J Mol Evol (2009) 69:142–149
DOI 10.1007/s00239-009-9257-y

Length Variation in 18S rRNA Expansion Segment 43/e4 of *Daphnia obtusa*: Ancient or Recurring Polymorphism?

Seanna J. McTaggart · Teresa J. Crease

Received: 13 March 2009 / Accepted: 2 June 2009 / Published online: 7 July 2009

Abstract Expansion segments in ribosomal DNA (rDNA) can show length variation at the level of the individual, yet our understanding of the evolutionary forces shaping this variation is incomplete. Previous studies of expansion segment 43/e4 of the 18S rRNA gene in *Daphnia obtusa* have examined this variation in six individuals; however, it is not known if the variation documented at this locus is representative of variation across the species' geographic range. Furthermore, it is unclear whether length variants found in multiple individuals share common ancestry, or were generated de novo through recombination. We quantified expansion segment length variant frequencies in 134 individual *D. obtusa* from 33 populations at 15 sites across the species range in the US, and used a phylogeographic approach to determine whether recombination continues to add to the standing crop of variation at this locus. We identified seven length variants across the sampling range, which spans almost 3000 km. Based on the phylogeographic distribution of length variants in the expansion segment, we conclude that they are shared ancient polymorphisms that have persisted despite the operation of molecular mechanisms that cause the concerted evolution of multigene families such as rDNA.

Keywords rDNA · Expansion segment · Daphnia · Concerted evolution · Phylogeophraphy · Multigene family

Electronic supplementary material The online version of this article (doi:10.1007/s00239-009-9257-y) contains supplementary material, which is available to authorized users.

S. J. McTaggart (✉)
Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK
e-mail: smctagga@staffmail.ed.ac.uk

S. J. McTaggart · T. J. Crease
Department of Integrative Biology, University of Guelph, Guelph N1G 2W1, Canada

Introduction

Ribosomal DNA (rDNA) is a large, multigene family encoding ribosomal RNA (rRNA), which, when coupled with proteins, comprises the ribosome. Hundreds to thousands of gene copies lie in tandem along one or more chromosomes, and generally show high sequence similarity since they evolve in concert through unequal crossing over and gene conversion (reviewed in Eickbush and Eickbush 2007). There are four rRNA molecules in most eukaryotic ribosomes; the 28S in the large subunit, and the 18S, 5.8S, and 5S in the small subunit. They fold into stable secondary structures that are necessary for normal ribosome function (Wuyts et al. 2001). While the sequence and overall secondary structure of the 18S and 28S rRNAs are generally very conserved across taxa, certain regions called expansion segments are highly variable in length and nucleotide sequence, both within and between species, and even within individuals (Crease and Taylor 1998).

One such variable region in the 18S rRNA, expansion segment 43/e4, has been previously characterized within six individual *Daphnia obtusa* from four populations in North America, and shown to be length variable within and among individuals. Generally, this expansion segment ranges from 222 to 236 nucleotides in length (McTaggart and Crease 2005) and folds into a stable helical structure. The majority of the length variation is restricted to six short regions in the nucleotide sequence that pair with one another in this helix (Fig. 1). Furthermore, it appears that the presence/absence of nucleotides at length variable sites is not random as there are often substitutions at two

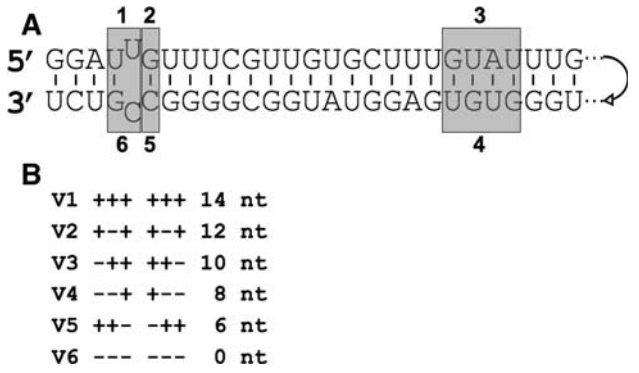

Fig. 1 Length variation in expansion segment 43/e4 of the 18S rRNA gene of *Daphnia obtusa*. **a** Sequence of the length-variable region of the expansion segment. The locations of six polymorphic indels are indicated with gray boxes, and numbered from 5′–3′ in the primary sequence. Vertical lines represent base pairing and unpaired nucleotides are offset from the helical structure. **b** Compensated indel genotype (see text) of each variant and its length relative to V6 (all indel nucleotides absent). "+" = indel nucleotides present. "−" = indel nucleotides absent

positions in the sequence that maintain the integrity of the helical structure (McTaggart and Crease 2005). We refer to this type of length variant as 'compensated', while length variants resulting in energetically destabilizing bulges in the secondary structure are called 'uncompensated'.

