Gradual Evolution of a Specific Satellite DNA Family in *Drosophila ambigua, D. tristis,* and *D. obscura*[1]

Lutz Bachmann and Diether Sperlich

Department of Population Genetics, University of Tübingen

The highly repetitive satellite DNA family "ATOC180" is specific for the three closely related species *Drosophila obscura, D. ambigua,* and *D. tristis* but does not occur in their closest relatives *D. subsilvestris* and *D. bifasciata.* Approximately 10,000 copies/haploid genome of ~180-bp repetition units are tandemly arranged in the centromeric heterochromatin of all chromosomes of all three species. Molecular analysis of 29 cloned repeats shows much intra- and interspecific sequence homogeneity. Single nucleotide changes are the main source of variability and distinguish the sequence-, subfamily- and species-specific ATOC180 repeats from each other. Based on these nucleotide differences, phylogenetic dendrograms were constructed and compared with published trees for other traits. The data indicate that the sequences of the ATOC180 satellite DNA family probably arose in a phylogenetically "short period" during the anagenetic evolution of the common ancestor of *D. obscura, D. tristis,* and *D. ambigua,* as a consequence of a process of genome reorganization, followed by a "long period" of entirely gradual sequence evolution. For the latter period, an evolutionary rate of 3×10^{-8} substitutions/site/year was calculated.

Introduction

Highly repetitive satellite DNA (satDNA) is a characteristic component of the genomes of almost all eukaryotic organisms. Only some fungi have been reported to lack this class of DNA (Timberlake 1978). Several common structural attributes characterize this specific component of the genome, regardless of the actual nucleotide sequence of the basic repeat in question (Beridze 1986, pp. 109–113). Typically, highly repetitive DNA is noncoding. It is restricted to the heterochromatic regions of the chromosomes, near the centromeres and/or telomeres. Several thousand or even several million copies of repetition units are tandemly arranged and form large homogeneous arrays. Among different organisms, the length of the repetition unit may vary considerably. For example, only 2-bp repeats were found in crabs (Gray and Skinner 1974), and 5-bp repeats were found in *Drosophila melanogaster* (Lohe and Roberts 1988), but 40-kb repeats were found in *Muntiacus vaginalis* (Benedum et al. 1986), to mention a few extreme cases. However, most repetition units of satDNA that have so far been described range in length between ~100 bp and 400 bp (Hsieh and Brutlag 1979; Hörz and Altenburger 1981).

Surely, satDNA is most curious with respect to its intraspecific sequence conservation and its special mode of evolution. Some satDNA families are clearly species

1. Key words: *Drosophila obscura* group, highly repetitive DNA, DNA evolution, genome evolution.

Address for correspondence and reprints: Lutz Bachmann, Department of Population Genetics, University of Tübingen, Auf der Morgenstelle 28, 7400 Tübingen, Germany.

Mol. Biol. Evol. 10(3):647–659. 1993.

0737-4038/93/1003-0011$02.00

specific (Bachmann et al. 1989), and some are amplified to a high copy number in one species but appear in only moderate copy numbers in closely related species (Bachmann et al. 1990). Other satDNAs may be restricted to a specific phylad of related species (Ganal et al. 1986; Cremisi et al. 1988), while others may be present in all species of a whole genus (Vignali et al. 1991).

Several models have been proposed to explain these phenomena. Smith (1976) was the first to demonstrate by computer simulation that "unequal crossing-over" can generate and maintain tandemly repeated homogeneous DNA clusters. However, his model only works when the copy number of tandemly arranged sequences is balanced by natural selection. More complex simulations (Charlesworth et al. 1986; Stephan 1986) showed that arrays of tandemly arranged satDNA accumulate most probably in chromosome regions where the recombination rate is low. Unequal crossing-over is frequently accepted as the mechanism that conserves the high intraspecific homogeneity of an already existing tandemly arranged DNA. In addition to unequal crossing-over, two other mechanisms are considered for the origin of arrays of tandemly arranged repetitive DNA. While slippage replication might be an important mechanism for the amplification of simple-sequence DNA such as minisatellites in humans (Walsh 1987; Stephan 1989), the finding of extrachromosomal circular alphoid satDNA (Okumura et al. 1987) suggests that rolling-circle amplification (Hourcade et al. 1973) might also play an important role in the amplification processes of satDNA.

In this paper we will describe the specific properties of a satDNA family common to three species of the *D. obscura* group (*D. obscura, D. ambigua,* and *D. tristis*). The results will be discussed in the context of satDNA organization and evolution.

Material and Methods

Drosophila Strains

The strains of *D. tristis* (1978) and *D. obscura* (1977) were derived from wild flies collected in Tübingen (Germany), and the *D. ambigua* (1973) strain was derived from flies collected in Vienna (Austria). All strains have been kept in the laboratory at constant 18°C in continuous light.

DNA Isolation and Cloning

Genomic DNA of *Drosophila* was extracted as described by Preiss et al. (1988). Highly repetitive DNA was isolated from restriction satellite bands visible after 5% polyacrylamide gel electrophoresis of *Eco*RI- or *Pst*I-digested genomic DNA of all the three species mentioned above. The DNA fragments were eluted from the gel, purified, and ligated into the plasmid pUC19 (King and Blakesley 1986) and were transformed to *Escherichia coli* JM 103 cells. Recombinant clones were selected as white colonies on ampicillin plates containing X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) and IPTG (isopropyl-β-D-thiogalactopyranoside).

DNA Hybridization

Digested genomic DNA separated on agarose gels was blotted to Hybond N membranes (Amersham) according to the method of Southern (1975). Bacterial colonies were transferred to nitrocellulose membranes (Davis et al. 1986, pp. 227–229). Labeling of probe DNA, hybridization, and detection of the hybridization signals were performed using the DIG DNA labeling and detection nonradioactive kit (109 36 57; Boehringer) as described in the manual. In some experiments ^{32}P-labeled probes

(Feinberg and Vogelstein 1983) were used for hybridizations (Davis et al. 1986, pp. 84–87).

