the same, we separated the variants from one another based on their mobilities on the SSCP gels when possible. Otherwise, the total number of variants was divided equally between the two length classes.

## Data Analysis

We aligned all the *D. obtusa* sequences using Sequencher 4.0.5 (Gene Codes Corp., Ann Arbor, Mich.) and corrected the alignment by eye with reference to the secondary structure proposed by Crease and Colbourne (1998). A consensus sequence was then generated using only the unique sequences from this alignment. We compared each unique sequence to this consensus to calculate the frequency of transitions and transversions. To estimate the effect of each mutation on the secondary structure, we assumed that nucleotides adjacent to the mutations did not shift from the proposed secondary structure suggested by Crease and Colbourne (1998). We used RNAdraw V1.1.b2 (Matzura and Wennborg 1996) to calculate the secondary structures of each unique expansion segment sequence that we identified.

In order to more closely examine the effect of indels on the stability of the secondary structure, we used RNAdraw V1.1.b2 (Matzura and Wennborg 1996) to calculate the minimum free energies (MFE) at 16°C of sequences having different combinations (haplotypes) of either the presence or absence of nucleotides at indel sites. We limited these calculations to the indels that occur in more than one unique

**Table 1**  
Nucleotide Frequencies of the Long Consensus Sequence of the Core Region and Expansion Segments 43/e1 + e4 in a 589-nt Fragment of the *Daphnia obtusa* 18S rRNA Gene

Region	Nucleotide Frequency				Total
	G	A	T	C	
Core	0.298	0.201	0.257	0.244	299
Expansion segments	0.355	0.121	0.303	0.221	290
Total					589

sequence. The temperature at which the calculations were done was chosen to approximate the environments in which *D. obtusa* is found. To ensure that base substitutions did not confound the analysis, we used the consensus sequence at all other sites. We used a one-tailed *t*-test, assuming unequal variances, to test the prediction that the MFE of haplotypes with compensated indel combinations (no bulges relative to the consensus sequence) would be significantly lower (i.e., have greater stability) than the MFE of haplotypes that have uncompensated indel combinations (bulges relative to the consensus sequence).

We calculated the level of linkage disequilibrium (D) between all pairs of indel variants that affect base pairing in the expansion segment relative to the consensus sequence and used a  $\chi^2$  test to determine if the values of D are significantly different from 0.

We used a  $\chi^2$  test to determine if the average nucleotide composition of the consensus sequence differs between the core and the expansion segment and if the nucleotide composition of unpaired regions differs from that of paired regions in the expansion segment. We used the program SIMPLE 3.1 (Albà, Laskowski, and Hancock 2002) to determine whether or not the consensus sequence contains repetitive nucleotide motifs. This program calculates the relative simplicity factor of a sequence, which is computed by comparing the number of repetitive motifs in the query sequence to a specified number of sequences (100 in this case) generated by randomly shuffling the nucleotides of the query sequence. We used a sliding window size of 15 and assigned a score of 1 to tetranucleotide repeats and 0 to all other types.

We performed an analysis of molecular variance (AMOVA) using Arlequin 2.000 (Schneider, Roessli, and Excoffier 2000) to partition length variation in the expansion segment into three hierarchical levels: (1) within individuals, (2) between individuals within a population (CI and MC only), and (3) between populations.

## Results

The sequences of the cloned fragments range in length from 579 to 589 nt. Of these, the core comprises 299 nt and the expansion segments, 43/e1 + e4, range from 280 to 290 nt. The construction of the sequence alignment was unambiguous due to the low level of sequence divergence among all haplotypes. If possible, we clustered single-nucleotide indels that are in very close proximity to one another so that they could be represented as a single mutational event. In such cases, we referred to the secondary structure to determine the location of these indels.

**Table 2**  
Frequency of Base-Paired and Unpaired Nucleotides in the Long Consensus Sequence of Expansion Segments 43/e1 + e4 of the *Daphnia obtusa* 18S rRNA Gene

Nucleotide State	Nucleotide Frequency				Total
	G	A	T	C	
Base paired	0.411	0.084	0.297	0.208	236
Unpaired	0.093	0.296	0.352	0.259	54
Total					290

One hundred and three unique sequences were identified among 2,588 cloned fragments that were screened from the six individuals (GenBank accession numbers AY887541–AY887643). The consensus sequence generated from these sequences is 585 nt long. One hundred and four nucleotide substitutions were identified among these unique sequences, of which 87 (84%) are transitions. When the error rate of the high fidelity *Taq* polymerase is taken into account, there is a very small probability that one of these substitutions is an artifact. We divided the substitutions into four categories based on comparisons of the secondary structures of each variant to that of the consensus sequence: (1) 27 substitutions (26%) maintain a bulge or a loop (2) 41 substitutions (39%) maintain base pairing, (3) 35 substitutions (34%) cause an internal loop (mismatch) in a base-paired region, and (4) 1 substitution (1%) decreases the size of a bulge or an internal loop by creating a new base pair. The transition from T to C occurred 2.6 times more often than the transition from C to T, whereas the transition from G to A only occurred 1.4 times as often as A to G.

