The intron boundaries and flanking rRNA coding sequences of *Calliphora erythrocephala* rDNA

Vana L. Smith and Kathy Beckingham

Department of Biochemistry, Rice University, P.O. Box 1892, Houston, TX 77251, USA

Received 3 October 1983; Revised and Accepted 10 January 1984

ABSTRACT

We have sequenced the available cloned examples of the intron-coding sequence junctions for the rDNA of the higher Dipteran, *Calliphora erythrocephala*. The introns interrupt the rDNA at the same position as the type 1 intron family detected in *Drosophila melanogaster* and *D. virilis* (10,11). A duplication of 14 base pairs of the 28S rRNA coding sequence surrounds a short version of the major genomic length class of introns. This same duplication is associated with boundaries of the type 1 introns in *D. virilis* and *D. melanogaster* (10, 13,14). We have detected considerable homology between the 3' intron sequences of *C. erythrocephala* and *D. virilis*. The rRNA coding sequences flanking the introns are extremely homologous in *C. erythrocephala*, *D. melanogaster* and *D. virilis*, with only one small region of significant divergence. This corresponds to a variable stem region previously identified in eukaryotic 28S rRNA at a site analogous to the L1 ribosomal protein binding site of prokaryotic 23S rRNA (27).

INTRODUCTION

In most species of Dipteran fly examined, a significant fraction of the rDNA cistrons contain introns (1-7). Amongst intron-containing (intron⁺) genes, these Dipteran intron⁺ rRNA genes are exceptional in that, in all cases investigated, they are essentially non-functional (reviewed in 8). Gross positioning of the introns by EM and endonuclease S1 mapping techniques indicates that, in all species, they are located at a similar position within the 28S rRNA coding sequence (8). However, it is already apparent that these Dipteran rDNA introns do not have a single evolutionary origin. Thus, in two of the species investigated (*Drosophila melanogaster* (9) and *Sciara coprophila* (5)) it has been shown that two distinct, non-related sequence classes of introns may occupy this position with the 28S rRNA gene.

A limited number of cross-hybridization studies using standard Southern blot hybridization techniques have been performed to investigate the sequence-relatedness of the rDNA introns of some pairs of Dipteran species. Thus, some homology between the major sequence class of rDNA introns present

in D. melanogaster (the so-called type 1 intron family) and the single sequence class of introns present in the two species D. virilis and D. hydei has been established (3). No homology could be detected between the D. melanogaster type 1 introns and the cloned rDNA intron fragments of the more distantly related higher Dipteran, Calliphora erythrocephala (8).

To date, sequence data for the rDNA intron-coding sequence junctions are available for only two species of Diptera, D. virilis and D. melanogaster. The rDNA introns of D. virilis have thus been shown to be surrounded by a duplication of 14 base pairs (bp) of the 28S rRNA coding sequence (10). Studies from three laboratories have examined the junctions of the two sequence classes of D. melanogaster rDNA introns (11-14), and have confirmed the relatedness of the type 1 introns to the D. virilis sequences. Thus, the type 1 introns interrupt the rRNA coding sequence at precisely the same position in D. melanogaster as in D. virilis and the shorter forms (1.0 kb and 0.5 kb) of this intron sequence family in D. melanogaster are bordered by the same 14 bp duplication (or a derivative thereof) as the D. virilis introns. However, this 14 bp sequence is missing at the 5' boundary of the major length class (5 kb) of these sequences in D. melanogaster, together with an additional nine residues of the 28S rRNA coding sequence. No obvious homology between the extreme 3' intron sequences of the type 1 D. melanogaster introns and the D. virilis introns was revealed by the sequence data, although the Southern hybridization experiments mentioned above localized some of the cross-hybridization between the two species, close to the 3' termini of both introns (3). The type 2 D. melanogaster introns were found to interrupt the rDNA at a different location - 60 residues upstream from the type 1 interruption point - and to possess a very different boundary structure showing no duplication of rRNA coding sequences (11,14).

We report here DNA sequence studies for the rDNA intron-coding sequence junctions of the species C. erythrocephala. In this species, ~14% of the embryonic rDNA consists of a single sequence class of intron⁺ cistrons (15, 16) which are preferentially and severely under-replicated in polyploid tissues (15, and Belikoff and Beckingham, unpublished observations). This suggests a non-functional character for these genes as is indicated for all other Dipteran intron⁺ rDNA cistrons. The sequence studies presented here establish that the introns within the rDNA cistrons of C. erythrocephala are related to the type 1 sequences seen in D. melanogaster and D. virilis and suggest that this intron sequence family will prove to be the major class of such structures within the Diptera.