DNA Sequencing

Plasmid DNA was purified according to the manual of the Diagen Plasmid Kit Hi-purity (41014; Diagen). The chain-termination method (Sanger et al. 1977) was used to sequence both strands of the pUC19 clones. It was performed as described in the instructions of the T7 sequencing kit (27-1682-01; Pharmacia).

Computer Analysis

DNA sequences were analyzed by DNASIS (version 5.0). Genetic distances were calculated according to Kimura's (1980) two-parameter method, with the DNADIST (version 1.0) program from J. C. Miller. The programs NJTREE (version 2.0) from N. Saitou and L. Jin, UPGMA (version 2.0) from L. Jin, TDRAW (version 1.14) from J. W. H. Ferguson, and PHYLIP from J. Felsenstein were used for constructing and plotting the dendrograms.

Results

Genomic DNA of *Drosophila ambigua, D. obscura,* and *D. tristis,* digested with the restriction nuclease *Pst*I or *Eco*RI and electrophoresed on 5% polyacrylamide gels, shows clearly visible bands of highly concentrated restriction satDNA fragments in the background smear of single-copy DNA fragments at ~180, ~360, and ~540 bp. The fragments of the *Pst*I (*D. ambigua, D. obscura*) or *Eco*RI bands (*D. tristis*) of ~180 bp were eluted from the gel and ligated into the pUC19 plasmid vector DNA. To show that the recombinant plasmids contain a specific class of satDNA, they were probed with labeled DNA of the respective fraction extracted from *D. obscura.* Almost all recombinant clones obtained from *D. ambigua* and *D. obscura* DNA gave positive signals, but few of those from *D. tristis* DNA did so. This can be explained by the presence of another 181-bp telomeric *Eco*RI restriction satDNA family in *D. tristis* (Bachmann et al. 1990). Four positive *Pst*I clones from *D. ambigua* (pAPC180/1–4), 5 *Eco*RI clones from *D. tristis* (pTEC181/1–5), and 18 *Pst*I clones from *D. obscura* (pOPC182/1–18) were selected for sequence analyses.

The alignment of the nucleotide sequences of these 27 clones (fig. 1) proved that they all belong to the same family of satDNA. The family was termed "ATOC180." Furthermore, it became evident that another 328-bp *Eco*RI-*Hae*III clone (pAEH328) from genomic DNA of the same strain of *D. ambigua,* selected in an earlier cloning experiment on a different topic, also belongs to the same class of DNA. The sequence comparison revealed that the insert of this clone contains two almost complete tandemly arranged copies of the ATOC180 family. The two repetition units of pAEH328 were termed "pAEC180/1" and "pAEC180/2," respectively, and are included in figure 1.

The nucleotide sequences of the ATOC180 clones are slightly AT rich (55%–60%). The sequences of only two clones (pOPC182/4 and pOPC182/5) are identical; all others differ from each other, to varying extents. The differences are mainly single nucleotide substitutions. Some of the base substitutions are autapomorphic, and others are common to two or more sequences. In several cases, the nucleotide substitutions were present in all clones extracted from only one species (e.g., G at position 26 in all clones from *D. tristis*).

*Eco*RI-clones:

```
pTEC 181/1  CTGTTGAGTT ATGGTGAAAC AGGGAG*TTT A*TTGCGGTT AACGT*TATA GAGGCTATAT CCTTTCTGTC CTGAAATCCA AATTGATTGC AAACGATCCA
pTEC 181/2  ---------- ---------- ------*--- -C-------- T----*---- ---------- --C--A---- T--------- ---------- -C--------
pTEC 181/3  --A------- ---------- ------*--- -C-------- T---C*---- ---------- ----*A---- ---------- ------CG-- ---A------
pTEC 181/4  ---------- ---------- ------*--- -C-------- T----*---- -T-------- -----A---- ---------- ---------- ---G------
pTEC 181/5  A--------- --T------- ------*--- -C---G---- T----*---- -T-------- -T---A---- ---------- ---------- ----------

pAEC 180/1  -----T-AA- ------T--- -----T*--A -T---GAG-- T----*A--- -T-------- -A---A---- ---T------ -------C-- --------A-
pAEC 180/2  -------AA- ---A--T--- -----T*--C -C---GAT-- T----*---- -T-------- A--C-TA-G- -
```