Nucleotide composition of the consensus sequence is significantly different in the core and the expansion segments (table 1;  $\chi^2 = 8.79$ , *df* = 3, *P* = 0.032) due to an underrepresentation of adenine (and to a lesser extent cytosine) and a substantial increase in guanine and thymine in the expansion segments. The nucleotide composition of the consensus sequence with respect to base-pairing nucleotides and unpaired nucleotides in the expansion segments is also significantly different (table 2;  $\chi^2 = 29.5$ , *df* = 3, *P* < 0.005). In this case, there is an overrepresentation of adenine and an underrepresentation of guanine in the bulge and loop regions.

The alignment of the 103 unique sequences (available as supplementary data file 1) shows that 19/124 polymorphic nucleotides are indels. Four of the indels involve a single nucleotide, three involve dinucleotides, and three involve trinucleotides. All the indels, except for three single nucleotides in the core, occur in the expansion segment 43/e4. No adenine nucleotides are present at any indel site, whereas cytosine and guanine make up 84% of these nucleotides. Based on the secondary structure of the consensus sequence of expansion segment 43/e4 suggested by Crease and Colbourne (1998), two dinucleotide indels (denoted as indel sites 2 and 3) complement each other at the distal end of the helix (fig. 2). Two trinucleotide indels (denoted as indel sites 1 and 4) are also located across from each other in the helix. However, they are not perfect complements of each other and form a single-nucleotide bulge flanked on either side by a base pair. All the polymorphic nucleotides are always present

or absent in the indels at these four sites. For example, there are always either 3 or 0 nt at indel site 1. These four sites account for more than 95% of the length variants identified on the denaturing gels. All length variants involving these four sites were classified as either compensated (when nucleotides are present or absent at both sites that pair with each other in the secondary structure) or uncompensated (when nucleotides are present at one indel site but absent at the other site with which it would pair) (table 3). The average MFE of compensated length variants is significantly lower (more stable) than the average MFE of uncompensated combinations of length variants ( $t = 5.738$ ,  $df = 8$ ,  $P < 0.0002$ ).

We did the sliding window analysis of sequence simplicity using SIMPLE 3.1 on three sequences: (1) the strict consensus, which is 585 nt long and contains the insertions at indel sites 1 and 4 and the deletions at indel sites 2 and 3, (2) the “long” consensus sequence, which is 589 nt long and contains the insertions at all four indel sites, and (3) the “short” consensus sequence, which is 579 nt long and contains the deletions at all four indel sites. All three sequences have a significant relative simplicity factor (data not shown). However, only the long consensus sequence has a significant motif associated with it, TTGT, which overlaps both indel sites 1 and 2 (fig. 2).

Analysis of the 2,588 clones on denaturing gels showed that 12 of the 16 possible haplotypes involving the two pairs of complementary indels in segment 43/e4 were observed in the six individual *D. obtusa*, but not all haplotypes were found in all individuals (table 3). Haplotypes containing uncompensated indel combinations were only present at low frequencies (i.e., less than 5%), while compensated combinations were present at a wide range of frequencies. For example, haplotype [+ - - +] (where [+] indicates the presence and [-] indicates the absence of nucleotides at indel sites 1 to 4, respectively) was not found in CIb, while haplotype [+ + + +] occurred at a frequency of 93% in MCb.

Analysis of linkage disequilibrium among the four indel sites revealed that there is a very significant association between complementary sites in every individual, that is, sites 1 and 4 are in significant linkage disequilibrium with one another, as are sites 2 and 3 (table 4). In addition, in all but two individuals (CIa and DO), all the other pairwise combinations of indel sites also show significant linkage disequilibrium, although to a lesser extent than that between complementary sites (table 4). Three of the individuals are in complete linkage disequilibrium for almost every combination of indel sites (table 5). In the other three individuals,  $D/D_{\max}$  is higher between complementary indel sites (e.g., sites 1-2 and 3-4) than between noncomplementary indel sites (e.g., 1-2, 1-3, 2-4, 3-4), even though some of the noncomplementary sites are physically closer to each other in the linear sequence.

All pairwise  $\Phi_{ST}$  values, which represent the proportion of variation in indel haplotype frequencies that is due to differences between individuals, are significantly different from 0 (table 6). Interestingly, the  $\Phi_{ST}$  values do not show a consistent relationship with geographic distribution; indeed, both the largest (0.703) and the smallest (0.033) values are found between the two individuals from within a single population (MC and CI, respectively). The AMOVA

