Ribosomal DNA Evolution at the Population Level: Nucleotide Variation in Intergenic Spacer Arrays of *Daphnia pulex*

Teresa J. Crease

Department of Zoology, University of Guelph, Guelph, Ontario, Canada N1G 2W1

ABSTRACT

Nucleotide variation was surveyed in 21 subrepeat arrays from the ribosomal DNA intergenic spacer of three *Daphnia pulex* populations. Eighteen of these arrays contained four subrepeats. Contrary to expectations, each of the four positions within the array had a different consensus sequence. However, gene conversion, involving sequences less than the length of a subrepeat, had occurred between subrepeats in different positions. Three arrays had more than four subrepeats and were undoubtedly generated by unequal crossing over between standard-length arrays. The data strongly suggested that most unequal exchanges between arrays are intrachromosomal and that they occur much less frequently than unequal exchanges at the level of the entire rDNA repeat. Strong associations among variants at different positions allowed the recognition of five groups of arrays, two of which were found in more than one population. Five of the seven individuals surveyed had arrays from more than one group. Analysis of the distribution of nucleotide variation suggested that the populations were quite divergent, a result that is concordant with previous surveys of allozyme and mitochondrial DNA variation. It was suggested that some of the subrepeat array types are quite old, at least predating the recolonization of pond habitats in the midwestern United States after the last glaciation.

THE observation that sequence variants can spread horizontally among the members of tandemly repeated multigene families has been termed concerted evolution (ARNHEIM 1983). The mechanisms thought to be responsible for this phenomenon include unequal crossing over, gene conversion, and replication slippage (DOVER 1982; ARNHEIM 1983). The interaction between such DNA exchange at the molecular level and population-level processes such as genetic drift and natural selection can result in the relative homogeneity of a multigene family within species despite the divergence of the family between species. However, the spread of new variants within a species is expected to take a considerable amount of time (DVOŘÁK *et al.* 1987) and thus intraspecific multigene family variation is also expected to occur. The pattern of such variation will depend on a variety of factors, including the rate of DNA exchange, the size of the gene family, the number of chromosomes on which it occurs, and the population structure (size, migration rate, breeding system) of the species.

At present, there are limited data on the distribution of multigene family sequence variation within and among conspecific populations. What studies there are (LANDESBERGER *et al.* 1992; BACHMANN *et al.* 1994; ELDER and TURNER 1994) have tended to survey variation in individual repeat units of satellite DNA. These studies provided information about the level of variation within and among individuals or populations; however, they can provide no information about variation among repeats within and between tandem arrays. Studies of variation among tandem repeats tended to focus on only one or two arrays (DVOŘÁK *et al.* 1987; HIBNER *et al.* 1991; BALDRIDGE *et al.* 199; and references within). However, LINARES *et al.* (1994) surveyed sequence variation in ribosomal DNA tandem repeats in two populations of *Drosophila melanogaster.* The purpose of this study is to examine the level and distribution of sequence variation in a tandemly repeated multigene family from several conspecific populations and to determine the relative importance of the different mechanisms that are responsible for these patterns.

Ribosomal DNA (rDNA) is a multigene family whose structure and function are well known (reviewed in LONG and DAWID 1980; GERBI 1985). In eukaryotes, rDNA is composed of tandem arrays of a basic repeat consisting of a transcription unit and an intergenic spacer (IGS). In most species studied to date, the IGS of each repeat contains a tandem array of short subrepeats that are thought to be involved in the control of rRNA transcription (GERBI 1985; BALDRIDGE *et al.* 1992; and references within). The subrepeats are usually on the order of hundreds of nucleotides in length, whereas the subrepeat arrays are on the order of a few thousand nucleotides in length, making them a tractable system on which to use sequence analysis.

Daphnia pulex is microcrustacean that inhabits freshwater ponds and lakes across North America and Europe. There are two modes of reproduction in this species: cyclic and obligate parthenogenesis. During the spring and summer, females produce eggs via ameiotic parthenogenesis. When environmental conditions deteriorate, males and diapausing or "ephippial" eggs are produced. The ephippial eggs of the cyclic partheno-

Author e-mail: tcrease@uoguelph.ca

gens are produced meiotically and require fertilization. The obligate parthenogens produce their ephippial eggs ameiotically, so they are strictly clonal. In temporary ponds, Daphnia populations must produce ephippial eggs at least once a year. Consequently, there is at least one sexual generation per year in cyclical parthenogens that inhabit such ponds. The existence of both types of reproduction in this species provides a unique opportunity to study the relative importance of meiosis in the concerted evolution of tandemly repeated multigene families.

Previous surveys of allozyme variation, and of restriction site variation in mitochondrial DNA (mtDNA) and rDNA in cyclically parthenogenetic populations of *D. pulex* from the midwestern United States have shown that they are highly differentiated, presumably as a result of limited dispersal ability and fluctuating selection on allozyme loci (LYNCH 1987; CREASE *et al.* 1989, 1990; CREASE and LYNCH 1991). In this study I report the results of a survey of DNA sequence variation in the rDNA-IGS subrepeat arrays of three of these previously sampled populations.

MATERIALS AND METHODS

D. pulex isolates from each of three populations in Illinois and Indiana were used in this analysis. The populations were sampled in the spring, soon after ephippial hatch. Consequently, the animals collected were either themselves the products of sexual reproduction or were the recent parthenogenetic descendants of sexually produced ephippial hatchlings.

The populations are numbered 1, 2 and 5 and correspond to PA, KA and BU, respectively, in previous studies (CREASE *et al.* 1990; CREASE and LYNCH 1991) . Population 2 is approximately midway between the other two, which are ~80 km apart. Daphnia collected from the three ponds were used to initiate laboratory cultures that were maintained parthenogenetically until sufficient numbers of animals were available for DNA extraction. Four isolates from population 1, one from population 2 and two from population 5 were randomly chosen from the isolates established in culture for the previous studies. Total DNA was extracted from each isolate using phenol-chloroform extraction and ethanol precipitation as described in CREASE (1986).