*Pst*I-clones:

```
            PstI    10         20         30         40         50         60         70         80         90         100
pAPC 180/1  ---CA----- --A------- -----T*--- -C--*-AT-- T----*---T AT-T------- ---C-A---G ---T-----G -T-A------ ----------
pAPC 180/4  ---CA----- --A------- -----T*A-- -C--*-AT-G T----*---- AT-------- ---C-A---- ---T-----G ---A------ --------G
pAPC 180/2  ---CA---G- --A------- -----T*--- -C--*-AT-- T----*---- AT-------- ---C-AC--- ---T------ ---G------ ----------
pAPC 180/3  ---CA----- --A------- -----T*--- -C--*-AT-- T----*---- -T-------- ---C-A-T-- ---T--AA-- ---G------ ----------

pOPC 182/4  ---CA----- -GA--A---- -----TA--- -A----A--- T----*--C- TG-------- -----A---- ---T------ --------A- ----C--AA-
pOPC 182/5  ---CA----- -GA--A---- -----TA--- -A----A--- T----*--C- TG-------- -----A---- ---T------ --------A- ----C--AA-
pOPC 182/3  ---CA----- -GA------- -A---TA--- -C-G--C-GA T----*---- -T-------- -----A---- ---C------ ---------- ---A----T
pOPC 182/1  ---CA----- -GA------A -----TA--- -C-G--A-A- T----*---- -TA------- -----A-T-- ---C------ ---------- ---A----T
pOPC 182/15 ---CA----- -GA------- -----TA--- -C-G--A-A- T----*---- -T-------- -----A---- ---C------ ---------- ---A----T
pOPC 182/2  ---CA----- -GA------- ------A-A- -C-G-AA-A- T----G---- -T-A------ -----A---- ---------- ---------- --GAC--T
pOPC 182/14 ---CA----- -GA-----T- T----TA--- -CAG-AAAA- T----*---- -T--A----- -----A---- ---C---T-- ---------- ---A----T
pOPC 182/10 ---CA----- -GA------- -----TA--- -C-G--A-A- T---C*A--- -T-------- -----A---- ---CC----- ---------- ---A----T
pOPC 182/12 ---CA----- -GA------- -----TA--- -C-G--A-A- T---C*A--- -T-------- -----A---- ---CC----- ---------- ---A----T
pOPC 182/16 ---CA----- -GA------- -----TA--- -C-G-TA-A- T----*A--- -T--A----- -----A---- ---CC---T- ---------- ---A----T
pOPC 182/6  ---CA----- -GA------A -----TA--- -C-G--A-A- T----*---- -T-------- -----A---- ---CC----- ---------- ---A----T
pOPC 182/7  ---CA----- -GA------A G----TA--- -C-G--A-A- T----*---- -T-------- -----A---- ---------- ---------- ---A----T
pOPC 182/8  ---CA----- -GA------- -----TA--A -C-G--A-A- T--A-*---- -T-------- -----A---- ---CC----- ---G---A-- ---A----T
pOPC 182/17 ---CA----- -GA-----G- -----TA--- -C-G--C-G- T----*---- -T-------- --A--A---- ---CC----- ----------G ---A----T
pOPC 182/18 ---CA----- -GA-----G- -----TA--- -C-G--C-GA T----*---- -T-------- -----A---- ---C------ ---------- ---A----T
pOPC 182/9  ---CA----- -GA------- -----TA--- -C-G--A-A- T----*---- -T-------- -----A---- ---C------ --------T- ---A----T
pOPC 182/11 ---CA----- -GA-----G- -----TA--- -C-G--C-GA T----*---- -T-------- -----A-A-- ---CC----- ---------- ---A----T
pOPC 182/13 ---CA----- GGA-----GT -----TA--- -C-G--C-GA T----*---- -T--A----- -----A---- ---CC----- ---------- ---A----T
```

*Eco*RI-clones:

```
                EcoRI
pTEC 181/1  GC*GAATTC* **AGAAGGCA TTGACACATT CCGGAAATTC TAGAGATGGC TGGGAGGCGT GGCATGGCCT GT*TCCGATA AGAAA
pTEC 181/2  --*------T TC--G---G- ---------- -----C---- ---------- -A-------- -----A---- --*------- -----
pTEC 181/3  --*------* **-------- ---------- ---A------ ---------- -A-------- ---------- --*------- -----
pTEC 181/4  --*------T TC-------- ---------- ---------- ---------- -A-----A-- ---------- --*----T-- -----
pTEC 181/5  ---------T TC-------- ---------- ---------- ---------- -A-----A-- ---------- --*------- ----T

pAEC 180/1  --*------* **-------- ----G----- -A---C--CG --C------- -A-------- ------TATG -A*------- -----
pAEC 180/2     -----TT CA-------- ----G----- -A---C---- --C-----A-- -A-------- ---------- CA*------- ----C
```

*Pst*I-clones:

```
                    110         120         130         140         150         160         170         180
pAPC 180/1  --TCA----T CC-------- ---C------ -A---C---- --CT------ -A-------- --G---TATG -A*------- -----
pAPC 180/4  --*-C---TT CC-----A-- A-------C- TA---C---- --C------- -A-------- --G---TATG -A*------- -----
pAPC 180/2  --*-C----T CC----A--- ---C------ -A---T---- --C------G -A-C------ T-G--AT--- -AA------- -----
pAPC 180/3  --*-C----T CC-------- --------C- TA---C---- --C------- -A-------- --G---T--- -A*------- -----

pOPC 182/4  --*AC----T CC-------- --ACA----- -A---T---- --C------- -A-------- ------TG-A -A*------- -A---
pOPC 182/5  --*AC----T CC-------- --ACA----- -A---T---- --C------- -A-------- ------TG-A -A*------- -A---
pOPC 182/3  A-*AC----T CC-------- ---------- -A---CG--G -TC------- -A-------- ------T--- -A*------- -----
pOPC 182/1  --*AC----T CC-------- ---------- -A--CCGG-G --C------- -A-------- ------T--- -A*------- -----
pOPC 182/15 --*AC-G--T CC-------- ---------- -A---CG--G --TT------ -A-------- ------T--- -A*------- -----
pOPC 182/2  --*AC----T CC-------- -----T---- -A---CG--G --C------- -A-------- ------T--- -A*----C-- -----
pOPC 182/14 --*AC----T CC-------- -----TT--- -A---CG--G --C------- -A-------- ------T--- -A*------- -----
pOPC 182/10 --*AC----T CC-------- ---------- -A---CG--G --C-----A- -A-------- ------T--- -A*------- -----
pOPC 182/12 --*AC----T CC-------- ---------- -A---CG--G --C-----A- -A-------- ------T--- -A*------- -----
pOPC 182/16 --*AC----T CC-------- --TG---G-- -A---CG--G --C------- -A-T------ ------T--- -A*------- -----
pOPC 182/6  --*AC----T CC-------- C-T-----A- -T---CG--G --C------- -A-------- ------T--- -A*------- -----
pOPC 182/7  --*AC----T CC-------- ---------- -A---CG--G -GC------- -A-------- ------T--- -A*------- -----
pOPC 182/8  --*AC-G--T CC-------- -----T---- -A---CG--G --C------- -A-------- ------T--- -A*------- -----
pOPC 182/17 --*AC-G--T CC-------- -----TT--- -A---CG--G --C------- -A-------- ------T--- -A*------- -----
pOPC 182/18 --*AC----T CC-------- ---------- -A---CG--G --C------- -A-------- ------T--- -A*------- -----
pOPC 182/9  --*AC----T CC-------- ---------- -A---TG-CG -TC------- -C-------- ------T--- -A*------- -----
pOPC 182/11 --*AC----T CC-----T-- ---------- -A---CG--G -TC------- -A-------- ------T--- -A*------- -----
pOPC 182/13 --*AC----T CC-------- ---------- -A---TG-CG -TC------- -A-------- ------T--- -A*------- -----
```