In a previous study, we (McTaggart and Crease 2005) found that compensated variants occur at a wide range of frequencies but uncompensated variants are always rare. We proposed that the former are selectively neutral with respect to one another, but the latter do not reach high frequency within an individual, or a population, because selection acts against them. Additionally, we found the four most common compensated length variants were shared among populations that are thousands of kilometers apart. Based on this initial analysis of only six individuals, we were unable to determine whether these shared expansion segment length variants are old and have 'escaped' the mechanisms of concerted evolution, or if they have been generated de novo by crossing over and gene conversion after the populations diverged from one another. The goal of this study is to distinguish between these two hypotheses by conducting a broad geographic survey of intraindividual length variation in expansion segment 43/e4. If the observed rDNA length variants represent old, neutral variation, we predict that their frequency distributions will show a phylogeographical pattern. On the other hand, if recombination continually generates the variation, then no phylogeographic pattern will be observed.

Several investigators have already examined phylogeographical patterns of genetic diversity in *D. obtusa* at a continental scale in order to better understand the evolutionary history of the genus *Daphnia* (Hebert and Finston 1996; Penton et al. 2004). These studies revealed that patterns of genetic diversity are not concordant between markers; allozyme data revealed groupings associated with ecological differentiation, presumed to be the result of selection, whereas mtDNA phylogroups appear to be a consequence of dispersal patterns from different glacial refugia (Penton et al. 2004). We compare the phylogeographic patterns observed with rDNA expansion segment 43/e4 with both the mtDNA and allozyme variation.

Methods

Sampling

We analyzed *D. obtusa* individuals from 33 ponds in 15 different sites across the majority of the species range in North America. Samples were collected by Hebert and Finston (1996) and flash frozen in liquid nitrogen in the field, after which they were stored at −80°C until processed for molecular analyses. In general, temporary ponds were sampled in the spring soon after re-establishment from resting eggs, which are produced sexually.

Whole animals were used for the allozyme analyses, so DNA was isolated from other individuals from the original collections (Penton et al. 2004), and again for this study. Thus, the results from all three studies represent a single time point but different individuals were analyzed for each genetic marker and we were not able to obtain individuals from all of the ponds that were analyzed in the previous two studies. For example, individuals from only one site in allozyme group II (Hebert and Finston 1996) were available for our study. We analyzed from 1 to 6 ponds per site and between 2 and 7 individuals per pond (Table 1). We assessed length variant frequency of expansion segment 43/e4 in a total of 134 individuals, which are denoted by a site code, pond number, and individual number. For example, RR1-5 is individual 5 from pond 1 in the Red Rock site. Penton et al. (2004) discovered that *D. obtusa* populations in North America belong to two very divergent mtDNA lineages, North America1 and North America2, which likely deserve recognition as separate species. All the populations included in this study belong to lineage 1.

Molecular Analysis

DNA extractions were performed using the GenElute mammalian Genomic DNA Miniprep kit (Sigma). Expansion segment 43/e4 was amplified from genomic DNA extracts using primers 18S1522F (Hex-ATTCCGATAAC GAACGAG) and 18S1880R (GAAGACTGCGTGACG-GAC) (both of which flank the expansion segment in the conserved core regions) in a 10 µl reaction, using the conditions described in McTaggart and Crease (2005).

Table 1 *Daphnia obtusa* samples examined in this study

Site location	State	Site code	Number of ponds	Number of individuals	Allozyme group[a]	MtDNA lineage[b]
Ada	Oklahoma	AD	3	10	IB	1B
Bellwood	Louisiana	BL	6	25	IB	1B
Big Wells	Texas	BW	3	12	IB	1B
Buffalo	South Dakota	BU	3	16	IA	1A
Canyonlands	Utah	CL	1	7	III	1A
Cisco	Texas	CI	2	9	IB	1B
Coconino	Arizona	CO	1	5	IA	1B
Dodson	Montana	DO	1	4	IA	1A
Miles City	Montana	MC	3	13	IA	1A
Natchez	Tennessee	NA	1	2	–	1C
Princeton	Arkansas	PR	1	7	II	1A,1B,1C
Red Rock	Nevada	RR	2	9	III	1B
Springfield	Missouri	SP	3	9	IB	1A,1B,1C
Oswega	Kansas	OS	1	2	1B	1C
Thunder Basin	Wyoming	TB	2	4	IA	1A
Total			33	134		