MATERIALS AND METHODS

The rDNA clones used in this study (pKB29, pKB31 and pKB33) have been described previously (16). To simplify sequencing, several subclones were prepared. The subclones pKB118 and pVS2 are Hind III-Kpn I fragments from pKB29 and pKB31, respectively, ligated to the large Hind III-Kpn I fragment of vector pKB111 (17). The remaining subclones, pVS3 and pVS12, are Hind III-Bgl II fragments from pKB31 and pKB33, respectively, ligated to the large Hind III-Bam HI fragment of vector pBR322 (18). The relevant restriction maps for these plasmids are shown in Figure 1.

The mapping technique of Smith and Birnstiel (19) was used to locate restriction sites suitable for DNA sequencing. DNA fragments were isolated from polyacrylamide gels by elution (20) or from agarose gels by either "freeze-squeeze" (21) or gel disruption followed by hydroxyapatite chromatography (22). DNA fragments were labeled at 3' termini by incubating 10 pmoles of DNA ends with 50 μ Ci of the appropriate α^{32} P-dNTP (2000-3000 Ci/mmol) and one unit of the Klenow fragment of DNA polymerase I in 5 mM Tris-HCl, pH 7.9, 0.5 mM MgCl₂, 1 mM β -mercaptoethanol and 5 μ g/ml bovine serum albumin at 20°C for thirty minutes. All fragments generated for DNA sequencing had termini produced by cleavage with two different restriction enzymes such that in the presence of the appropriate α^{32} P-dNTP only one of the two 3' termini would be labelled. All DNA sequencing was performed as described by Maxam and Gilbert (23). These reactions were supplemented on occasion with the osmium T reaction (24). For each DNA fragment, the sequencing data presented here was derived from at least two sets of sequencing chemistry and from multiple sequencing ladders such that various regions of each sequence shown were determined between 2 and 8 times. Any residues at the extreme termini of fragments which could not be unequivocally resolved, have been omitted.

Computerized DNA sequence analyses were performed using graphic matrix analysis programs developed by Maizel and Lenk (25) and Pustell and Kafatos (26) and further modified by Dr. C. Lawrence. All computer analyses were done on a genetic engineering work-station developed by Dr. C. Lawrence in the Department of Cell Biology, Baylor College of Medicine.

RESULTS

In the Clever strain of C. erythrocephala, a single sequence class of rDNA introns has been detected. There are however, several size variants of this sequence class. Data from Southern blot hybridizations have shown that

the major genome intron size class is 6.1 kb and also indicate the presence of a small number of longer, sequence-related introns (16).

The cloned examples of rDNA intron sequences from C. erythrocephala consist of i) a 4.9 kb segment from the 3' region of a 6.1 kb intron including the 3' intron-coding sequence junction (pKB29) and ii) a complete 2.8 kb intron and its adjacent 28S rRNA coding sequences (pKB31). As is the case for the short type 1 introns of D. melanogaster, this shorter intron contains a length of the 3'-most sequences of the major genome intron length class and, as a length class, is too rare in the total genomic rDNA for detection in Southern blots. With the aid of a cloned 28S rRNA coding region from an intron⁻ rDNA cistron (pKB33), the exact intron-coding sequence junctions within these two intron⁺ rDNA clones have been identified and sequenced. The subclones prepared for these studies and the sequencing strategies applied are shown in Figure 1.

1. 28S rRNA coding sequences surrounding the intron position

A small Sau 96-Hind III DNA fragment was identified within the intron⁻ subclone pVS12 which spans the point at which the introns interrupt the 28S rRNA coding region (see Figure 1). This fragment has been sequenced in its entirety yielding 440 bp of rRNA coding sequence as shown in Figure 2. This entire rRNA coding region has also been sequenced from the intron⁺ subclone pVS3 and found to be identical to the functional intron-free sequence.

The *Drosophila* rDNA introns are known to interrupt a region of the 28S rRNA coding sequence which is highly conserved in evolution (10,27) and not surprisingly therefore, a comparison of this rRNA coding region from C. erythrocephala to the equivalent sequences from D. virilis (11) and D. melanogaster (12) reveals overall, extremely high homology (Figure 2). An unexpected finding however, was that essentially all of the divergence between the three sequences is confined to one small region of 24 bp (residues +298 - +322) localized ~125 bp downstream of the type 1 intron interruption position (see Figure 2).