FIG. 1.—Nucleotide sequences of pOPC182/1–18 from *Drosophila obscura,* pAPC180/1–4 and pAEC180/1–2 from *D. ambigua,* and pTEC181/1–5 from *D. triauis,* aligned to pTEC181/1. The *Pst*I clones start at position 1, and the *Eco*RI clones start at position 114. Asterisks indicate gaps (i.e., indels) introduced to increase sequence similarity.

In addition, five single nucleotide insertions and/or deletions (indels) were found. Three of these indels are autapomorphic (at position 46 in pOPC182/2, at position 103 in pAPC180/1, and at position 173 in pAPC180/2). Two gaps were observed in all the clones belonging to a group of related sequences (at position 27 in pOPC182/1–18 and at position 35 in pAPC180/1–4). One particular gap, a three-base deletion, is common to pTEC181/1 and to pTEC181/3, which were isolated from *D. tristis,* and to pAEC180/1, which was isolated from *D. ambigua.*

No sequence cloned via *Pst*I has an *Eco*RI site, and no sequence cloned via *Eco*RI has a *Pst*I site. This is not due to the cloning procedure, because the presence of an *Eco*RI site does not necessarily exclude a priori the absence of a *Pst*I site, and vice versa. Furthermore, the *Pst*I site is modified in all *Eco*RI clones (pTEC181/1–5 and pAEC180/1–2) by the substitution of two nucleotides (CTGCAG becomes CTGTTG). In addition, in all *Pst*I clones derived from *D. obscura* (pOPC182/1–18), the *Eco*RI site is changed by the substitution of two nucleotides (GAATTC becomes ACATTC). The modification of the *Eco*RI site in the *Pst*I clones from *D. ambigua* by a 1-bp substitution (GAATTC becomes GCATTC) seems to be an intermediate link between the correct *Eco*RI site and the consensus sequence of *D. obscura.* The mutually exclusive occurrence of the *Eco*RI and *Pst*I restriction sites was consequently taken as argument for dividing the ATOC180 family into the two subfamilies ATEco180, which comprises the *Eco*RI clones of *D. ambigua* and *D. tristis,* and AOPst180, which comprises the *Pst*I clones of *D. ambigua* and *D. obscura.* Furthermore, Southern-blot analyses (data not shown) confirmed that the exclusive occurrence of the *Pst*I and *Eco*RI restriction sites is not true only for the isolated clones used in these experiments (it could be argued that the analyzed sequences are biased because of the cloning of restriction fragments of a given length). A "ladder-like" hybridization pattern, typical for tandemly arranged satDNA, with signals at fragment lengths of 180, 360, 540 bp, etc., was only obtained by hybridization of a labeled ATOC180 probe to *Pst*I-digested genomic DNA of *D. obscura* and to *Eco*RI-digested genomic DNA of *D. tristis.* Hybridizing the same probe to reciprocally digested DNA (*Eco*RI for *D. obscura* DNA and *Pst*I for *D. tristis* DNA) gave only signals with fragments of high molecular weight. A ladder-like hybridization pattern could be obtained with either *Pst*I- or *Eco*RI-digested genomic DNA of *D. ambigua,* because members of both ATOC180 subfamilies (ATEco180 and AOPst180) are present in the DNA of this species (see fig. 1). As will be shown later, this subdivision reflects the evolutionary history of the ATOC180 satDNA family, as illustrated in the dendrogram (fig. 2).

Trials of the hybridization of pAPC180/1 DNA to filterbound *Eco*RI-digested genomic DNA of *D. ambigua, D. tristis,* and *D. obscura;* of several other *D. obscura* group species (i.e., *D. subobscura, D. guanche, D. madeirensis, D. bifasciata, D. subsilvestris, D. pseudoobscura,* and *D. azteca*); and of *D. melanogaster* proved that the sequences of the ATOC180 family are specific for the species triad *D. obscura, D. ambigua,* and *D. tristis.* In situ hybridization of DNA probes from the same clone to mitotic chromosomes of *D. ambigua, D. tristis,* and *D. obscura* identified the presence of clusters of ATOC180 sequences around the centromeres of all chromosomes (M. Raab, personal communication).