EMBL3 lambda phage libraries were constructed from *Bam*HI-digested DNA from each of the seven isolates using standard techniques (SAMBROOK *et al.*1989). Previous restriction mapping of the rDNA of *D. pulex* (CREASE and LYNCH 1991) showed that this enzyme produces an ~11-kb fragment containing all of the IGS and much of the coding sequences (Figure 1). The libraries were screened via filter lifts (SAMBROOK *et al.* 1989) and a digoxigenin-labeled probe (HÖLTKE *et al.* 1988) containing the entire IGS of *D. pulex* (CREASE and LYNCH 1991). At least 48 recombinant phage plaques that hybridized with the probe were collected from each library. Three of these were randomly chosen for further sequence analysis.

The *Xho*I fragment containing the IGS (Figure 1) from each of these 21 recombinant phage was subcloned into the phagemid Bluescript. Exonuclease III followed by S1 nuclease treatment was used to generate a set of overlapping deletion subclones from this fragment (Erase-a-Base kit, Promega Scientific, Madison, WI). This method preserves the order of subrepeats in the IGS array in the subsequent sequence analysis. The subclones were sequenced using the Sequenase 2.0 kit (United States Biochemical, Cleveland, OH) and [α^{35}S]dATP (1000 Ci/mmol, Amersham, Arlington Heights, IL). Sequencing products were separated on 6% denaturing polyacrylamide gels (Sequagel 6, National Diagnostics, Atlanta, GA). The DNA deletion subclone sequences were assembled with the aid of the SEQMAN program in the DNAStar software package (DNAStar, Madison, WI) and aligned using the ALIGN and MULIGN programs. A published sequence (CREASE 1993) included as array 132 in this study (GenBank accession number L07948) was used as a reference.

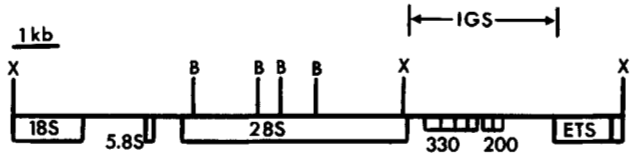

FIGURE 1.—Organization of the rDNA repeat unit of *D. pulex*. 18S, coding region for the small subunit (18S) rRNA; 5.8S, coding region for the 5.8S rRNA; 28S, coding region for the large subunit (28S) rRNA; ETS, external transcribed spacer; IGS, intergenic spacer; 330, the 330-bp subrepeats sequenced in this study; 200, 200-bp subrepeats (not related to the 330-bp subrepeats); B, *Bam*HI sites; X, *Xho*I sites.

Sequence divergence among pairs of individual IGS subrepeats and among entire subrepeat arrays was calculated using the Kimura-2 parameter model in the program DNADIST from the PHYLIP software package (FELSENSTEIN 1993). Dendrograms showing the relationships among individual subrepeats and among entire arrays were generated using the NEIGHBOR program of SAITOU and NEI (1987), available in PHYLIP. Nucleotide diversity (average number of substitutions per nucleotide site) within and between populations was calculated using the method of LYNCH and CREASE (1990). In addition, the degree of genetic differentiation among populations at the level of the subrepeat and of the array was calculated using their fixation index, *N*. This *N*-statistic is analogous to Wright's *F*-statistic in that it measures the amount of genetic variation within a megapopulation that is attributable to genetic differentiation among its subpopulations (HARTL 1980). However, in addition to differences among allele frequencies, *N* also explicitly accounts for variation between alleles at the nucleotide level.

RESULTS

Each IGS subrepeat array was assigned a three-digit number. The first digit refers to the population (1, 2 or 5) from which the Daphnia isolate was collected, the second digit refers to the individual from which the array was cloned, and the third digit refers to the particular array. Thus, arrays 141, 142, and 143 were cloned from individual 4 from population 1. Individual subrepeats are indicated by their position within an array. For example, 141-1 refers to the first subrepeat in array 141.

Of the 21 arrays sequenced, 18 contained three repeats of ~330 bp and one of ~200 bp. As has been previously described (CREASE 1993), the short repeat is likely the result of an intrastrand exchange event involving a 41-bp direct repeat that occurs twice in the longer subrepeats. The other three arrays were all

longer than four subrepeats in length: one (214) contained an additional 200-bp repeat and the other two (534 and 535) each contained one additional 200-bp and one additional 330-bp repeat for a total of six subrepeats. A total of 89 IGS subrepeats were present in the sample of 21 arrays.

When all of the subrepeats in arrays of size four were aligned with one another (Figure 2) and grouped according to their position in the array, it was apparent that variation was not randomly distributed among subrepeats but that each position within the array had a characteristic sequence. The fourth subrepeat in the array was always the short one. A total of 58 nt positions showed variation that was associated with differences among the four subrepeat types.

Based on these data, the additional repeats in the long arrays could be assigned to one of the four subrepeat types. Array 214 was composed of subrepeats in the order 1-2-3-3s-4, where 3s refers to a 200-bp subrepeat with sequence characteristic of a subrepeat normally found in position 3. Arrays 534 and 535 were composed of subrepeats in the order 1-2-3-4-3-4, where 4 is always a 200-bp subrepeat. The additional subrepeats in these arrays have been labeled 534-3b, 534-4b, 535-3b, 535-4b in Figure 2. These array configurations can easily be explained as the result of unequal crossing over between two standard-size arrays (Figure 3). Both events seemed to have involved a short repeat.

Although each position in the arrays had a characteristic sequence, there were occasionally short sections of sequence within a particular subrepeat that were more characteristic of another position. For example, subrepeats 111-3, 121-3 and 133-3 each had a section of sequence (nt 70–94, Figure 2) that was identical to the sequence normally found in subrepeats in position 2. The sequence on either side of this region was that normally found in subrepeats in position 3. Because the usual order and number of subrepeats is maintained in these arrays, the most likely explanation for such a pattern is a gene conversion event, without accompanying recombination, between misaligned subrepeats.

There were three other such "conversion tracts." One occurred in subrepeats 513-2 and 515-2 (nt 256–296, Figure 2) and involved a short section of sequence identical to that normally found in subrepeats in position 1. In the other two cases, it was not possible to determine which sequence was "unconverted" and which was "converted." Thirteen of the subrepeats in position 4 had a section of sequence (nt 286–292, Figure 2) that was characteristic of subrepeats in position 3, whereas the other 10 subrepeats were similar to subrepeats in position 2 in that region. Similarly, 13 subrepeats in position 1 contained a section of sequence (nt 70–78, Figure 2) that was characteristic of subrepeats in position 2, whereas the other 8 were similar to subrepeats normally found in position 3 in that region. In all cases, the sequences flanking the "conversion tract" were characteristic of the subrepeat normally found in that position.