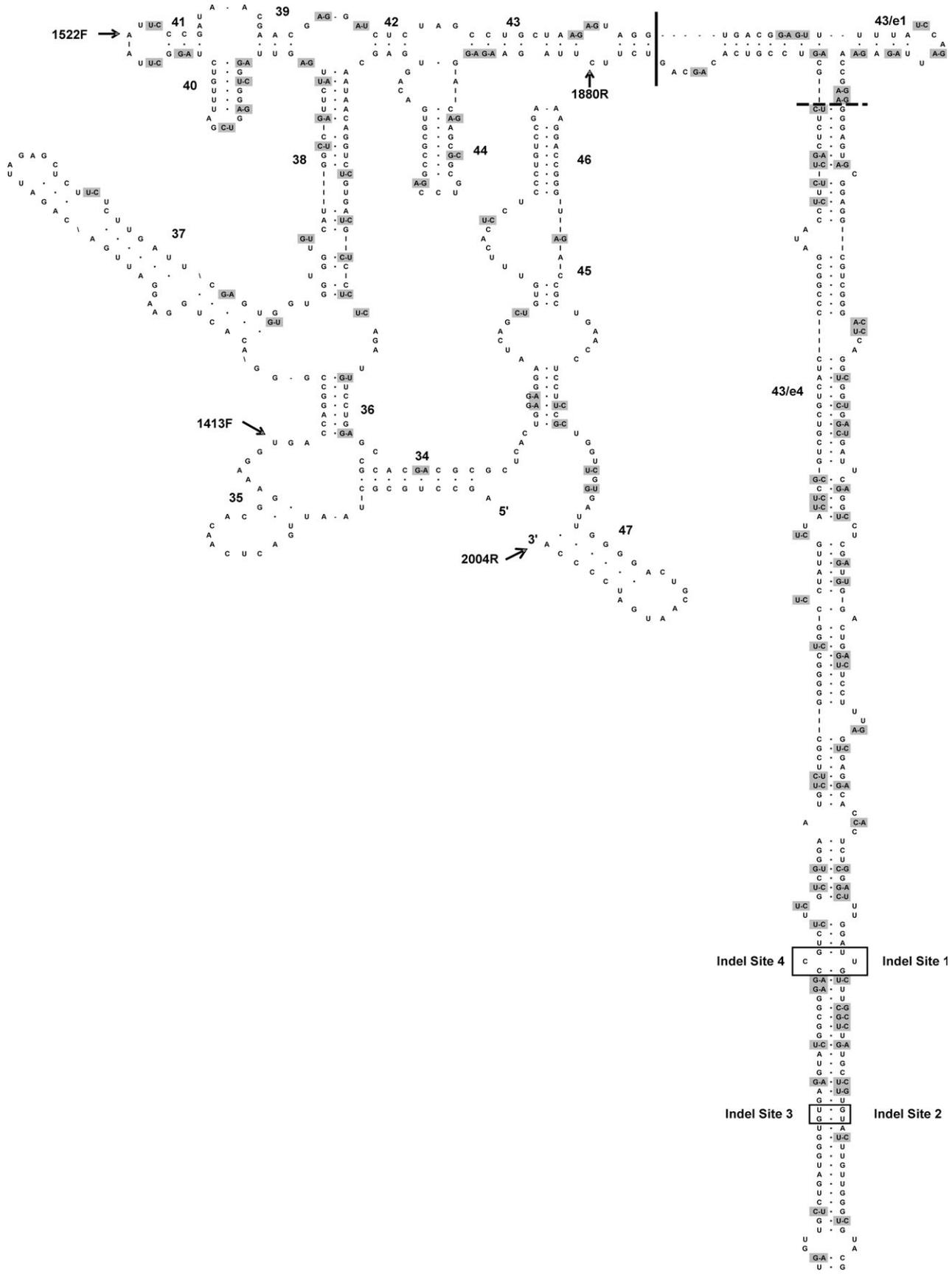
showed that all the variations in haplotype frequency occur within (59.2%) and between (42.3%) individuals and not between the different populations (table 7).

## Discussion

The data from this study support the hypothesis that selection acts to maintain the helical structure of expansion segment 43/e4 in *D. obtusa*. First, haplotypes with compensated indel sites are present at a wide range of frequencies (table 3) as expected for selectively neutral (or nearly neutral) variation. Second, all haplotypes with uncompensated indel sites are only present at low frequencies. The fact that the highest observed frequency of an uncompensated haplotype is only ~4% (table 3) is strong support for the hypothesis that they are selectively unfavorable. Even so, the presence of several types of uncompensated haplotypes suggests that selection is not able to eliminate them altogether, presumably due to the high copy number of rDNA. Moreover, their persistence could also be a function of the fact that they can be regenerated by recombination (both gene conversion and crossing-over) between different compensated haplotypes. Finally, the existence of strong linkage disequilibrium between complementary pairs of indels is strong evidence that selection is acting to reduce the frequency of uncompensated haplotypes.

The physical closeness of the two pairs of complementary indel sites makes it very likely that hitchhiking is a factor in this system, and it is probably responsible for the significant associations between noncomplementary indel sites. Indeed, when the physical proximity of the indel sites is taken into consideration, it is remarkable that such a large difference in linkage disequilibrium is sometimes observed between indel sites that do and do not pair with one another. For example, three of the individuals reached maximum linkage disequilibrium at five or six of the six pairs of indel sites (table 5), but the deviation from linkage equilibrium is still much higher for the complementary sites (table 4). Notably, most or all of the associations between unpaired indel sites are not significant in two individuals (CIa and DO), despite the fact that there is complete or nearly complete linkage disequilibrium between the paired sites. Given that sites 2 and 3 are located between sites 1 and 4, selection must be acting on the products of very localized recombination events, most probably the result of gene conversion.