By means of dot-blot hybridizations of pTEC181/1 DNA to genomic DNA of *D. tristis,* we estimated that 1.7% of the genome of this species consists of ATOC180 sequences. If it is assumed that the genome of *D. tristis* is approximately the same size as *D. melanogaster* (150×10^6 bp/haploid genome), this number corresponds

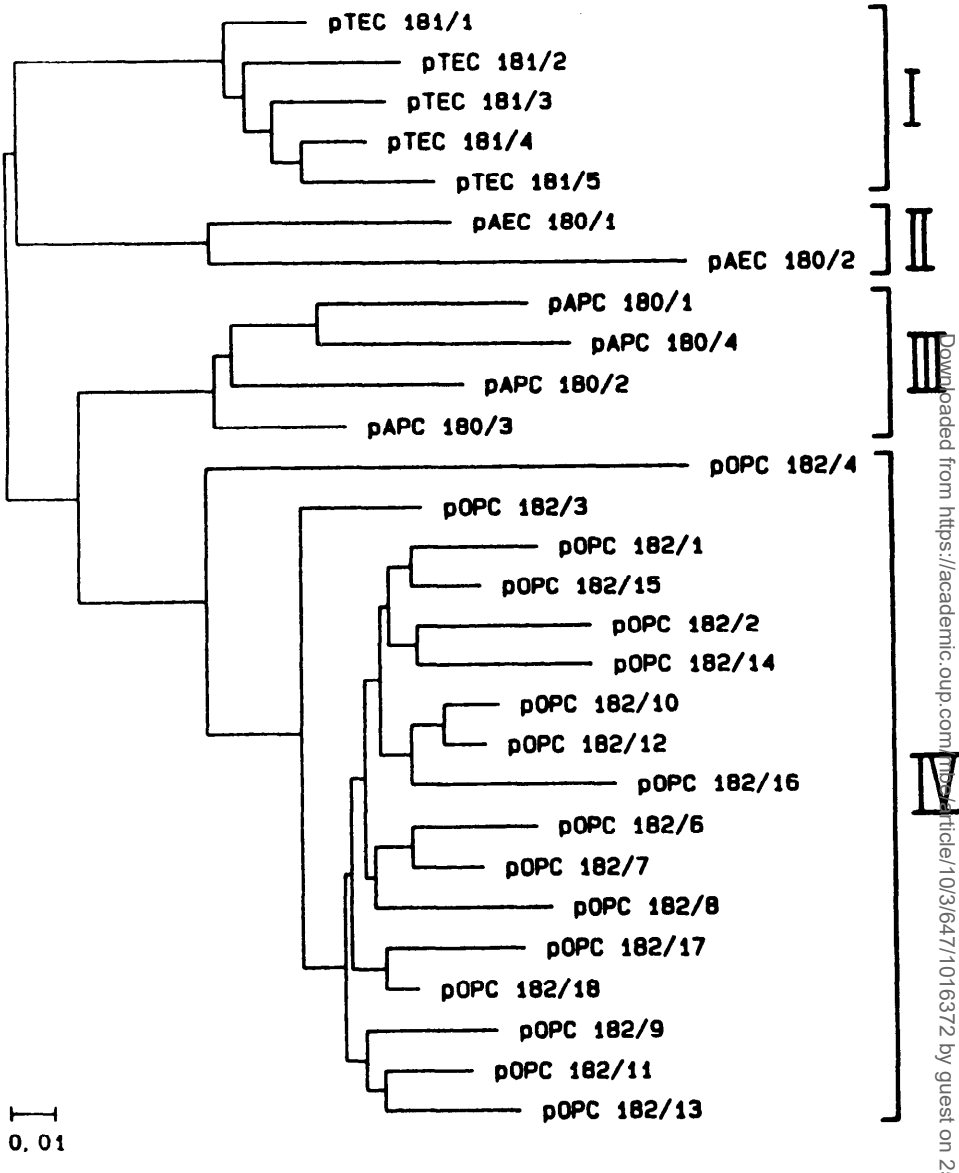

FIG. 2.—Unrooted neighbor-joining dendrogram of relationship of the nucleotide sequences of pOPC182/1–18, pTEC181/1–5, pAPC180/1–4, and pAEC180/1–2. Scale bar represents a genetic distance *D* of 0.01 as the frequency of nucleotide substitutions in pairwise comparison of two nucleotide sequences according to Kimura's (1980) two-parameter method.

to ~14,000 copies/haploid genome. Though no direct experiments were performed to determine the proportion of ATOC180 in the genomes of *D. ambigua* and *D. obscura,* similar intensities of hybridization signals indicate that copy numbers are equivalent to those of *D. tristis.*

Genetic distances of the cloned sequences (except for pOPC182/5, which is identical to pOPC182/4) were calculated according to Kimura's (1980) two-parameter method. The data (which can be provided on request) were used to construct two

dendrograms. A rooted tree (not shown) was constructed according to the UPGMA method of Sneath and Sokal (1973, pp. 230–234), and a second unrooted network (fig. 2) results from the neighbor-joining method of Saitou and Nei (1987). Although the UPGMA and neighbor-joining methods are based on different evolutionary assumptions, the resulting dendrograms are practically identical; in both cases, four clusters of sequences become apparent (I—pTEC181/1–5, II—pAEC180/1–2, III—pAPC180/1–4, and IV—pOPC182/1–3 and 6–18), and each cluster only includes ATOC180 sequences derived from one of the three species. All clones isolated from *D. tristis* are located exclusively in cluster I, and the clones derived from *D. obscura* all appear in cluster IV—only pOPC182/4 ranges slightly apart, thus linking clusters III and IV. The sequences of *D. ambigua* are found in the two separate clusters II and III. Cluster III is placed, in both cases, in the proximity of the cluster of the *D. obscura* clones, whereas the position of cluster II differs slightly in the two dendrograms. The 62 most probable maximum-parsimony trees (Eck and Dayhoff 1966, p. 164; Kluge and Farris 1969; not shown) calculated by the DNAPARS computer program (Felsenstein 1988) differ from the distance trees only by minor rearrangements within the four groups of clones.

Discussion

In recent publications concerning the phylogeny of the *Drosophila obscura* group, the three species *D. ambigua, D. tristis,* and *D. obscura* have been shown to constitute a monophyletic triad (Cabrera et al. 1983; Felger and Pinsker 1987; Cariou et al. 1988), although the exact phylogenetic relationships within this triad are still unclear. Allozyme data (Lakovaara and Keränen 1980) allow us to calculate the time of divergence of these three species to be ~3–5 Mya, on the basis of Nei's (1987, p. 234) substitution rate for electrophoretically detectable variants.