In comparing the highly conserved sequences of the rDNA introns in the two species Tetrahymena pigmentosa and Tetrahymena thermophila, Kan and Gall (28) found three small non-homologous regions and noted that each was flanked by inverted repeats. In the secondary structure deduced by Cech et al. (29) for the RNA derived from these introns, these non-homologous regions are mainly represented as divergent single-strand loops on conserved stems formed by the inverted repeats. This prompted us to perform a computer-assisted search for inverted repeats within the entire rRNA coding sequence

Figure 1. Subcloning and sequencing to identify the rDNA intron-coding sequence junctions of *C. erythrocephala*. pKBs 33, 31 and 29 are cloned chromosomal fragments of the rDNA containing respectively i) an intron 28S rRNA coding sequence ii) a 28S rRNA coding sequence containing a 2.8 kb intron and iii) 4.9 kb from the 3' end of the major genome length class of introns. The subclones prepared from these fragments are shown below their maps using a 10-fold expanded scale. Asterisks and arrows beneath the subclones indicate the start points, directions, and extents of DNA sequencing. Coding regions within all clones are shown as solid blocks and vector sequences as the terminal, open, discontinuous bars. Only those restriction enzyme sites used for subcloning or DNA sequencing are displayed. (Ba = Bam HI; Bg = Bgl II; Hd = Hind III; Hf = Hinf I; K = Kpn I; S = Sau 96)

Figure 2. Comparison of the 28S rRNA coding sequence which flanks the introns in three Dipteran species. The 28S rRNA coding sequence determined here for *C. erythrocephala* (C.e.) is compared to the equivalent sequences for *D. melanogaster* (D.m.) and *D. virilis* (D.v.) (references 11 and 10 respectively). Only changes from the *C. erythrocephala* sequence are shown with missing residues indicated by Δ . The brackets at +65 and +86 of the *D. virilis* sequence surround a region which has not been sequenced for this species. Long inverted repeats which flank the variable region +298 - +322 are indicated by arrows. The boxed region is the 14 bp junction sequence which is duplicated at both termini of the short *C. erythrocephala* introns (see text).

Figure 3. Two 3' intron-coding sequence junctions from *C. erythrocephala* rDNA. The 3' junctions of the short intron (pVS2) and an example of the major genome intron length class (pKB118) are shown. The 14 bp of rRNA coding sequence which are duplicated at the 5' boundary of the short intron (see Figure 4) are boxed. Residues of the 28S rRNA coding sequence are numbered as in Figure 2. The 3' (right-hand) intron-coding sequence junction is designated 0 R.

shown in Figure 2. This revealed that the only substantial pair of inverted repeats within the sequence does indeed flank the variable region +298 - +322 (see Figure 2). We then scanned other eukaryotic 28S rRNAs to determine whether this potential stem and loop is evolutionarily conserved, and discovered that this region of the 28S rRNA is the eukaryotic analogue of the L1 ribosomal binding site of prokaryotic 23S rRNA (30,31). The secondary structure of this region is reviewed in the Discussion.

2. The intron-coding sequence junctions

The 3' intron-coding sequence junctions of i) the long intron class (present in subclone pKB118) and ii) the short 2.8 kb intron (present in subclone pVS2) were identified by comparing sequences derived from these subclones with the pVS12 intron⁻ rRNA coding sequence. The DNA sequences at these 3' junctions are shown in Figure 3. As can be seen, both the point at which the intron sequences interrupt the rRNA coding sequence and the intron sequences 5' to the interruption point, are identical in the two subclones.

Figure 4. The 5' intron-coding sequence junction of a short rDNA intron of *C. erythrocephala*. Residues of the 28S rRNA coding sequence are numbered as in Figure 2. The 5' (left-hand) intron-coding sequence junction is designated 0 L. Residues +160 - +173 of the rRNA coding sequence (enclosed in box) are duplicated at this boundary (compare Figure 3).

Further, the interruption point is exactly the same as that found for all three length classes of the type 1 introns of *D. melanogaster* (11-14) and the single rDNA intron class of *D. virilis* (10).

The 5' intron-coding sequence junction of subclone pVS3 was similarly identified with the aid of the intron⁻ rRNA coding sequence and is shown in Figure 4. Comparison of this sequence to the 3' junctions shown in Figure 3 reveals that 14 bp of the rRNA coding sequence are duplicated at this 5' boundary. The same 14 bp of rRNA coding sequence are duplicated at the 5' and 3' intron-coding sequence junctions of 1) the *D. virilis* rDNA introns (10) and 1) the 0.5 kb short type 1 introns of *D. melanogaster* (13,14).