It is important to emphasize that only *polymorphic* uncompensated bulges in segment 43/e4 appear to be selectively unfavorable. The rest of the unpaired nucleotides in this segment (fig. 2) are highly conserved and thus appear to be maintained by strong selection. In fact, contrary to neutral expectations, 26 of 27 nucleotide substitutions that occur at an unpaired nucleotide maintain the unpaired state. We suspect that these conserved unpaired nucleotides play a similar role to their counterparts in the core regions, such as involvement in the folding of the expansion segment into stable tertiary and quaternary conformations. Unpaired nucleotides have been particularly implicated in RNA-RNA interactions, although they may also play a role in RNA-protein interactions (e.g., Moore 1999; Lee, Cannone, and Gutell 2003). For example, ribosomal protein S9 binds to unpaired nucleotides in bacterial helix H39, which



**Table 3**  
**The Frequency of the 16 Possible Indel Haplotypes of Expansion Segment 43/e4 in Six Individual *Daphnia obtusa* from Four Populations in the United States**

Code <sup>a</sup>	Comp <sup>b</sup>	MFE <sup>c</sup>	Frequency						
			CO (N = 354) <sup>d</sup>	CIa (N = 467)	CIb (N = 466)	MCA (N = 426)	MCb (N = 408)	DO (N = 467)	
1	----	Y	-161.53	0.028	0.006	0.034	0.148	0.037	0.570
2	++++	Y	-164.58	0.726	0.383	0.515	0.028	0.926	0.009
3	-++-	Y	-163.92	0.031	0.529	0.376	0.045	0.037	0
4	+--+	Y	-162.19	0.147	0.015	0	0.737	0	0.373
5	+++-	N	-159.47	0	0.032	0.039	0	0	0
6	-+++	N	-159.47	0	0.034	0.036	0	0	0
7	+-+-	N	-152.89	0	0	0	0	0	0
8	+---	N	-157.08	0.014	0	0	0	0	0.024
9	--++	N	-157.08	0.017	0	0	0.002	0	0.024
10	--++	N	-153.00	0	0	0	0	0	0
11	-+--	N	-157.34	0	0	0	0	0	0
12	+--+	N	-158.11	0.017	0	0	0	0	0.002
13	++-+	N	-158.00	0.020	0	0	0	0	0
14	--+-	N	-157.45	0	0	0	0	0	0
15	+--+	N	-152.33	0	0	0	0.019	0	0
16	-+++	N	-152.89	0	0	0	0.021	0	0

<sup>a</sup> The code of each haplotype indicates if nucleotides are absent (-) or present (+) at each of the four complementary indel sites (fig. 2).

<sup>b</sup> Comp = compensated, Y = yes, N = no.

<sup>c</sup> MFE = minimum free energy in kilocalories per mole.

<sup>d</sup> N = number of clones scored.

corresponds to eukaryotic helix 43 (Brodersen et al. 2002). In this context, it is interesting that we found adenine nucleotides to be overrepresented in the unpaired regions of the expansion segment. McLain (2001) has suggested that this pattern may be indicative of selection acting to reduce the potential for unpaired regions to form hydrogen bonds in the event of a mutation because adenine only pairs with uracil in the rRNA molecule. Conversely, the high proportion of guanine in the expansion segment may reflect its ability to pair with both uracil and cytosine, which would increase the probability that a sequence could fold into a stable secondary structure. Gutell et al. (2000) also found that a high proportion of adenine nucleotides are unpaired and conserved in a large set of bacterial 16S and 23S rRNA genes and suggested that they play important roles in several different structural motifs.

The observation that 66% of the nucleotide substitutions found in this study result in a secondary structure that is as stable as, or more stable than, that of the consensus sequence is strong indirect evidence that expansion segment stability is a trait under selection. Indeed, there are several studies that demonstrate that expansion segments may play an important role in the function of the ribosome. For example, in vivo mutational analysis of expansion segment 6 in variable region 1 of the 18S rRNA of *Saccharomyces cerevisiae* showed that a relatively small increase in the size of the base-paired regions or the terminal loop

stopped the formation of the mature 18S rRNA molecule (van Nues et al. 1997). Furthermore, other studies have shown that mutations in different expansion segments of both the large- and small-subunit rRNAs can have remarkably different consequences. For example, the fitness costs in *Tetrahymena thermophila* range from none, to the inability of a mutant rRNA to support cell growth, to cell death (Sweeney, Chen, and Yao 1993, 1994). Clearly, we are currently aware of only some of the selective pressures impacting expansion segments, and much work remains to be done before the role that these regions play in ribosome function is elucidated.

The conclusion that selection acts to maintain or even to increase the stability of expansion segment secondary structure is not concordant with previous observations of uncompensated indels in rRNA genes, even though relatively few clones have been sequenced in other studies (Gonzalez et al. 1985; Holzmann, Piller, and Pawlowski 1996; García, Nicholson, and Nickrent 2004). Therefore, it appears that there might be fundamental differences in the selective pressures acting on these regions in different species. For example, in those species which have uncompensated indels at relatively high frequencies, either the selective pressure maximizing expansion segment stability is relaxed compared to *Daphnia* or the rate of concerted evolution/replication slippage is high enough to outpace the rate at which they are lost by selection. A more thorough

←

FIG. 2.—Putative secondary structure of the long consensus sequence of a portion of the 18S rRNA of *Daphnia obtusa* containing expansion segments 43/e1 and 43/e4. This structure is based on the one proposed by Crease and Colbourne (1998). Nucleotides at variable complementary indel sites in segment 43/e4 are boxed. Dots represent base pairing. Dashes connect adjacent nucleotides where necessary. Shaded nucleotides are polymorphic with the nucleotide state of the consensus sequence given first. Helix numbering follows Wuyts, Van de Peer, and De Wachter (2001). The 5'-end of the two forward primers and the 3'-end of the two reverse primers are indicated by arrows and the primer names. The boundary between the core and the expansion segments is indicated by a single vertical line with the expansion segments located to the right of the line. MFE of the secondary structure of expansion segment 43/e4 was calculated for the region below the dashed line. Nucleotides upstream of primer 1413F and the sequences corresponding to primers 1413F and 2004R were not sequenced in this study but were taken from the 18S rRNA gene sequence of the closely related species, *Daphnia pulex* (GenBank accession number AF014011).