[a] Based on Hebert and Finston 1996. We recognize two subclades (A and B) within group I based on their UPGMA tree

[b] Based on Penton et al. 2004

Fifty microliter of formamide (Unison Biotek) and bromophenol blue were then added to each reaction. A size standard was made by amplifying and pooling the amplicons from six plasmid clones (McTaggart and Crease 2005), each containing an insert of one of the six most commonly found length variants of expansion segment 43/e4 in *D. obtusa*, which are longer than the shortest variant by 2–14 nt. The size standard was diluted with formamide and bromophenol blue, as were the other samples. Two microliter of each sample and the size standard were loaded onto a 6% acrylamide denaturing gel (37.5:1 acrylamide:bis-acrylamide) and electrophoresed for 4 h in 1× TBE buffer at a constant 35 W. All bands were visualized with an Hitachi FMBIO II scanner on channel 2 (585 nm), and marked by eye. The relative intensity of each fragment, which has been shown to be a good approximation of the relative frequency of each length variant (McTaggart et al. 2007), was calculated using the 1D-Gel Analysis tool in the FMBIO Analytical software, version 8.0.

We identified a novel length variant in one individual from a BL population and one individual from each of two BU populations. In order to determine its sequence, we cloned the amplified expansion segment from two individuals (BL3-7 and BU6-4) using the TOPO TA Cloning kit for Sequencing (Invitrogen) according to the manufacturer's instructions. Inserts were prepared for sequencing by colony amplification (McTaggart and Crease 2005) using the M13F and M13R vector primers, and then sequenced in one direction with the 18S1522 primer and the Big Dye Terminator v3.1 sequencing reaction mix (Applied Biosystems).

Data Analysis

Variant frequencies were averaged across all individuals within a site and used to calculate Nei's genetic distance (*D*) between sites with the Gendist program in Phylip 3.86 (Felsenstein 2004). Some pairs of sites share no length variants with one another, which results in an undefined value in the calculation of Nei's *D*. Thus, we added a second locus to each site with the same allele frequency in each (0.99 and 0.01). We conducted an analysis of molecular variance (AMOVA) using the software package Arlequin 3.1 (Excoffier et al. 2005) on each geographic site at which two or more ponds had been sampled. The input data consist of the length variant frequencies for each individual. We also performed a subsequent AMOVA on the 15 sites by pooling individuals within sites. In addition, we used Arlequin 3.1 to calculate the average number of nucleotide differences per nucleotide site between sampling sites in the mitochondrial *cytochrome* c *oxidase* I (COI) gene by pooling the sequences generated by Penton et al. (2004) from all individuals analyzed at a particular site. The number of sequences available for each site ranges from 1 to 6. The original Nei's *D* matrix based on allozyme variation was not available to us, so we generated

a distance matrix using branch lengths from the UPGMA tree in Hebert and Finston (1996). Although the distances in this matrix are distorted by averaging, the relative differences between populations are preserved.

In order to visualize the matrices of genetic variation, and of geographic distance between sites, we constructed Neighbor-joining trees using Phylip 3.86 (Felsenstein 2004) and then used MEGA4 (Tamura et al. 2007) to draw them. In addition, Mantel tests were carried out in Arlequin 3.1 (Excoffier et al. 2005) to determine whether or not there is a significant correlation between the genetic and geographic distance matrices, and between pairs of genetic distance matrices.

Results

Seven length variants were resolved on the acrylamide gels and all of them can be described in terms of the presence (+) or absence (−) of nucleotides at the six indel locations numbered from 5′ to 3′ in the primary sequence (Fig. 1). Indels 1 and 6 are each 2 nt long, indels 2 and 5 are each a single nucleotide, while 3 and 4 are each 4 nt. When the secondary structure is formed, nucleotides at indel 1 pair with nucleotides at indel 6, 2 with 5, and 3 with 4. For example, length variant 1 (V1), which has nucleotides at all indel locations, can be described as +++ +++, whereas V3, which lacks nucleotides at indel positions 1 and 6, is written as −++ ++−. Extensive sequencing in a previous study (McTaggart and Crease 2005) showed that the vast majority of variants (>95%) either have nucleotides present or absent at both complementary indel locations. As previously mentioned, these type of variants are described as compensated. All of the length variants except V1 (all indel nucleotides present) and V6 (all indel nucleotides absent) can represent more than one sequence including one compensated and one or more uncompensated variants. For example, the compensated length variant V5 (++− −++ which is 6 nt [2 + 1 + 0 + 0 + 1 + 2] longer than V6) is the same length as the uncompensated length variant (−− + −−+ [0 + 0 + 4 + 0 + 0 + 2]). In contrast, the lengths of all compensated variants are different from one another. Because alleles at low frequency (i.e., all uncompensated length variants) will not qualitatively change the outcome of our analyses, we considered each band on our gels to represent only one compensated length variant.