The actual limits of these conversion tracts cannot be determined, but their maximum length can. In the first two cases (111-3, 121-3, 133-3 and 513-2, 515-2), the maximum extent of the conversion event was on the order of 130 bp. In the case of the two classes of type 1 subrepeats, the maximum extent of the conversion event was on the order of 30 or 65 bp, depending on whether the exchange occurred with a type 2 or with a type 3 subrepeat, respectively. In the case of the two classes of type 4 subrepeats, the maximum extent of the conversion event was on the order of 115 or 125 bp, depending on whether the exchange occurred with a type 3 or with a type 2 subrepeat, respectively. In all four cases, the maximum extent of a conversion tract between two subrepeats in different positions within an array was much less than the length of a subrepeat.

Sequence variation not associated with differences between major types of subrepeats also occurred and took the form of base substitutions as well as short insertions/deletions (indels). In most cases, indels involved changes of one or two nucleotides in the length of mononucleotide runs and can most easily be explained by replication slippage. However, a unique trinucleotide insertion occurred in subrepeat 132-4 (between nt 169 and 170, Figure 2) and a unique tetranucleotide deletion occurred in subrepeat 133-1 (nt 174–177, Figure 2).

Forty-seven (64%) of the 73 variants that were not associated with differences among subrepeat types were unique to a single subrepeat. Of the remaining 26 variants, only four occurred in more than one *type* of subrepeat. Subrepeat 214-3s shared the absence of a C at nt 154 (Figure 2) with subrepeat 122-4. Subrepeats 133-3, 112-4 and 535-4b shared a G at nt 300, and subrepeats 141-2, 212-2, 114-4, 122-4, 143-4, 212-4, 532-4 shared the loss of a G at nt 175 (Figure 2). Subrepeats in position 2 were characterized by a seventh A in a run of 6 As starting at nt 94. All five of the type 3 subrepeats from individual 53 (532-3, 534-3, 534-3b, 535-3, 535-3b) also had an additional A in this run. In addition, eight type 2 subrepeats (including 111-2, 121-2 and 133-2) had an eighth A in this run. The three type 3 subrepeats with a type 2 conversion tract (111-3, 121-3, 133-3) also had eight As in this region. Assuming that the differences in the number of As in this region did not all arise independently via replication slippage, the shared sequence variants in these subrepeats would indicate that the conversion event was probably intrachromosomal.

CREASE (1993) suggested that the sequence 5′-TATATAGGGAAG (nt 264–275, Figure 2) is an RNA polymerase promoter, such as is commonly found within the IGS subrepeats of other organisms (BALDRIDGE *et al.* 1992; and references within). Despite the clear divergence among the four types of subrepeats, the only variation in this sequence was an additional A (TAT-