**Table 4**  
Chi-Square Values<sup>a</sup> of the Deviation from Linkage Equilibrium for All Pairs of the Four Complementary Indel Sites in Expansion Segment 43/e4 from Six Individual *Daphnia obtusa* from Four Populations in the United States

Sites <sup>b</sup>	CO	CIa	CIb	MCA	MCb	DO
1-4	213.1	349.2	335.2	331.0	408.0	380.3
2-3	283.3	467.0	466.0	263.3	408.0	372.8
1-2	23.0	<b>3.0</b>	20.6	79.9	196.2	<b>5.9</b>
1-3	22.7	<b>3.0</b>	20.6	27.8	196.2	7.4
2-4	20.3	<b>3.0</b>	20.4	14.4	196.2	<b>5.9</b>
3-4	19.8	<b>3.0</b>	20.4	38.0	196.2	7.4

<sup>a</sup> Values in bold are not significant at the  $P = 0.01$  level.

<sup>b</sup> Indel sites are denoted 1, 2, 3, and 4 as in figure 2. Sites 1-4 and 2-3 are the complementary pairs.

examination of length variation in the expansion segments of other organisms will be required to address this issue.

The fact that three of the four compensated indel haplotypes of segment 43/e4 were isolated from individuals from all four populations indicates that there may be sufficient gene flow between the four populations to prevent their divergence from one other. Conversely, the fact that there are haplotypes unique to each population except for DO suggests that gene exchange between the other three populations may be limited. A recent study of mitochondrial DNA variation that included these *D. obtusa* populations (Penton, Hebert, and Crease 2004) shows that populations DO and MC belong to a different mitochondrial lineage than do populations CO and CI. Average sequence divergence between these mitochondrial lineages is 1.5%, suggesting that they diverged within the last Myr, likely as a consequence of isolation in different glacial refugia during the Pleistocene (Penton, Hebert, and Crease 2004). Even so, these lineages now show weak differentiation over broad geographic areas, which is consistent with recent long-range dispersal following retreat of the last glacial advance.

If we accept that the four populations were isolated from each other, at least during the Pleistocene glacial advances, then the shared indel haplotypes were either present in the common ancestor or they arose independently. Due to the fact that all the nucleotides at the indel sites are identical in both populations, it is most parsimonious to assume that these variants were present in the common ancestor. On the other hand, it is more difficult to infer

**Table 5**  
D/D<sub>max</sub> for All Pairs of the Four Complementary Indel Sites in Expansion Segment 43/e4 from Six Individual *Daphnia obtusa* from Four Populations in the United States

Sites <sup>a</sup>	CO	CIa	CIb	MCA	MCb	DO
1-4	0.760	0.861	0.844	1	1	0.902
2-3	0.887	1	1	1	1	1
1-2	0.476	0.471	1	0.673	1	1
1-3	0.474	0.471	1	0.494	1	1
2-4	0.455	0.471	1	0.251	1	1
3-4	0.453	0.471	1	0.521	1	1

<sup>a</sup> Indel sites are denoted 1, 2, 3, and 4 as in figure 2. Sites 1-4 and 2-3 are the complementary pairs.

**Table 6**  
Pairwise  $\Phi_{ST}$  Values Based on the Indel Haplotype Frequencies (table 3) of Expansion Segment 43/e4 in Six Individual *Daphnia obtusa* from Four Populations in the United States

	CO	CIa	CIb	MCA	MCb
CO	—				
CIa	0.2582 <sup>a</sup>	—			
CIb	0.1409	<b>0.0328<sup>b</sup></b>	—		
MCA	0.4764	0.4722	0.4678	—	
MCb	0.1053	0.4212	0.2727	<b>0.7025<sup>b</sup></b>	—
DO	0.4489	0.4391	0.4229	0.2445	0.6436

<sup>a</sup> All values are significantly different from 0 ( $P \ll 0.001$ ).

<sup>b</sup> The highest and the lowest values, shown in bold, are between individuals from the same population.

whether haplotypes that are unique to a particular population arose de novo or if they were lost in the other populations due to selection and/or drift because, as mentioned above, recombination could regenerate these haplotypes.