Six of the seven length variants we observed were reported previously by McTaggart and Crease (2005). The novel variant, V7 (225 nt) is only present in three individuals (BL3-7, BU1-6, and BU3-4) at frequencies of 6, 3 and 4%, respectively (Table S1). We were not successful in cloning it and so we do not know if it is compensated or not. However, since it is only found at low frequencies, we suspect that it is uncompensated. Because we could not sequence this length variant, it is not included in Fig. 1, but it was included in all statistical analyses, which are based solely on frequencies. Because it represents less than 1% of the total length variation in either the BL or the BU sites in which it is found, we also omitted this variant from Fig. 2. In addition, we found a V2 sequence that was not observed by McTaggart and Crease (2005). Prior to this study, all observed sequences of V2 were uncompensated, and this variant was only ever observed at low frequency. However, because we found V2 at frequencies over 10% in four of the seven individuals from BL3, and one individual from PR (Table S1), we cloned length variants from two of these individuals; BL3-4 (62%) and BL3-5 (44%) to determine the sequence of their V2 variant. Sequencing re-revealed a novel indel site polymorphism, which results in a new compensated version of V2 (Fig. 1).

Twenty-two of the 134 individuals we analyzed possess a single length variant (15 individuals with V1, 1 individual with V4, and 6 individuals with V6), which includes all individuals from three ponds with sample sizes of 6, 3, and 5 individuals (Table S1). The mean number of variants per individual is 2.6 (range = 1–6). A comparison of mean length variant frequencies shows that overall, V7 (mean = 0.1%) and V2 (mean = 2.0%) occur at very low frequency but V2 does occur at high frequency in some individuals (e.g., 62% in individual BL3-4). Indeed, every variant except V7 was observed in at least one individual at a frequency of over 50%. V1 is most common with a mean frequency of 33.2% while the mean frequencies of V3, V4, and V5 are about half of that value (17.1, 19.6, and 17.3%). The mean frequency of V6 is 10.7%.

The AMOVA conducted on ponds within a site indicates that population subdivision is only significant at two of the sites: BL (6 ponds) and BW (3 ponds) (Table 2). In contrast, mean variant frequencies differ substantially among sites (Fig. 2). The AMOVA indicates that the majority of the genetic variance (53%) occurs within individuals, but a substantial portion, 33%, also occurs among sites. The remaining 14% occurs among individuals within collection sites, and all three values are highly significant (Table 3).

There are substantial differences among the topologies of NJ trees based on the three genetic markers and geographic distance. However, three groups of sites (BU + DO + MC, CI + BL, and BW + CO) do belong to the same major cluster in all three trees based on genetic variation (Fig. 3). While sites in the BU + DO + MC and CI + BL groups are in relatively close geographic proximity, the BW and CO sites are not (Figs. 2, 3).

The Mantel tests show a highly significant positive correlation between genetic distance (based on UPGMA branch lengths) and geographic distance among sites for both rDNA and allozyme variation (Table 4). Even so,