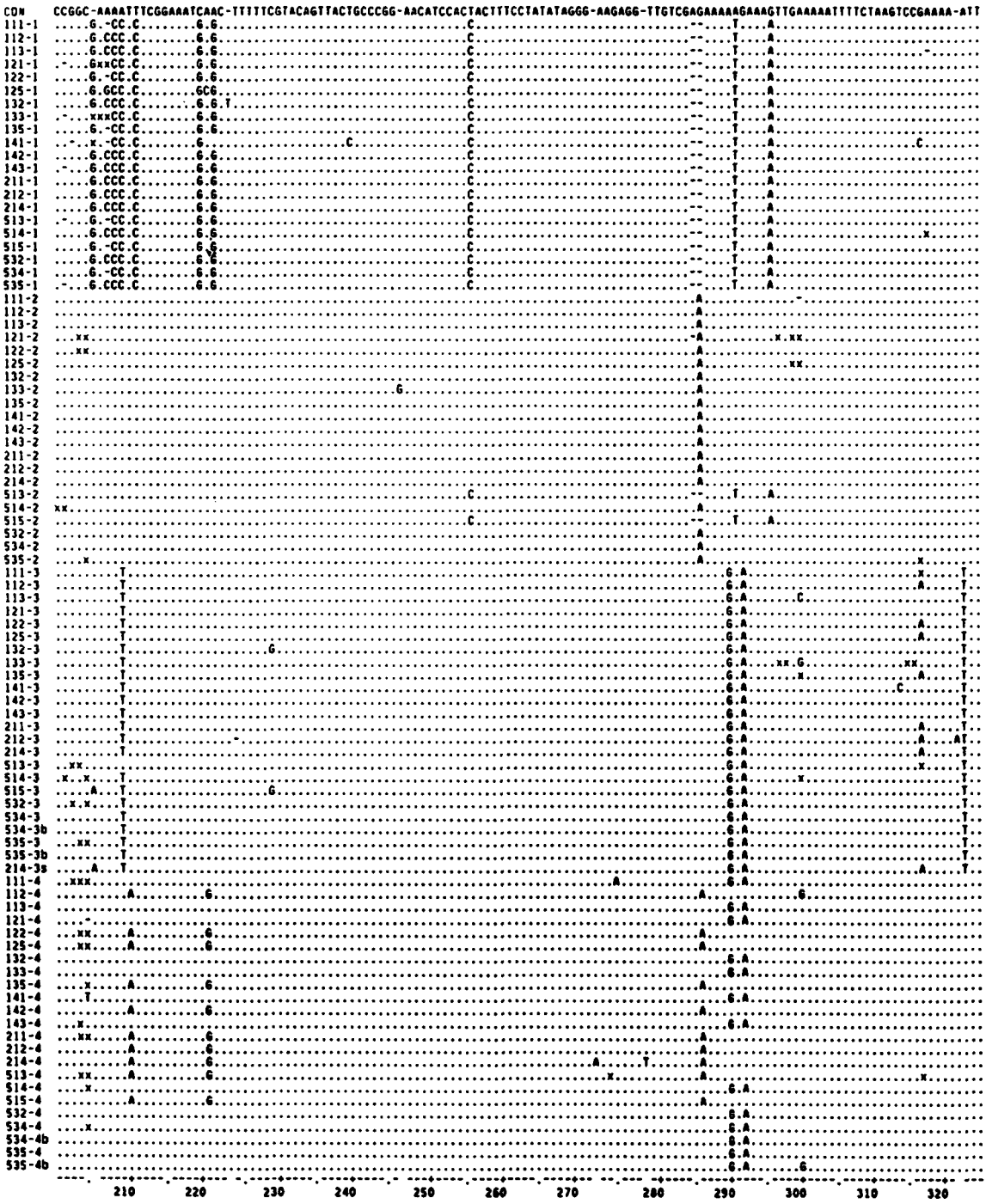

FIGURE 2.—Alignment of 89 IGS subrepeats from 21 arrays cloned from *Daphnia pulex*. The top sequence is the consensus (CON). Gaps introduced into this sequence to accommodate insertions present in some of the subrepeats have not been numbered. Dots indicate agreement with the consensus sequence, dashes indicate deletions and X's indicate unresolved nucleotides. The subrepeats have been grouped according to their position in the arrays.

ATAGGGAAAG) in subrepeat 214-4. The closest variant at the 5′ end of this sequence involved the substitution of the usual T for a C in type 1 subrepeats at nt 256 (Figure 2). Thus, there is a section of 16 bp that is perfectly conserved in all 89 subrepeats. There is only one other highly conserved section of sequence of the same or longer length: a section of 19 bp (17 bp in short repeats) from nt 128 to 146 (Figure 2). Whether or not this also might indicate some sort of functional constraint on this region is unknown.

Mean sequence divergence among subrepeats within types ranged from 0.69% (type 2) to 1.96% (type 4, Table 1), whereas mean sequence divergence between types ranged from 3.63% (2 *vs.* 4) to 14.15% (1 *vs.* 3). A dendrogram constructed from the matrix of pairwise sequence divergences clearly illustrates the distinctive

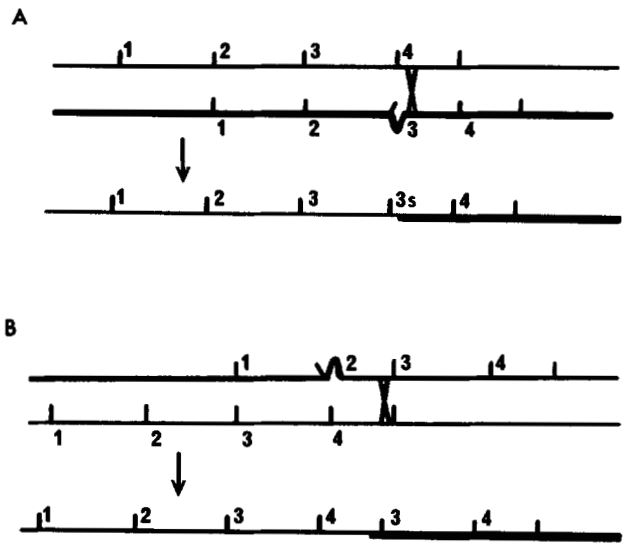

FIGURE 3.—Diagram showing possibilities for the generation of long arrays 214 (A) and 534 and 535 (B) via unequal crossing over between arrays of standard length.

nature of each of the four subrepeat types (Figure 4). The type 1 subrepeats fell into two groups: one including the subrepeats with the type 2-like sequence and the other including the subrepeats with the type 3-like sequence. The type 4 subrepeats also fell into two groups: one including the subrepeats with the type 3-like sequence and the other including the subrepeats with the type 2-like sequence. The divergence between these two subgroups was quite high because the subrepeats with the type 2-like region all shared several other variants not associated with differences between subrepeat types (Figure 2).

A preliminary analysis of intra- and interpopulation sequence divergence was carried out to determine whether there was any population structure with respect to subrepeat variation. However, because of the presence of all four types of subrepeats in approximately equal frequency in each population, intrapopulation diversity effectively overshadowed any diversity present among populations (data not shown).

It is clear from the alignment of intact arrays (Figure 5) that there are associations between subrepeat variants at each position along the array. In other words, arrays are not composed of a random assortment of subrepeat variants. A dendrogram constructed from the matrix of pairwise sequence divergence estimates between arrays suggests the existence of five groups or types of arrays, two of which are quite closely related to one another (Figure 6). If it is assumed that all of the variants in the sample only arose once, then some exchange must be occurring between arrays of different types. Indeed, type 1 arrays appear to have been created by recombination between a type 2 and a type 5 array. In addition, there is evidence of short exchanges, most likely the result of gene conversion, between array types. For example, arrays 112 and 113 share a C at nt 163 in subrepeat 3 (Figure 5). However, the extent of the conversion event was not sufficient to obscure the characteristics of either array type.

The diversity of distinct arrays types within populations, and even within individuals, was substantial given the number of arrays sampled. For example, four types of arrays were recognized among the 12 sequenced from population 1. All four individuals from this population had at least two types of arrays and two individuals had three types. Two very distinct types of arrays were found among the six sequenced from two individuals in population 5. One individual had both types, whereas all three arrays from the other individual were of the same type. The "postrecombination" portion of long arrays 534 and 535 were most similar to the "prerecombination" portion of the same arrays, again suggesting an intrachromosomal recombination event.

Nucleotide diversity at the level of arrays was highest within population 1 (Table 2), but this may be partly due to the larger sample of arrays sequenced from this population. Nucleotide diversity between populations was highest between populations 2 and 5, which did not share any array types. However, this pattern could change with a larger sample from these two populations. The analysis of population structure using the *N*-statistic approach showed that there was significant differentiation among the three populations with respect to nucleotide variation in IGS subrepeat arrays (N_{ST} = 0.31 ± 0.10, D = 10.07, $P < 0.01$).

An analysis of nucleotide diversity at the array level, within and among individuals, was also done for population 1 from which four individuals were analyzed. Mean

TABLE 1

Mean pairwise sequence divergence within and among IGS subrepeat types in *Daphnia pulex*

	Subrepeat 1	Subrepeat 2	Subrepeat 3	Subrepeat 4
1	0.0123 ± 0.0082	0.1168 ± 0.0092	0.1415 ± 0.0100	0.0999 ± 0.0078
2		0.0069 ± 0.0068	0.0784 ± 0.0101	0.0363 ± 0.0067
3			0.0079 ± 0.0062	0.0930 ± 0.0189
4				0.0196 ± 0.0172

Values are means ± SD.