In our studies on repetitive DNAs, the satDNA family ATOC180 was identified as a genome component exclusively common to *D. ambigua, D. tristis,* and *D. obscura,* located in the centromeric heterochromatin of all chromosomes. The genomic organization of the ATOC180 sequences shows homogeneous arrays of tandemly arranged repetition units of ~180 bp, and the same may also be assumed for the common ancestor of this species triad.

The comparison of nucleotide sequences of 29 cloned repetition units of the ATOC180 satDNA family (18 from *D. obscura,* 5 from *D. tristis,* and 6 from *D. ambigua*) shows great intra- and interspecific sequence similarity. If the assumption of convergent sequence evolution is excluded, it is possible to reconstruct the consensus sequence of the ATOC180 satDNA repeats of the common ancestor for 156 nucleotide positions (~85% of the sequence). Twenty-six positions (~15%) remain uncertain because of the fixation of different nucleotides or indels in the consensus sequences of the extant species (fig. 3).

UPGMA, neighbor-joining, and maximum-parsimony dendrograms show four clearly separated clusters of related sequences. The existence of sequence clusters in the dendrograms can easily be explained by the occurrence of cluster (species)-specific nucleotide substitutions shared by all sequences of a cluster (species), while the branching within the clusters (species) is due to randomly occurring sequence variability. Speciation is the main reason for the evolution of species-specific sequence characters.

The agreement among the trees obtained by UPGMA, neighbor joining, and

```
        10          20          30          40          50          60
CTG??GAGTT A??GTGAAAC AGGGA??TTT ACT??C???T TACGTTATAG TGGCTATATC

        70          80          90         100         110         120
CT?TATGTCC TG?AATCCAA ATTGATTGCA AA?GATCC?G C??ATTCT?C AGAAGGCATT

       130         140         150         160         170         180
GACACATTC? GGA??TT?TA ?AGATGGCTA GGAGGCGTGG ?ATG?CCTG? TCCGATAAGA

AA
```

FIG. 3.—Consensus sequence of the ATOC180 satellite DNA family in the common ancestor of *Drosophila ambigua, D. tristis,* and *D. obscura,* deduced from sequence comparisons of pOPC182/1–8, pTEC181/1–5, pAPC180/1–4, and pAEC180/1–2. Question marks indicate nucleotides that remain open because of variability in the extant species.

maximum parsimony leads us to believe that the ATOC180 satDNA family sequences evolved gradually at an almost constant rate. Otherwise, the neighbor-joining and the maximum-parsimony networks should, for the ATOC180 sequences, give trees that are different from those given by the UPGMA method. If so, one can reconstruct the evolution of the ATOC180 satDNA family. The highest interspecific distance values appear between pTEC181 and pOPC182, and the lowest occur between pAPC180 and pOPC182 sequences. This indicates that the cladogenesis of *D. tristis* should have taken place before the speciation of *D. ambigua* and *D. obscura.* This interpretation is supported by mitochondrial restriction data (Gonzalez et al. 1990) and by transpositions and duplications of the histone gene cluster in the chromosome elements of these species (Felger and Pinsker 1987).

According to this argument, the *Eco*RI clones shared by *D. tristis* and *D. ambigua* are regarded as phylogenetically older than the *Pst*I clones from *D. ambigua* and *D. obscura.* After the separation of *D. tristis,* the AOPst180 subfamily probably evolved by the partial homogenization of substitutions at the *Eco*RI site, in two steps (from GAATTC in clusters I and II and in their common ancestor to GCATTC in cluster III to ACATTC in cluster IV), while at the *Pst*I site CTGTTG changed to CTGCAG, along with the substitution of a G by an A at position 13. While AOPst180 sequences became homogenized in *D. obscura,* both types of sequences (ATEco180 and AOPst180) are still present in *D. ambigua.*

A 3-bp deletion at positions 110–112, common to pTEC181/1, pTEC181/3, and pAEC180/1, indicates that the ATEco180 subfamily repeats must have already existed as two length variants (A and B) in the common ancestor: ATEco180 A with the deletion and ATEco180 B without it. However, this deletion became neither fixed nor lost by homogenization in *D. tristis* and *D. ambigua.* Therefore, this 3-bp deletion characterizes only length variants, not subfamilies of repeats. Since the homogenization of other sequence characters seems not to be restricted by the presence or absence of this deletion, ATEco180 A and B sequences do not appear as a monophyletic group in the dendrogram, although it can be assumed that this particular 3-bp deletion results from a single mutation in the common ancestor of *D. ambigua* and *D. tristis.*

The observed sequence diversity between the species-specific consensus sequences of pTEC181 and pOPC182 is ~10%–15%, and the assumed time for the separation of *D. tristis* and *D. obscura* is 3–5 Mya. This corresponds roughly to a substitution rate (α) of 3×10^{-8} substitutions/site/year. This value is very close both to the

1–1.5 $\times 10^{-8}$ substitutions/site/year, deduced from ADH sequences of several Hawaiian *Drosophila* species (Rowan and Hunt 1991), and to 1×10^{-8} substitutions/site/year, deduced from DNA-DNA hybridization data of coding DNA of several species of the *D. obscura* group (Caccone and Powell 1990). As fossil records are not available, the estimation of substitution rates in *Drosophila* might appear to be very approximate. On the other hand, the example of Hawaiian drosophilids shows that biogeographic data can be related to geologic data, with great accuracy. The only estimate of $\alpha > 25 \times 10^{-8}$ substitutions/site/year that is available for noncoding and nonrepetitive DNA is also based on DNA-DNA hybridization data (Caccone and Powell 1990). However, it should be emphasized that the estimate of 3×10^{-8} substitutions/site/year considers only the diversity between the consensus sequences. The actual intraspecific sequence variability already ranges between 7% and 15%. This variability is mainly due to randomly distributed nucleotide changes, which seem to allow an identification of individual repetition units but which are not at all characteristic of the average.