There is no consistent pattern in the frequency of the four compensated indel haplotypes among individuals, which suggests that random genetic drift may have a larger impact on their frequencies than selection. For example, if the relative fitness of the compensated haplotypes is based on the stability of their secondary structures, then we would expect their relative frequencies to be  $[- - - -] < [+ - - -] < [- + + -] < [+ + + +]$ , which is very different from the observed pattern. Furthermore, neither the absolute stability of the secondary structure (table 3) nor the length of the sequence can explain the frequency distribution of these haplotypes. Thus, the inconsistent pattern among the six individuals suggests that fluctuations in compensated haplotype frequencies are more strongly influenced by drift than by selection. If so, then we predict that the frequency of compensated haplotypes will fluctuate randomly over time in *D. obtusa* populations. However, if selection on these haplotypes is based on their energetic stability or some other parameter, then we would expect to see their frequencies change in a consistent direction (i.e., the longest length variant  $[+ + + +]$  would always increase over time). We are testing this hypothesis by screening changes in the relative frequency of length

**Table 7**  
AMOVA Based on the Indel Haplotype Frequencies (table 3) of Expansion Segment 43/e4 in Six Individual *Daphnia obtusa* from Four Populations in the United States

Level of Variation	df	Variance Components		P Value <sup>a</sup>	Fixation Indices <sup>b</sup>
		Absolute	%		
Between populations	3	-0.0058	-1.48	NS <sup>c</sup>	$F_{CT} = -0.0148$
Within populations (CI, MC)	2	0.1660	42.25	$\ll 0.001$	$F_{SC} = 0.4164$
Within individuals	2,593	0.2326	59.22	$\ll 0.001$	

<sup>a</sup> Levels of significance are based on 1,000 iteration steps.

<sup>b</sup>  $F_{SC}$  = differentiation between individuals within a population (CI and MC), and  $F_{CT}$  = differentiation between populations.

<sup>c</sup> NS = not significant.

variants in expansion segment 43/e4 in multiple lines that were initiated from the parthenogenetically produced offspring of a single female *D. obtusa* and maintained parthenogenetically for over 90 generations.

The AMOVA results indicate that the majority of the indel haplotype frequency variation occurs within an individual, followed closely by the variation between individuals (table 7). The  $\Phi_{ST}$  value between the two individuals from MC is very large, indicating that the frequency of indel haplotypes within individuals is not necessarily representative of frequencies in the population as a whole. However, these results must be interpreted with caution as our sampling strategy is heavily biased toward finding variation within individuals, as opposed to within or between populations. Despite the potential bias in our sampling regimen, these results are comparable to those of a study of rDNA variation surveyed in five populations of *D. pulex* (Crease and Lynch 1991), where 37 unique rDNA repeat types were identified using restriction-site mapping. While the number of repeat types per population varies from 5 to 17 (mean = 10.4), the number of haplotypes per individual averages only 2.6, indicating that there is substantial variation among individuals not only with respect to the relative frequency of repeat types but also with respect to the complement of repeat types that are present. Crease and Lynch (1991) estimated differentiation among individuals within populations to be 0.34, while that among populations is much lower at 0.06, which is similar to the results of our study with values of 0.42 and -0.01, respectively (table 7). Many repeat types were found in only one population; however, the high variance in frequency of the four most common haplotypes among individuals within populations overshadowed the variation among populations, as was also the case in our study.

A more recent study that examined sequence variation in the ITS1 region of rDNA in the meadow grasshopper, *Chorthippus parallelus* (Parkin and Butlin 2004), also found that individuals within populations typically share the same haplotypes, but their frequencies vary significantly among individuals. This high level of interindividual variation is concordant with previous work showing that intrachromosomal exchange is more common than interchromosomal exchange in rDNA (Arnheim et al. 1982; Seperack, Slatkin, and Arnheim 1988; Crease and Lynch 1991; Schlötterer and Tautz 1994; Crease 1995; Petes and Pukkila 1995) and thus causes variants to spread through arrays faster than the chromosomes carrying the arrays spread through populations.

It is well documented that slippage events during DNA replication are more likely to occur in regions of repetitive nucleotide composition (Levinson and Gutman 1987). Our analysis of sequence simplicity of the *D. obtusa* expansion segment revealed one significantly repetitive motif in the long consensus sequence that overlaps indel sites 1 and 2. The observation that repetitive motifs occur at these sites is strong evidence that replication slippage is responsible for the changes in length in this region. It is interesting that only two of the four indel sites are associated with a repetitive motif, as the compensatory slippage model of Hancock and Dover (1990) suggests that the second compensatory mutation event is the result of replication slippage, like the first event. However, our data do not support this model, which suggests that alternatives should be considered. For exam-

ple, it has been documented that palindromic or quasipalindromic DNA can form secondary structures during replication, which may cause them to incorrectly serve as the template for subsequent strand synthesis (Fieldhouse and Golding 1991). As these structures are sequence dependent and short lived, the mutation hotspot that they create may be extremely localized. DNA foldback (Wang and Ripley 1994) and DNA strand switching (Rosche, Trinh, and Sinden 1997) are two models that explain this type of complex mutation. The expansion segment is an ideal candidate for this type of mutational change because the entire sequence is a quasipalindrome, and at a smaller scale, its high degree of sequence simplicity gives it the ability to form many internal secondary structures (e.g., hairpins) during DNA replication. Thus, we think that it is more probable that this type of mechanism is responsible for the generation of compensatory change in rDNA for two reasons. First, the probability of occurrence of two independent slippage events that compensate each other is very low, yet it is frequently observed, while uncompensated changes are very rare. Second, this type of mechanism explains the fact that the compensated sites that we observed are in linkage disequilibrium. If the presence of one mutation, perhaps caused by slippage, is directly involved in the occurrence of the second mutation, then they would be in perfect linkage disequilibrium initially. This association could be subsequently broken down by recombination and/or maintained by selection. The fact that the *D. obtusa* repetitive motifs were only identified when nucleotides are present at the indel sites indicates that the shorter versions of segment 43/e4 are less prone to slippage events, and as a consequence, this segment may be undergoing a reduction in size.