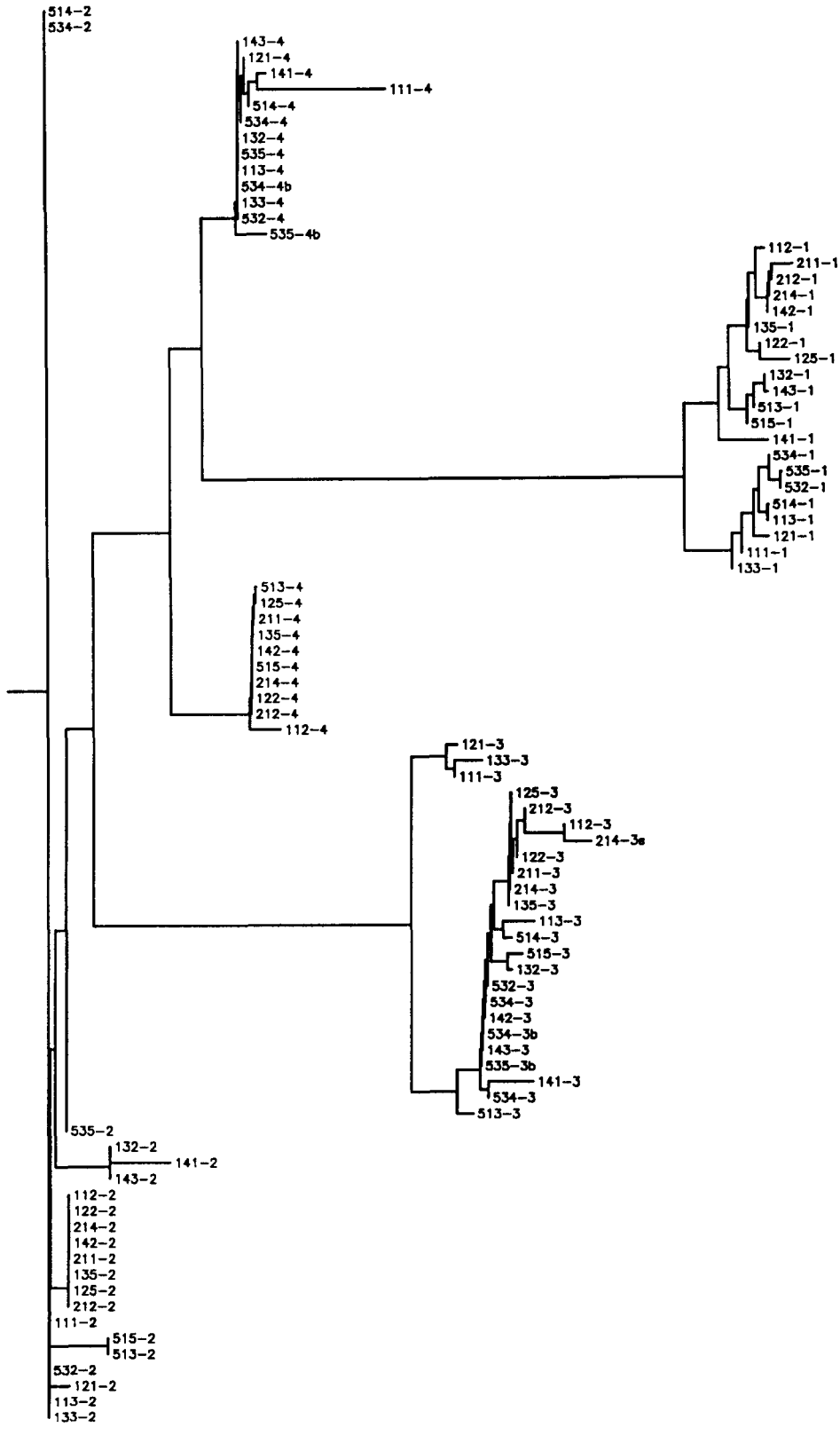

FIGURE 4.—Unrooted dendrogram showing the relationship among 89 IGS subrepeats from *Daphnia pulex*. Branch lengths are proportional to sequence divergence.

nucleotide diversity within individuals (0.0210 ± 0.0039, Table 3) was essentially the same as the diversity within the population as a whole (0.0193 ± 0.0030, Table 2). The estimate of between-individual diversity is corrected by subtracting the mean diversity within the two individuals under consideration (LYNCH and CREASE 1990). The negative values in Table 3 reflect the fact that intraindividual diversity exceeded the uncorrected estimate of interindividual diversity in this population.

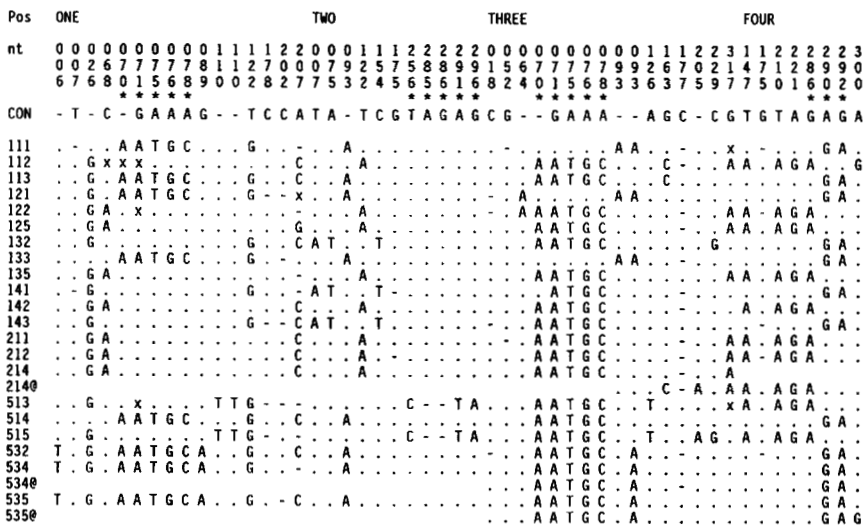

FIGURE 5.—Segregating nucleotide sites in 21 IGS arrays of *Daphnia pulex*. The numbers above the sequence (nt) correspond to the nucleotide positions for individual subrepeats in Figure 2. Dots indicate agreement with the consensus sequence and dashes indicate deletions relative to the consensus. The 5′ flanking nt position is given in the case of insertions relative to the consensus sequence. The subrepeat position in the array is indicated above the nucleotide positions (Pos). The "postrecombination" portions of long arrays 214, 534 and 535 (indicated by @) have been aligned with the corresponding region in the standard-length arrays. Asterisks indicate nucleotide positions involved in conversion tracts (see text).