The evolution of already existing tandemly repeated satDNA is driven by two main processes. Mutation leads to variability between repeated sequences by single nucleotide changes. This provides the basis for the evolution of the consensus sequences through random fixation by homogenization processes such as unequal crossing-over and/or gene conversion. Our data imply, however, that the base-pair diversity between repeats is rather low. This might be due either to selection or to the efficiency of the homogenization processes. As long as no selective mechanisms acting on satDNAs are known, we strongly favor the hypothesis of efficient homogenization processes.

Two apparent questions remain unresolved. If satDNA really tends to be as conservative during evolution as has been argued above, and if it is possible to trace the ATOC180 satDNA family back to the common ancestor of *D. obscura, D. ambigua,* and *D. tristis,* it should be possible to detect ATOC180 sequences in the closest relatives of these species—e.g., *D. bifasciata* and *D. subsilvestris*—as well. This raises the question, What initiates, regulates, and terminates the upheavels of genome evolution responsible for the exchange of an entire set of satDNA in an evolutionary short period of time? In this special case, how is it possible to explain the sudden appearance and amplification of the ATOC180 satDNA family after the cladogenetic splitting of the ancestors of *D. subsilvestris* and *D. bifasciata?* Of course, the same question may be raised the other way round: How did the ATOC180 satDNA family get substituted by another satDNA family in *D. subsilvestris* and *D. bifasciata,* if we assume that it was present in the common ancestor of all five species? If the formula $t \approx 24/c \ln(i_o/2)$ is applied (Stephan 1986), where c is the crossing-over rate per array per generation, i_o is the initial copy number of the ATOC180 satDNA family (10,000), and t is the time in generations, it is possible to estimate that there was sufficient time, since the cladogenesis of *D. subsilvestris* and *D. bifasciata,* for loss of the ATOC180 sequences in these species, under the forces of random genetic drift and unequal crossing-over alone. Even if we assume that there are only two generations per year and a rate of crossing-over as low as 2×10^{-11}/nucleotide/generation, the original ATOC180 satDNA arrays might be lost within ~5 Myr. Such calculations illustrate the possibility that the ATOC180 satDNA family got lost in *D. subsilvestris* and *D. bifasciata,* but they cannot explain why this should have happened and, furthermore, why this occurred only in these species and not in one of the triad of *D. ambigua, D. tristis,* and *D. obscura.* The existence of several species-specific or subgroup-specific satDNA fam-

ilies in other species of the *D. obscura* group—i.e., pGH290 of *D. guanche* (Bachmann et al. 1989), pTET181 of *D. tristis* (Bachmann et al. 1990), KM190 of *D. microlabis* and *D. kitumensis* (Bachmann et al. 1992), a species-specific satDNA of *D. bifasciata,* and another one of *D. subobscura* (M. Raab, personal communication)—seems to support the assumption that satDNA evolution proceeds by introducing new satDNA families by saltatory amplification replacing the ancient sequences, rather than by a slow destruction of arrays of ancient satDNAs. Therefore, we assume that the ATOC180 satDNA family arose after the separation of *D. subsilvestris* and *D. bifasciata* in the genome of a common ancestor of *D. ambigua, D. tristis,* and *D. obscura.*

The second question is, Why are there two types of ATOC180 (ATEco180 and AOPst180 sequences) repeats in the genome of *D. ambigua,* if the supposed homogenization processes are really as effective as proposed above? Although there is no experimental evidence, a localization on different chromosomes or different sections of chromosomes prohibiting or reducing nonhomologous crossing-over might be a plausible explanation. Since the example of 18 independent clones from *D. obscura,* which correspond to five chromosomes, shows no comparable splitting into a similar number of sequence clusters, this explanation seems insufficient. On the other hand, there is no evidence to take today's coexistence of the two subfamilies ATEco180 and AOPst180 in the genome of the extant *D. ambigua* as the final outcome of ATOC180 satDNA evolution. Since we do not believe in an exceptional mode of evolution of the ATOC180 satDNA family, we propose that the evolution of tandemly repeated satDNA is governed by the alternation of gradual and saltatory modes of evolution.

Acknowledgments

We thank E. Müller for excellent technical assistance during this work, and we thank D. Graur and J. Gutknecht for critical discussions and valuable comments on the manuscript. The work was supported by grant Sp 146/6 of the 'Deutsche Forschungsgemeinschaft.'

LITERATURE CITED

BACHMANN, L., E. MÜLLER, M. L. CARIOU, and D. SPERLICH. 1992. Cloning and characterization of KM190, a specific satellite DNA family of *Drosophila kitumensis* and *D. microlabis.* Gene **120**:267–269.

BACHMANN, L., M. RAAB, and D. SPERLICH. 1989. Satellite DNA and speciation: a species specific satellite DNA of *Drosophila guanche.* Z. Zool. Syst. Evol.-Forsch. **27**:84–93.

———. 1990. Evolution of a telomere associated satellite DNA sequence in the genome of *D. tristis* and related species. Genetica **83**:9–16.

BENEDUM, U. M., H. NEITZEL, K. SPERLING, J. BOGENBERGER, and F. FITTLER. 1986. Organization and chromosomal distribution of a novel repetitive DNA component from muntjak *Muntiacus vaginalis* with a repeat length of more than 40 kb. Chromosoma **94**:267–272.

BERIDZE, T. 1986. Satellite DNA. Springer, Berlin, Heidelberg, and New York.

CABRERA, V. M., A. M. GONZALEZ, J. M. LARRUGA, and A. GULLON. 1983. Genetic distance and evolutionary relationships in the *Drosophila obscura* group. Evolution **37**:657–689.