It has been postulated that the rate of compensatory mutation will be inversely proportional to the distance between two indel sites along a linear molecule (Stephan 1996). Coupled with the occurrence of repetitive motifs, this phenomenon could explain why we only see compensatory mutations in the distal portion of segment 43/e4, where the indel sites are separated by only 72–76 nt. This result is similar to those of McLain (2001), who found that indels are more common toward the distal end of expansion segment D3 in the 28S rRNA gene of ticks, and to those of Crease and Taylor (1998), who found that sequence alignment among closely related species of cladoceran crustaceans became more difficult in the distal part of expansion segments. This phenomenon could also explain why we did not see indels at an appreciable frequency anywhere else in the *D. obtusa* expansion segment; uncompensated indels will remain at low frequency due to selection where they are susceptible to loss by drift, which becomes more likely as the rate of compensatory mutation events decreases.

It may seem counterintuitive that intraindividual variation exists in rDNA, given that this gene family is known to undergo the homogenizing effects of concerted evolution. It is possible that this process differs between species that have intraindividual variation compared to those that do not. For example, a close relative of *D. obtusa*, *D. pulex*, does not appear to exhibit length variation in expansion segment 43/e4 (Crease and Taylor 1998). The difference between the two species could be due to a much lower rate of concerted evolution in *D. obtusa* than in *D. pulex* or a

higher rate of indel mutations in *D. obtusa* than in *D. pulex*. If the rate of concerted evolution is lower in *D. obtusa*, then the average sequence divergence among haplotypes, based on nucleotide substitutions, should be higher in *D. obtusa* than in *D. pulex*. Conversely, if the different levels of intra-individual length variation are caused by an increase in the rate of indel mutations, then the average sequence divergence between haplotypes based on nucleotide substitutions should be the same in both species, all else being equal. We are currently investigating these possibilities.

### Supplementary Material

Both sequential FASTA and interleaved Nexus files of the sequences generated in this study are available at *Molecular Biology and Evolution* online ([www.mbe.oupjournals.org](http://www.mbe.oupjournals.org)). Supplementary Data File 1: Nucleotide FASTA file of 103 partial 18S rRNA gene sequences from *Daphnia obtusa* and interleaved alignment file of 103 partial 18S rRNA gene sequences from *D. obtusa*.

### Acknowledgment

This study was funded by a grant to T.J.C. from the Natural Sciences and Engineering Research Council of Canada and an Ontario Graduate Scholarship in Science and Technology to S.J.M. We thank Angela Holliss and Janet Horseman for sequencing of the plasmid clones, Paul Hebert for kindly providing the *Daphnia* samples, Karim Gharbi for providing the SSCP conditions, Sara Northrup for her excellent laboratory assistance, Michael Lynch for insightful discussion of the data, and Brian Golding and an anonymous reviewer for helpful comments on an earlier draft of the manuscript.