DISCUSSION

It is clear from the results of this study that unequal crossing over and gene conversion, which are thought to lead to the concerted evolution of tandemly repeated multigene families, are operating at the level of the IGS subrepeat arrays (IGSSA) in *D. pulex*. Despite the fact that exchanges occur among subrepeats, a distinct consensus sequence has evolved at each position within the array (Figure 2). The occurrence of position-specific differences among IGSSA has also been observed in *D. melanogaster*. LINARES *et al.* (1994) aligned sequences of the first two 95-bp subrepeats of IGSSA cloned from four strains of *D. melanogaster* and found differences that were shared by repeats in the same position in all four strains. In addition, they found two major types of 240-bp subrepeats that were not randomly distributed along the IGSSA: the 5′ half of each array was composed of type B subrepeats and the 3′ half of each array was composed of type A subrepeats. In contrast, LASSNER and DVOŘÁK (1986) sequenced two IGSSA from wheat and found that each array had a different subrepeat consensus, suggesting that variants were spreading within arrays much more quickly than they were spreading between arrays. Additional data on intraspecific sequence variation in IGSSA will be required to determine which, if either, of these patterns is most characteristic of IGSSA evolution.

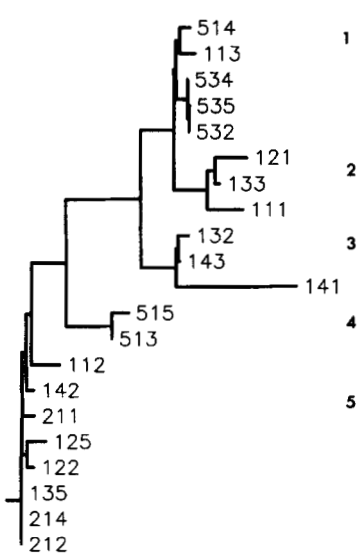

FIGURE 6.—Unrooted dendrogram showing the relationship among 21 IGS subrepeat arrays from *Daphnia pulex*. Branch lengths are proportional to sequence divergence. Only the "prerecombination" portion of long arrays 214, 534, and 535 was included in the analysis.

A previous restriction site survey (CREASE and LYNCH 1991) showed that one particular size class of rDNA repeats predominates in *D. pulex* and in no case did individuals lack this size class. This and a previous sequence study (CREASE 1993) have shown that this repeat size class contains an IGSSA with four subrepeats. CREASE and LYNCH suggested that the predominance of one size class of repeat could reflect a relatively recent origin of repeat types of other sizes or that there is stabilizing selection in favor of IGSSA of that particular length. Such selection on IGSSA length has been suggested for *Drosophila melanogaster* (WILLIAMS *et al.* 1987). Even if selection was acting to maintain an IGSSA size of four, each of the four positions would not necessarily be expected to have a distinct consensus sequence. Presumably, unequal crossing over between IGSSA of nonstandard length could generate new arrays with four subrepeats whose order is not 1-2-3-4 with respect to the consensus sequence. However, both types of long arrays sequenced in this study could be most easily explained via unequal crossing over between IGSSA of standard length as the standard order of subrepeats was maintained on either side of the crossover point (Figure 3). If this is true for most of the long arrays that occur in these populations, it may reflect their recent origin relative to arrays with four subrepeats. However, there is also the possibility that selection is acting to maintain not only the number of subrepeats in an array but also their position-specific sequence.

TABLE 2

Estimates of nucleotide diversity in IGS subrepeat arrays within and between populations of *D. pulex* reproducing via cyclical parthenogenesis

Population(s)	Nucleotide diversity
Within populations	
1	0.0193 ± 0.0030
2	0.0026 ± 0.0011
5	0.0112 ± 0.0032
Mean	0.0119 ± 0.0049
Between populations	
1 and 2	0.0042 ± 0.0025
1 and 5	0.0018 ± 0.0018
2 and 5	0.0085 ± 0.0031
Mean	0.0048 ± 0.0016

Values are estimates ± SE.

TABLE 3

Estimates of nucleotide diversity in IGS subrepeat arrays within and between individuals of *D. pulex* population 1

Individual(s)	Nucleotide diversity
Within individuals	
11	0.0221 ± 0.0075
12	0.0186 ± 0.0083
13	0.0215 ± 0.0071
14	0.0219 ± 0.0080
Mean	0.0210 ± 0.0039
Between individuals	
11 and 12	−0.0020 ± 0.0061
11 and 13	−0.0029 ± 0.0059
11 and 14	−0.0010 ± 0.0066
12 and 13	−0.0013 ± 0.0069
12 and 14	−0.0014 ± 0.0064
13 and 14	−0.0037 ± 0.0061
Mean	−0.0020 ± 0.0048

Values are estimates ± SE.

Gene conversion between misaligned subrepeats could also lead to the "shuffling" of subrepeat order in IGSSA if conversion tracts were longer than one subrepeat in length. However, the results of this study suggest that the length of such exchanges is much less than a subrepeat. The observation that IGSSA conversion tracts are shorter than the length of the subrepeats has been made in wheat and mouse (DVOŘÁK *et al.* 1987) and *Drosophila* (LINARES *et al.* 1994).

The overall pattern of sequence diversity in *D. pulex* suggests that most exchanges among IGSSA are intrachromosomal. Variants that were segregating within one position were strongly associated with variants at other positions along the array, making it possible to identify distinct IGSSA types (Figures 5 and 6). Furthermore, most unequal exchanges seemed to occur between arrays of the same type despite the high diversity of array types within individuals (Table 3). SEPARAK *et al.* (1988) argued that such associations among variants can only develop when sister chromatid exchange occurs much more frequently than exchange between homologous or non-homologous chromosomes.