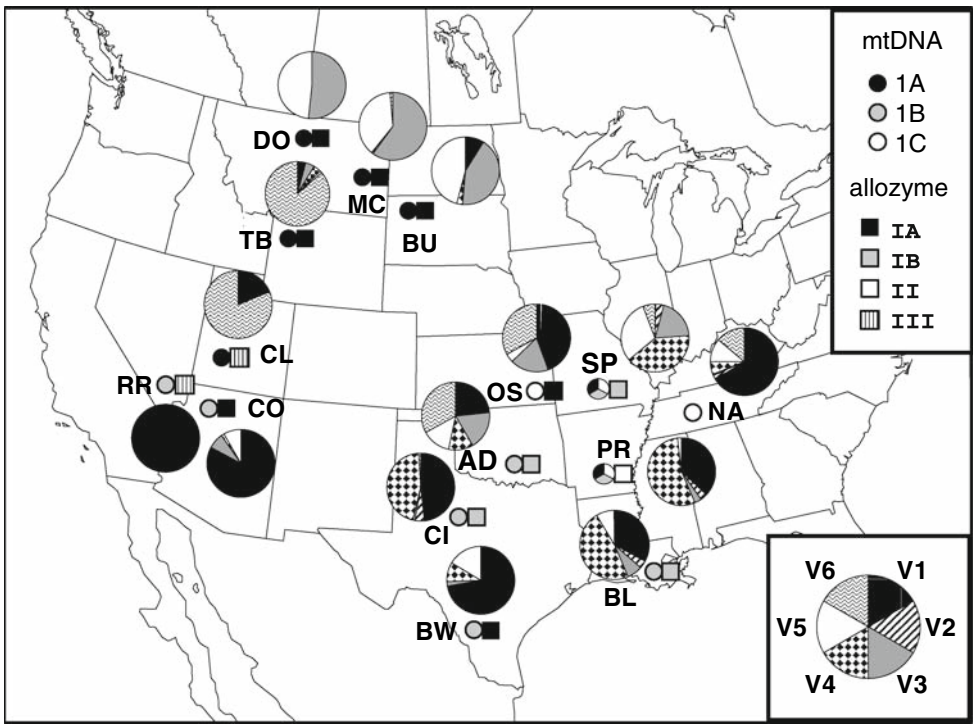

Fig. 2 Location of *Daphnia obtusa* sampling sites (*circles*) and expansion segment 43/e4 length variant frequencies (*pie charts*). The circles also indicate the mtDNA lineage (based on Penton et al. 2004) at each site, while the allozyme group (based on Hebert and Finston 1996) is indicated by squares. Allozyme data are not available for NA

sites in relatively close proximity sometimes belong to different lineages (e.g., RR and CO for allozymes, RR and CL for mtDNA; Fig. 2). In contrast, the correlation between geographic distance and mtDNA variation (based on average number of nucleotide differences) is not significant, which is consistent with the broad north–south distribution of lineage 1A (from Montana to Texas), and the broad east–west distribution of lineage 1B (from Louisiana to Nevada; Fig. 2). Although none of the Mantel tests between genetic markers are significant, the one between rDNA and allozymes is substantially higher than that between either of these and mtDNA, and is nearly significant (Table 4).

Discussion

Based on their analysis of the mitochondrial COI gene, Penton et al. (2004) concluded that mitochondrial phylogroups were probably restricted to different glacial refugia during the Pleistocene and subsequently expanded their ranges following deglaciation. They only detected one lineage at most sites, but they detected multiple lineages in some central sites (e.g., PR and SP in Fig. 2), and even within ponds, suggesting that lineages have come into secondary contact following postglacial expansion (Penton et al. 2004). However, they found the distribution of mtDNA phylogroups to be discordant with the three major allozyme groups identified by (Hebert and Finston 1996); a difference that was attributed to habitat niche shifts. Allozyme group I predominates throughout the Great Plains, while allozyme group II is primarily found east of the Mississippi River, and group III is restricted to the southwest US. Twelve of the 15 sites we sampled belong to allozyme group I. We were only able to analyze individuals from one site in group II (PR), which contains multiple mtDNA lineages and therefore may not be representative of this allozyme group with respect to rDNA variation.

Overall, the groups we identified using rDNA expansion segment length variation are more congruent with the allozyme than the mtDNA data (Table 4), although none of the Mantel tests between markers are statistically significant. Indeed, patterns of variant distribution in the south/central sites are likely complicated by the putative mixing of different refugial lineages (Penton et al. 2004; Paland et al. 2005). The AMOVA results indicate that a substantial portion of the variance in length variant frequencies occurs within individuals, which have as many as six of the seven observed variants. Even so, there is substantial population structure at the level of sampling sites (Table 3) and there are several distinct clusters of sites in the NJ tree (Fig. 3b). Although our sample sizes are small, it is unlikely that a

Table 2 Analysis of molecular variance in the frequency distribution of length variants in expansion segment 43/e4 of the 18S rRNA gene in *Daphnia obtusa* populations from the US

Site	F_{SC}^{a}	*P* value	F_{CT}^{b}	*P* value
AD	0.40	<0.0001	0.04	0.27
BL	0.21	<0.0001	0.05	0.03
BW	0.13	<0.0001	0.15	<0.0001
BU	0.14	<0.0001	−0.01	0.57
CL	0.29	<0.0001	NA[c]	
CI	0.09	<0.0001	−0.02	0.81
CO	0.08	<0.0001	NA	
DO	0.23	<0.0001	NA	
MC	0.07	<0.0001	−0.01	0.86
NA	0.00	0.67	NA	
PR	0.19	<0.0001	NA	
RR	0.00	1	0	1
SP	0.07	<0.0001	0.06	0.05
OS	0.19	<0.0001	NA	
TB	0.20	<0.0001	−0.06	0.52