### Literature Cited

- Ahn, S., B. Baek, T. Oh, C. Song, and B. Chatterjee. 2000. Rapid mini-scale plasmid isolation for DNA sequencing and restriction mapping. *BioTechniques* **29**:466–468.
- Albà, M., R. Laskowski, and J. Hancock. 2002. Detecting cryptically simple protein sequences using the SIMPLE algorithm. *Bioinformatics* **18**:672–678.
- Arnheim, N., D. Treco, B. Taylor, and E. Eicher. 1982. Distribution of ribosomal gene length variants among mouse chromosomes. *Proc. Natl. Acad. Sci. USA* **79**:4677–4680.
- Brodersen, D., W. Clemons Jr., A. Carter, B. Wimberly, and V. Ramakrishnan. 2002. Crystal structure of the 30S ribosomal subunit from *Thermus thermophilus*: structure of the proteins and their interactions with 16S RNA. *J. Mol. Biol.* **316**:725–768.
- Crease, T. 1995. Ribosomal DNA evolution at the population level: nucleotide variation in intergenic spacer arrays of *Daphnia pulex*. *Genetics* **141**:1327–1337.
- Crease, T., and J. Colbourne. 1998. The unusually long small-subunit ribosomal RNA of the crustacean, *Daphnia pulex*: sequence and predicted secondary structure. *J. Mol. Evol.* **46**:307–313.
- Crease, T., and M. Lynch. 1991. Ribosomal DNA variation in *Daphnia pulex*. *Mol. Biol. Evol.* **8**:620–640.
- Crease, T., and D. Taylor. 1998. The origin and evolution of variable-region helices in V4 and V7 of the small-subunit ribosomal RNA of branchiopod crustaceans. *Mol. Biol. Evol.* **15**:1430–1446.
- Dover, G. 1982. Molecular drive: a cohesive mode of species evolution. *Nature* **299**:111–117.
- Fieldhouse, D., and B. Golding. 1991. A source of small repeats in genomic DNA. *Genetics* **129**:563–572.
- García, M., E. Nicholson, and D. Nickrent. 2004. Extensive intra-individual variation in plastid rDNA sequences from the holoparasite *Cynomorium coccineum* (Cynomoriaceae). *J. Mol. Evol.* **58**:322–332.
- Gonzalez, I., J. Gorski, T. Campen, D. Dorney, J. Erickson, J. Sylvester, and R. Schmickel. 1985. Variation among human 28S ribosomal RNA genes. *Proc. Natl. Acad. Sci. USA* **82**:7666–7670.
- Gonzalez, I., and J. Sylvester. 2001. Human rDNA: evolutionary patterns within the genes and tandem arrays derived from multiple chromosomes. *Genomics* **73**:255–263.
- Gutell, R., J. Cannone, Z. Shang, Y. Du, and M. Serra. 2000. A story: unpaired adenosine bases in ribosomal RNAs. *J. Mol. Biol.* **304**:335–354.
- Hancock, J., and G. Dover. 1990. “Compensatory slippage” in the evolution of ribosomal RNA genes. *Nucleic Acids Res.* **18**:5949–5954.
- Hancock, J., D. Tautz, and G. Dover. 1988. Evolution of the secondary structures and compensatory mutations of the ribosomal RNAs of *Drosophila melanogaster*. *Mol. Biol. Evol.* **5**:393–414.
- Hartl, D., and A. Clark. 1989. *Principle of population genetics*. Sinauer, Sunderland, Mass.
- Hinkle, G., D. Leipe, T. Nerad, and M. Sogin. 1994. The unusually long small subunit ribosomal RNA of *Phreatamoeba balamuthi*. *Nucleic Acids Res.* **22**:465–469.
- Holzmann, M., W. Piller, and J. Pawlowski. 1996. Sequence variations in the large-subunit ribosomal RNA gene of *Ammonia* (Foraminifera, Protozoa) and their evolutionary implications. *J. Mol. Evol.* **43**:145–151.
- Juan, V., and C. Wilson. 1999. RNA secondary structure prediction based on free energy and phylogenetic analysis. *J. Mol. Biol.* **289**:935–947.
- Lee, J., J. Cannone, and R. Gutell. 2003. The lonepair triloop: a new motif in RNA structure. *J. Mol. Biol.* **325**:55–83.
- Leffers, H., and A. Anderson. 1993. The sequence of 28S ribosomal RNA varies within and between human cell lines. *Nucleic Acids Res.* **21**:1449–1455.
- Levinson, G., and G. Gutman. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**:203–221.
- Matzura, O., and A. Wennborg. 1996. RNAdraw: an integrated program for RNA secondary structure calculation and analysis under 32-bit Microsoft Windows. *Comput. Appl. Biosci.* **12**:247–249.
- McLain, D. 2001. Evolution of transcript structure and base composition of rDNA expansion segment D3 in ticks. *Heredity* **87**:544–557.
- Moore, P. 1999. Structural motifs in RNA. *Annu. Rev. Biochem.* **68**:287–300.
- Parkin, E. J., and R. K. Butlin. 2004. Within- and between-individual sequence variation among ITS1 copies in the meadow grasshopper *Chorthippus parallelus* indicates frequent intra-chromosomal gene conversion. *Mol. Biol. Evol.* **21**:1595–1601.
- Penton, E. H., P. D. Hebert, and T. J. Crease. 2004. Mitochondrial DNA variation in North American populations of *Daphnia obtusa*: continentalism or cryptic endemism? *Mol. Ecol.* **13**:97–107.
- Petes, T., and P. Pukkila. 1995. Meiotic sister chromatid recombination. *Adv. Genet.* **33**:41–62.
- Rosche, W., T. Trinh, and R. Sinden. 1997. Leading strand specific spontaneous mutation corrects a quasipalindrome by an

- intermolecular strand switch mechanism. *J. Mol. Biol.* **269**:176–187.
- Ruiz Linares, A., J. Hancock, and G. Dover. 1991. Secondary structure constraints on the evolution of *Drosophila* 28S ribosomal RNA expansion segments. *J. Mol. Biol.* **219**:381–390.
- Schlötterer, C., and D. Tautz. 1994. Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. *Curr. Biol.* **4**:777–783.
- Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin ver. 2.000: a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland.
- Seperack, P., M. Slatkin, and N. Arnheim. 1988. Linkage disequilibrium in human ribosomal genes: implications for multigene family evolution. *Genetics* **119**:943–949.
- Stephan, W. 1996. The rate of compensatory mutation. *Genetics* **144**:419–426.
- Sweeney, R., L. Chen, and M.-C. Yao. 1993. Phenotypic effects of targeted mutations in the small subunit rRNA gene of *Tetrahymena thermophila*. *Mol. Cell. Biol.* **13**:4814–4825.
- . 1994. An rRNA variable region has an evolutionarily conserved essential role despite sequence divergence. *Mol. Cell. Biol.* **14**:4203–4215.
- van Nues, R., J. Venema, R. Planta, and H. Raué. 1997. Variable region V1 of *Saccharomyces cerevisiae* 18S rRNA participates in biogenesis and function of the small ribosomal subunit. *Chromosoma* **105**:523–531.
- Wang, F. J., and L. S. Ripley. 1994. DNA sequence effects on single base deletions arising during DNA polymerization *in vitro* by *Escherichia coli* Klenow fragment polymerase. *Genetics* **136**:709–719.
- Wuyts, J., Y. Van de Peer, and R. De Wachter. 2001. Distribution of substitution rates and location of insertion sites in the tertiary structure of ribosomal RNA. *Nucleic Acids Res.* **29**:5017–5028.

Diethard Tautz, Associate Editor

Accepted February 21, 2005