DVOŘÁK *et al.* (1987) showed that gene conversion is distance dependant at the level of IGS subrepeats. If this is also the case at the level of the entire rDNA repeat unit, then distance-dependent frequency of exchange, along with the prevalence of intrachromosomal exchange, is expected to lead to the clustering of IGSSA types along a chromosome. Whether this is the case in *D. pulex* has not yet been tested. However, the observation that rDNA repeat types (identified via restriction site variation) that were rare within a population could occur with high frequency within an individual (CREASE and LYNCH 1991) is consistent with the idea that the variant repeat types were clustered. A similar pattern was observed by SEPARAK *et al.* (1988) for length variants in human rDNA. Thus, the most likely explanation for the divergence of subrepeat positions along the IGSSA, in addition to the nonrandom associations between subrepeat variants at each position, is a very low rate of unequal intrachromosomal exchange within IGGSA relative to the rate of exchange at the level of the entire rDNA repeat.

Despite the divergence of the subrepeats at different positions along the IGSSA, some regions remained highly conserved across positions. The sequence involving the spacer promoters was identical in all but one of the 89 subrepeats. Such conservation around spacer promoters has been observed in the IGSSA of other organisms. For example, TAUTZ *et al.* (1987) sequenced one IGS from each of four species of Drosophila and found that the sequence surrounding the spacer promoter was highly conserved within each array but differed between the species. Similarly, little or no variation was found at spacer promoters within IGSSA of the tsetse fly, *Glossina morsitans morsitans* (CROSS and DOVER 1987); the mosquito, *Aedes albopictus* (BALDRIDGE and FALLON 1992) and rice (TAKAIWA *et al.* 1990).

The occurrence of a second region of high conservation among *D. pulex* subrepeats suggests that this region may be under functional constraint as well. BALDRIDGE *et al.* (1992) found that the consensus subrepeat sequences of many species, including plants, invertebrates and vertebrates, have the potential to form strong secondary structure. In a more detailed analysis of the mosquito, *A. albopictus,* they found that most sequence variation among subrepeats occurred in unstable stem structures, bulges and loops. The possibility that similar constraints on secondary structure are responsible for sequence conservation among subrepeats in *D. pulex* deserves further consideration. Attempts to determine secondary structure in these subrepeats and to determine the impact of variation within and between subrepeat types are currently under way.

Substantial levels of intrapopulation variation occurred with respect to nucleotide variation in the IGSSA

of *D. pulex* (Table 2). The observation that nucleotide diversity was highest in population 1 is consistent with the restriction site survey of the entire rDNA repeat (CREASE and LYNCH 1991). Although more types of rDNA repeats were found in population 2 (13) than in population 1 (9) in that survey, sequence divergence was higher, on average, among the repeat types found in population 1, as was the case in the present study.

Although no two IGSSA were identical, they did fall into distinct groups, two of which had representatives in more than one population. Even so, there was clear differentiation among the three populations, suggesting restricted gene flow. In addition, some of the IGSSA types in population 1 were not closely related to one another, making it unlikely that they arose *in situ*. Previous studies of mtDNA and allozyme variation in the populations surveyed in this study (CREASE *et al.* 1990) also suggested restricted gene flow. However, most differences between the populations at allozyme loci were due to differences in allele frequencies rather than allelic substitutions. CREASE *et al.* attributed the widespread geographic distribution of alleles at allozyme loci in midwestern *D. pulex* populations to the fact that the midwest was probably recolonized from relict populations that persisted south of the extent of the last glaciation. Thus, current patterns of variation among populations most likely reflect continued divergence in the absence of substantial contemporary gene flow. If so, it is highly likely that some of the IGSSA types found in these populations have existed within the species for considerable periods of time.

The pattern of sequence variation in the IGSSA of *D. pulex* is not consistent with that found in studies of population genetic variation in satellite families. For example, ELDER and TURNER (1994) found little intraindividual or intrapopulation variation among 170-bp satellite monomers in the pupfish, *Cyprinodon variegatus*. However, there was significant differentiation between populations suggesting that concerted evolution of this satellite family was occurring at the level of individuals and populations. On the other hand, BACHMANN *et al.* (1994) found very little variation among 102-bp satellite monomers at any level in the cave cricket, *Dolichopoda schiavazzii*. Most of the variation they observed was due to rare variants scattered throughout the monomer sequence. In addition, there were no associations between variants at different positions within the subrepeat sequence. Gene flow is known to be extremely restricted, if not nonexistent, among some of the populations surveyed in this study. BACHMANN *et al.* suggested that the large size of this satellite family has effectively "buffered" the consensus monomer sequence against change, even between populations that rarely, if ever, exchange genes. The differences between the results of this study and those involving satellite families may reflect differences in the population structure of the organisms surveyed and/or differences in patterns of DNA exchange in satellites and in rDNA. Surveys of both types of gene families, in the same populations, will be required to address these possibilities.

The restriction site survey of CREASE and LYNCH (1991) suggested that recombination at the level of entire rDNA repeats occurred within obligately parthenogenetic lineages of *D. pulex* (which lack meiosis) but that, on average, individuals from such lineages carried fewer rDNA repeat types than did their cyclical relatives. Thus, an important consequence of meiotic recombination may be the redistribution of IGGSA variants among individuals thereby retarding the loss of intraindividual repeat-type variation. If so, then very little difference in the distribution of sequence variation within IGSSA types would be expected between obligate and cyclic parthenogens. To further investigate the impact of the loss of meiosis at the sequence level, a survey of IGSSA variation within and among obligately parthenogenetic clones with widespread geographic distributions has been undertaken.

In addition to the high levels of intraspecific rDNA variation observed within *D. pulex*, CREASE and LYNCH found species-specific differences between *D. pulex* and its close relative, *D. pulicaria*. However, it is noteworthy that very few fixed site differences were observed between the two species. In most cases, a site not found in one species was still polymorphic in the other, suggesting that the two species have diverged very recently or that it can take a very long time for species differences to accumulate. A comparison of the consensus subrepeat sequence and the distribution of IGSSA variation in these two species may provide additional insights into the mechanisms responsible for the unusual distribution of IGSSA variation in *D. pulex*.