[a] Proportion of variation that is due to differentiation between individuals within ponds

[b] Proportion of variation that is due to differentiation between ponds within sites

[c] F_{CT} could not be calculated for sites at which only a single pond was sampled

random sample of individuals over such a large geographic range would generate such a non-random distribution of length variants, and indeed, the overall variance is lowest among individuals within sites (14%, Table 3) even though variation between individuals within ponds (F_{SC}) is significant (Table 2). Moreover, the significant Mantel test supports a model of isolation by distance. However, this alone may not be sufficient to explain the distribution of length variants as frequencies are sometimes more similar between distant sites than they are between sites that are closer to one another. For example, TB is closer to MC (~300 km) than it is to CL (~750 km), but the variant frequency distribution in TB is much more similar to the one in CL than it is to the one in MC (Figs. 2, 3b).

The overall homogeneity of *Daphnia* gene pools across broad geographic regions, despite substantial differentiation at local scales (Crease et al. 1990, 1997; Ishida and Taylor 2007), has been attributed to the capacity for passive long-distance dispersal of *Daphnia* resting stages by both abiotic (e.g., wind) and biotic factors (e.g., birds). Indeed, it has been suggested that migratory waterfowl are an important mechanism of long-distance dispersal of aquatic invertebrates in North America (Figuerola et al. 2005; Charalambidou and Santamaria 2005), particularly along north–south axes (Taylor et al. 1998). Bird-mediated gene flow is certainly one explanation for the fact that most of the expansion segment length variants we observed have broad geographic distributions. This broad distribution could also occur if length variants are generated de novo on an ongoing basis by crossing over and gene conversion. In fact, crossing over between V1 and V6 could generate all the length variants that we observed, and they do co-occur at high frequency in several sites (Fig. 2). For example, three individuals from CL have high frequencies of V1 and V6, but no other variants were detected at this site suggesting that crossing over between them does not occur at a substantial rate. Moreover, the significant correlation between frequencies and geographic distance is unlikely to occur if length variants are continually generated de novo throughout the geographic range. Such a correlation is better explained by the divergence of established variant frequencies via genetic drift and restricted gene flow among populations that are geographically distant.

Another explanation for the geographic correlation with variant frequencies is natural selection, and our previous study has shown that uncompensated variants are likely to remain at low frequency as a result of selection acting against them (McTaggart and Crease 2005). However, this study also suggests that compensated variants are likely to be selectively neutral relative to one another. Thus, their frequencies will primarily be affected by genetic drift, gene flow, and the rate of de novo generation by crossing over and gene conversion, which appears to be low. Even so, it is possible that some crossing over and gene conversion occur between variants, but we are not able to detect it. For example, V3 and V5 occur in high frequency in individuals from ponds DO, MC, and BU, but crossing over between them will generate only uncompensated variants, which we expect to remain at low frequency due to negative selection. On the other hand, we have shown that rates of recombination at the level of the entire rDNA repeat can be substantial

Table 3 Analysis of molecular variance based on the frequency distribution of length variants in expansion segment 43/e4 of the 18S rRNA gene in *Daphnia obtusa* populations from the US

Source of variation	df	Variance component	Percentage of variation	*P* value	Fixation indices
Among sites	14	0.14	32.9	<0.0001	$F_{CT} = 0.33$
Among individuals within a site	119	0.06	13.9	<0.0001	$F_{SC} = 0.21$
Within individuals	13262	0.22	53.2	<0.0001	