CACCONE, A., and J. R. POWELL. 1990. Extreme rates and heterogeneity in insect DNA evolution. J. Mol. Evol. **30**:273–280.

CARIOU, M. L., D. LACHAISE, L. TSACAS, J. SOURDIS, C. KRIMBAS, and M. ASHBURNER. 1988. New African species in the *Drosophila obscura* species group: genetic variation, differentiation and evolution. Heredity **61**:73–84.

CHARLESWORTH, B., C. H. LANGLEY, and W. STEPHAN. 1986. The evolution of restricted recombination and the evolution of repeated DNA sequences. Genetics **112**:947–962.

CREMISI, F., R. VIGNALI, R. BATISTONI, and G. BARSACCHI. 1988. Heterochromatic DNA in *Triturus* (Amphibia, Urodela). II. A centromeric satellite DNA. Chromosoma **97**:204–211.

DAVIS, L. G., M. D. DIBNER, and J. F. BATTEY. 1986. Basic methods in molecular biology. Elsevier, New York.

ECK, R. V., and M. O. DAYHOFF. 1966. Atlas of protein sequence and structure. National Biomedical Research Foundation, Silver Spring, Md.

FEINBERG, A. P., and B. VOGELSTEIN. 1983. A technique for radiolabeling DNA restriction fragments to high specific activity. Anal. Biochem. **132**:6–13.

FELGER, I., and W. PINSKER. 1987. Histone gene transposition in the phylogeny of the *Drosophila obscura* group. Z. Zool. Syst. Evol.-Forsch. **25**:127–140.

FELSENSTEIN, J. 1988. Phylogenies from molecular sequences: inference and reliability. Annu. Rev. Genet. **22**:521–565.

GANAL, M., I. RIEDE, and V. HEMLEBEN. 1986. Organization and sequence analysis of two related satellite DNAs in Cucumber (*Cucumis sativus L.*). J. Mol. Evol. **23**:23–30.

GONZALEZ, A. M., M. HERNANDEZ, A. VOLZ, J. PESTANO, J. M. LARRUGA, D. SPERLICH, and V. M. CABRERA. 1990. Mitochondrial DNA evolution in the *obscura* species subgroup of *Drosophila*. J. Mol. Evol. **31**:122–131.

GRAY, D. M., and D. M. SKINNER. 1974. A circular dichroism study of the primary structures of three crab satellite DNAs rich in AT pairs. Biopolymeres **13**:843–852.

HÖRZ, W., and W. ALTENBURGER. 1981. Nucleotide sequence of mouse satellite DNA. Nucleic Acids Res. **9**:683–696.

HOURCADE, D., D. DRESSLER, and J. WOLFSON. 1973. The amplification of ribosomal RNA genes involving a rolling circle intermediate. Proc. Natl. Acad. Sci. USA **70**:2926–2930.

HSIEH, T., and D. BRUTLAG. 1979. Sequence and sequence variation within the 1.688 g/cm^3 satellite DNA of *Drosophila melanogaster*. J. Mol. Biol. **135**:465–481.

KIMURA, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Biol. **16**:111–120.

KING, P. V., and R. W. BLAKESLEY. 1986. Optimizing DNA ligation for transformations. Focus **8**:1–3.

KLUGE, A. G., and J. S. FARRIS. 1969. Quantitative phyletics and the evolution of anurans. Syst. Zool. **18**:1–32.

LAKOVAARA, S., and L. KERÄNEN. 1980. Phylogeny of the *Drosophila obscura* group. Genetika **12**:157–172.

LOHE, A. R., and P. ROBERTS. 1988. Evolution of satellite DNA sequences in *Drosophila*. Pp. 148–186 *in* R. S. VERMA. ed. Heterochromatin. Cambridge University Press, Cambridge.

NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

OKUMURA, K., R. KIYAMA, and M. OISHI. 1987. Sequence analyses of extrachromosomal Sau3A and related family DNA: analysis of recombination in the excision event. Nucleic Acids Res. **15**:7477–7489.

PREISS, A., D. A. HARTLEY, and S. ARTAVANIS-TSAKONAS. 1988. Molecular genetics of enhancer of split, a gene required for embryonic neural development in *Drosophila*. Embo J. **12**:3917–3927.

ROWAN, R. G., and J. A. HUNT. 1991. Rates of DNA change and phylogeny from the DNA sequences of the alcohol dehydrogenase gene for five closely related species of Hawaiian *Drosophila*. Mol. Biol. Evol. **8**:49–70.

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

SANGER, F., S. MICKLEN, and A. R. COULSON. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.

SMITH, G. P. 1976. Evolution of repeated DNA sequences by unequal crossover. Science **191:** 528–535.

SNEATH, P. H., and R. R. SOKAL. 1973. Numerical taxonomy. W. H. Freeman, San Francisco.

SOUTHERN, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:**503–517.

STEPHAN, W. 1986. Recombination and the evolution of satellite DNA. Genet. Res. **47:**167–174.

———. 1989. Tandem-repetitive noncoding DNA: forms and forces. Mol. Biol. Evol. **6:**198–212.

TIMBERLAKE, W. E. 1978. Low repetitive DNA content in *Aspergillus nidulans.* Science **202:** 973–975.

VIGNALI, R., M. R. FILIPPO, R. BATISTONI, D. FRATTA, F. CREMISI, and G. BARSACCHI. 1991. Two dispersed highly repetitive DNA families of *Triturus vulgaris meridionalis* (Amphibia, Urodela) are widely distributed among Salamandridae. Chromosoma **100:**87–96.

WALSH, J. B. 1987. Persistance of tandem arrays: implications for satellite and simple sequence DNAs. Genetics **154:**553–567.

BRIAN CHARLESWORTH, reviewing editor

Received March 17, 1992; revision received December 30, 1992

Accepted January 6, 1993