I thank JOE TRAKALO for technical assistance and M. M. FERGUSON for helpful comments on the manuscript. This work was supported by Natural Sciences and Engineering Research Council research grant OGP0130594 to T.J.C. and by National Science Foundation grant BSR 86-00487 to M. LYNCH. The sequences of the 21 IGS arrays have been deposited in GenBank under accession numbers U34752, U34782-U34792, and U34864-U34871.

LITERATURE CITED

ARNHEIM, N., 1983 Concerted evolution of multigene families, pp. 38–61 in *Evolution of Genes and Proteins*, edited by M. NEI and R. K. KOEHN. Sinauer, Sunderland, MA.

BACHMANN, L., F. VENANZETTI and V. SBORDONI, 1994 Characterization of a species-specific satellite DNA family of *Dolichopoda schiavazzii* (Orthoptera, Rhaphidophoridae) cave crickets. J. Mol. Evol. **39:** 274–281.

BALDRIDGE, G. D., and A. M. FALLON, 1992 Primary structure of the ribosomal DNA intergenic spacer from the mosquito *Aedes albopictus*. DNA Cell Biol. **11:** 51–59.

BALDRIDGE, G. D., M. W. DALTON and A. M. FALLON, 1992 Is higher-order structure conserved in eukaryotic ribosomal DNA intergenic spacers? J. Mol. Evol. **35:** 514–523.

CREASE, T. J., 1986 Mitochondrial DNA variation in *Daphnia pulex* Leydig reproducing by obligate and cyclic parthenogenesis. Ph.D. thesis. Washington University, St. Louis, MO.

CREASE, T. J., 1993 Sequence of the intergenic spacer between the 28S and 18S rRNA-encoding genes of the crustacean, *Daphnia pulex*. Gene **134:** 245–249.

CREASE, T. J., and M. LYNCH, 1991 Ribosomal DNA variation in *Daphnia pulex*. Mol. Biol. Evol. **8:** 620–640.

CREASE, T. J., D. J. STANTON and P. D. N. HEBERT, 1989 Polyphyletic origin of asexuality in *Daphnia pulex*. II. Mitochondrial DNA variation. Evolution **43:** 1016–1026.

CREASE, T. J., M. LYNCH and K. SPITZE, 1990 Hierarchical analysis of population genetic variation in mitochondrial and nuclear genes of *Daphnia pulex*. Mol. Biol. Evol. **7:** 444–458.

CROSS, N. C. P., and G. A. DOVER, 1987 Tsetse fly rDNA: an analysis of structure and sequence. Nucleic Acids Res. **15:** 15–30.

DOVER, G.A., 1982 Molecular drive: a cohesive mode of species coevolution. Nature **299:** 111–116.

DVOŘÁK, J., D. JUE and M. LASSNER, 1987 Homogenization of tandemly repeated sequences by distance-dependent nucleotide sequence conversion. Genetics **116:** 487–498.

ELDER, J. F., JR., and B. J. TURNER, 1994 Concerted evolution at the population level: pupfish *Hin*dIII satellite DNA sequences. Proc. Natl. Acad. Sci. USA **91:** 994–998.

FELSENSTEIN, J., 1993 PHYLIP: phylogeny inference package. Version 3.51c. University of Washington, Seattle, WA.

GERBI, S. A., 1985 Evolution of ribosomal DNA, pp. 419–517 in *Molecular Evolutionary Genetics*, edited by R. J. MACINTYRE. Plenum, New York.

HARTL, D., 1980 *Principles of Population Genetics*. Sinauer, Sunderland, MA.

HIBNER, B. L., W. D. BURKE and T. H. EICKBUSH, 1991 Sequence identity in an early chorion multigene family is the result of localized gene conversion. Genetics **128:** 595–606.

HÖLTKE, H. J., G. SAGNER, C. KESSLER and G. SCHMITZ, 1992 Sensitive chemiluminescent detection of Digoxigenin-labelled nucleic acids: a fast and simple protocol and its applications. BioTechniques **12:** 103–114.

LANDSBERGER, N., S. CANCELLI, D. CARETTONI, C. BARIGOZZI and G. BADARACCO, 1992 Nucleotide variation and molecular structure of the heterochromatic repetitive *Alu*I DNA in the brine shrimp *Artemia franciscana*. J. Mol. Evol. **35:** 486–491.

LASSNER, M., and J. DVOŘÁK, 1986 Preferential homogenization between adjacent and alternate subrepeats in wheat rDNA. Nucleic Acids Res. **14:** 5499–5512.

LINARES, A. R., T. BOWEN and G. A. DOVER, 1994 Aspects of nonrandom turnover involved in the concerted evolution of intergenic spacers within the ribosomal DNA of *Drosophila melanogaster*. J. Mol. Evol. **39:** 151–159.

LONG, E. O., and I. B. DAWID, 1980 Repeated gene in eukaryotes. Annu. Rev. Biochem. **49:** 727–764.

LYNCH, M., 1987 The consequences of fluctuating selection for isozyme polymorphisms in *Daphnia*. Genetics **115:** 657–669.

LYNCH, M., and T. J. CREASE, 1990 The analysis of population survey data on DNA sequence variation. Mol. Biol. Evol. **7:** 377–394.

SAITOU, N., and M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual*, Ed. 2. Cold Spring Harbour Laboratory Press, New York.

SEPERACK, P., M. SLATKIN and N. ARNHEIM, 1988 Linkage disequilibrium in human ribosomal genes: implication for multigene family evolution. Genetics **119:** 943–949.

TAKAIWA, F., S. KIKUCHI and K. OONO, 1990 The complete nucleotide sequence of the intergenic spacer between 25S and 17S rDNAs in rice. Plant Mol. Biol. **15:** 933–935.

TAUTZ D., C. TAUTZ, D. WEBB and G. A. DOVER, 1987 Evolutionary divergence of promoters and spacers in the rDNA family of four *Drosophila* species: implications for molecular coevolution in multigene families. J. Mol. Biol. **195:** 525–542.

WILLIAMS, S. M., G. R. FURNIER, E. FUOG and C. STROBECK, 1987 Evolution of the ribosomal DNA spacers of *Drosophila melanogaster*: different patterns of variations on *X* and *Y* chromosomes. Genetics **116:** 225–232.

Communicating editor: G. B. GOLDING