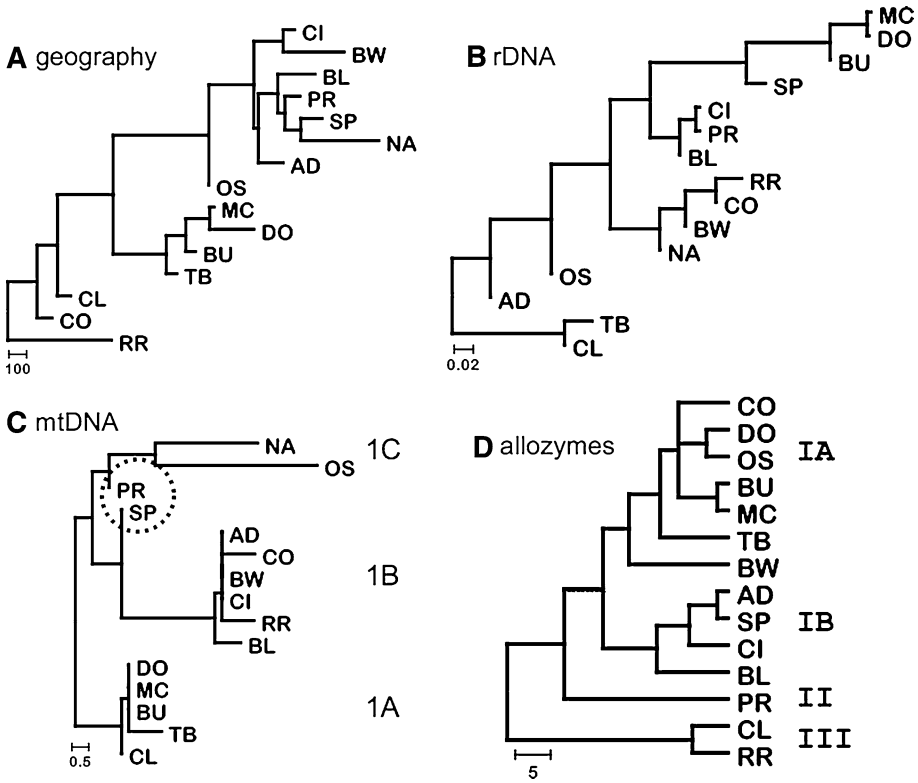

Fig. 3 Neighbor-joining trees constructed from genetic and geographic distances between *Daphnia obtusa* populations collected from 15 sites across the US. **a** Geographic distance among sites. **b** Nei's genetic distance based on expansion segment 43/e4 length variant frequencies. **c** Average number of nucleotide differences per nucleotide site based on the mitochondrial COI gene data of Penton et al. (2004). **d** Nei's genetic distance based on allozyme variation from Hebert and Finston (1996). The *dashed circle* around sites PR and SP in the mtDNA tree indicates that more than one mtDNA lineage was detected. Sequences sampled from the other sites all belong to a single mtDNA lineage

Table 4 Correlation, based on Mantel tests, between geographic distance and genetic distance in rDNA expansion segment 43/e4, the mitochondrial COI gene and allozymes in populations of *Daphnia obtusa* from the US

X matrix	Y matrix	*r* value	*P* value
Geographic distance	rDNA	0.289	0.008
Geographic distance	Allozymes	0.373	0.002
Geographic distance	mtDNA	0.150	0.13
Allozymes	rDNA	0.188	0.08
mtDNA	rDNA	0.035	0.40
Allozymes	mtDNA	−0.092	0.67

in apomictically propagated lines of *D. obtusa* (0.02–0.06 events per generation, McTaggart et al. 2007), which can lead to rapid changes in the relative frequency of length variants. However, we also found that there is no bias in the direction of change such that the average frequency of length variants among clonal lines derived from one female remains fairly constant. This is essentially the operation of genetic drift at the molecular level, which has the same impact on length variant frequencies as does genetic drift at the population level. Thus, our results are most consistent with the hypothesis that the length variants in expansion segment 43/e4 in *D. obtusa* are old, and are not generated de novo by crossing over or gene conversion at an appreciable rate.

The conclusion that length variants in *D. obtusa* are old implies that they are in a period of relative stasis relative to other expansion segments in species whose sequences contain repetitive motifs, and where uncompensated length variants can be observed at high frequencies (e.g., Gonzalez et al. 1985; Holzmann et al. 1996; Garcìa et al. 2004). Given enough time, *D. obtusa* populations may eventually go to fixation for a single variant through the combined effects of concerted evolution and genetic drift. However, if the length variants are truly neutral, then all populations may not go to fixation for the same variant. Thus, even infrequent gene flow could maintain some length variation in the species as a whole for a substantial period of time. Given the crucial role that ribosomes play in protein synthesis, intense selection clearly constrains their evolution. The fact that expansion segments are found in similar locations in the rRNA of a wide range of taxa (Wuyts et al. 2001) suggests that changes in secondary structure are limited to these regions. As long as the expanded sequences fold into stable secondary structures, and do not interfere with ribosome function, they will not be eliminated by selection. However, processes such as replication slippage may cause insertions and deletions in a repetitive sequence, which could generate new uncompensated length variants at a higher rate than their elimination by selection (Hancock and Vogler 2000). Generation of such variants would continue to accumulate until point mutations decrease the repetitive nature of the sequence (Hancock and Dover 1990), at which point selection would eventually purge energetically unstable variants, and the region would once

again enter a period of relative evolutionary stasis influenced primarily by stabilizing selection and segregation of neutral variation. Our results suggest that this is the current situation with expansion segment 43/e4 in *D. obtusa* from across the US.

In summary, our phylogeographic survey of length variation in rDNA expansion segment 43/e4 in *D. obtusa* supports the hypothesis that it is not presently undergoing substantial changes in sequence length. Instead, the length variation we observed is likely to be quite old, and simply segregating within and between populations. Over evolutionary time, this variation may be eliminated by genetic drift in the absence of new point mutations that increase the repetitive nature of the expansion segment. However, the occurrence of such mutations could possibly stimulate the generation of new length variants via replication slippage, which could in turn initiate a new phase of expansion segment length evolution.

Acknowledgments We thank Angela Holliss and Janet Topan for sequencing our plasmid clones, and Paul Hebert and Erin Penton for the *Daphnia* samples. This study was funded by a grant from the Natural Sciences and Engineering Research Council of Canada to TJC, and an Ontario Graduate Scholarship in Science and Technology to SJM.

References

Charalambidou I, Santamaria L (2005) Field evidence for the potential of waterbirds as dispersers of aquatic organisms. Wetlands 25:252–258

Crease TJ, Taylor DJ (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

Crease T, Lynch M, Spitze K (1990) Hierarchical analysis of population genetic structure in mitochondrial and nuclear genes of *Daphnia pulex*. Mol Biol Evol 7:444–458

Crease TJ, Lee S-K, Yu S-L, Spitze K, Lehman N, Lynch M (1997) Allozyme and mtDNA variation in populations on both sides of the *Daphnia pulex* complex from both sides of the Rocky Mountains. Heredity 79:242–251

Eickbush TH, Eickbush DG (2007) Finely orchestrated movements: evolution of the ribosomal RNA genes. Genetics 175:477–485

Excoffier LGL, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50

Felsenstein J (2004) PHYLIP (Phylogeny Inference Package) (Distributed by the author). Department of Genome Sciences, University of Washington, Seattle

Figuerola J, Green AJ, Michot TC (2005) Invertebrate eggs can fly: evidence of waterfowl-mediated gene flow in aquatic invertebrates. Am Nat 165:274–280

García M, Nicholson E, Nickrent D (2004) Extensive intraindividual variation in plastid rDNA sequences from the holoparasite *Cynomorium coccineum* (Cynomoriaceae). J Mol Evol 58:322–332

Gonzalez I, Gorski J, Campen T, Dorney D, Erikson J, Sylvester J, Schmickel R (1985) Variation among human 28S ribosomal RNA genes. Proc Natl Acad Sci USA 82:7666–7670

Hancock JM, Dover GA (1990) Compensatory slippage in the evolution of ribosomal-RNA genes. Nucleic Acids Res 18:5949–5954

Hancock JM, Vogler AP (2000) How slippage-derived sequences are incorporated into rRNA variable-region secondary structure: implications for phylogeny reconstruction. Mol Phylogenet Evol 14:366–374

Hebert PDN, Finston TL (1996) Genetic differentiation in *Daphnia obtusa*: a continental perspective. Freshw Biol 35:311–321

Holzmann M, Piller W, Pawloski J (1996) Sequence variations in the large sub-unit ribosomal RNA gene of *Ammonia* (Foraminifera, Protozoa) and their evolutionary implications. J Mol Evol 43:145–151

Ishida S, Taylor DJ (2007) Mature habitats associated with genetic divergence despite strong dispersal ability in an arthropod. BMC Evol Biol 7:52

McTaggart SJ, Crease TJ (2005) Selection on the structural stability of a ribosomal RNA expansion segment in *Daphnia obtusa*. Mol Biol Evol 22:1309–1319

McTaggart S, Dudycha JL, Omilian A, Crease TJ (2007) Rates of recombination in the ribosomal DNA of apomictically propagated *Daphnia obtusa* lines. Genetics 175:311–320

Paland S, Colbourne JK, Lynch M (2005) Evolutionary history of contagious asexuality in *Daphnia pulex*. Evol Int J Org Evol 59:800–813

Penton EH, Hebert PDN, Crease TJ (2004) Mitochondrial DNA variation in North American populations of *Daphnia obtusa*: continentalism or cryptic endemism? Mol Ecol 13:97–107

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Taylor DJ, Finston TL, Hebert PDN (1998) Biogeography of a widespread freshwater crustacean: pseudocongruence and cryptic endemism in the North American *Daphnia laevis* complex. Evol Int J Org Evol 52:1648–1670

Wuyts J, Van de Peer Y, Wachter RD (2001) Distribution of substitution rates and location of insertion sites in the tertiary structure of ribosomal RNA. Nucleic Acids Res 29:5017–5028