3. Comparison of Dipteran rDNA intron sequences at the 5' and 3' junctions

The 3' intron sequences derived from the *C. erythrocephala* rDNA were compared to the equivalent sequences for the *D. virilis* rDNA introns and the 3' intron sequences common to all three length classes of the type 1 introns in *D. melanogaster*. Since the 5' sequences we have presented here are derived from a shortened version of the major genomic length class of rDNA introns in *C. erythrocephala*, the most appropriate comparison for these sequences is to the 5' region of the two shorter length classes (approximately 1.0 kb and 0.5 kb) of the type 1 introns found in *D. melanogaster*. In addi-

Figure 5. Homology between the 3' rDNA intron sequences of *C. erythrocephala* and *D. virilis*. The alignment shown maximizes homology between the two sequences and renders 74% of the *C. erythrocephala* sequence and 68% of the *D. virilis* sequence homologous. The 14 bp junction sequences at the 3' intron coding sequence junctions is underlined and the 3' intron sequences of *C. erythrocephala* are numbered as in Figure 3. Residues identified as resulting from nucleotide changes are displaced one line from the main sequences. Residues designated as resulting from additions or deletions are displaced at least two lines from the main sequences: 65% of the non-homologous residues are thus designated as the products of addition or deletion events.

tion, however, these 5' intron sequences were also compared to the 5' sequences of the major type 1 intron length class of D. melanogaster and to the 5' intron sequences of D. virilis. The 5' intron sequences of i) D. virilis and the 5 kb type 1 intron length class of D. melanogaster and ii) the 1.0 kb and 0.5 kb short type 1 D. melanogaster introns were also compared.

Junction sequences for two examples of the 0.5 kb length class of D. melanogaster type 1 introns have been sequenced and found to differ in that one contains an additional 28 bp of the major 5 kb length class sequences at the 5' boundary (13,14). Both of these sequences were used for comparisons. For homology studies using the D. virilis introns, additional intron sequence data to that already published was kindly provided by P. M. M. Rae (personal communication).

Assuming a non-coding and probably totally non-functional character for these intron sequences, it is to be expected that some of the sequence divergence experienced in the different species would involve addition and deletion of residues. Thus the sequence matching strategy adopted was to perform computer-aided searches for regions of perfect or very high sequence homology located at approximately the same positions with respect to the intron-coding sequence junctions and then to attempt further alignment of more divergent intervening regions by inspection.

Of these various comparisons only one pair of intron sequences (the 3' intron sequences of C. erythrocephala and D. virilis) showed sufficient perfect matching to permit alignment throughout the entire regions examined. Figure 5 shows an alignment which maximizes homology between these sequences and yields a high overall homology of 71%. This alignment, in which most of the non-homologous residues appear as additions or deletions (see Legend, Figure 5) suggests that such changes are the major source of sequence divergence between the two species, as opposed to nucleotide substitutions. The sequence conservation is more pronounced towards the intron-coding sequence junctions with a 20 bp stretch of perfect homology located approximately 15 residues from the junction proper.

Much smaller regions of sequence homology were present between all other pairs of sequences examined which made unambiguous alignments impossible. In an attempt to assess the significance of the small amounts of sequence homology detected, the following parameter was quantitated for each of the sequence pairs examined. Using the intron-coding sequence junctions as positional reference points and stretches of homology no smaller than 6

Table 1. Quantitation of homology between the terminal rDNA intron sequences of three Diptera.

INTRON ^{a)} SEQUENCE PAIR	NUMBER OF INTRON RESIDUES COMPARED	PERCENTAGE OF SEQUENCES ^{b)} SHOWING 80% HOMOLOGY AT SAME RELATIVE POSITIONS (+/-25 bp)
D.m. (major) 3'	160	19%
D.v. (major) 3'	160	
C.e. (major) 3'	119 ^{c)}	59%
D.v. (major) 3'	130	54%
C.e. (major) 3'	119	22%
D.m. (major) 3'	119	
D.m. (major) 5'	160	25%
D.v. (major) 5'	160	
C.e. (2.8 kb) 5'	134	5%
D.m. (major) 5'	134	
C.e. (2.8 kb) 5'	134 ^{c)}	25%
D.v. (major) 5'	150	23%
C.e. (2.8 kb) 5'	134	8%
D.m. (0.5 kb) 5' ^{d)}	134	
C.e. (2.8 kb) 5'	134	19%
D.m. (1.0 kb) 5'	134	
D.m. (0.5 kb) 5' ^{d)}	160	19%
D.m. (1.0 kb) 5'	160	

Using a graphic matrix analysis program, regions of homology of no fewer than 6 bp and no lower than 80% homology were localized within each of the intron sequence pairs indicated. Those showing no greater than a 25 bp positional difference relative to the intron-coding sequence junctions were examined further. Some of these homologies were incompatible in that i) overlapping regions of one sequence appeared homologous to two different regions in the second sequence or ii) two different homologies within a given sequence pair showed a different 5' - 3' order in the two sequences. In case i), the longer of the two overlapping homologies was chosen. In case ii), that set of homologies showing the same 5' to 3' order within the two sequences which gave greatest overall sequence match was used.

a) Each sequence is designated by i) the species of origin, ii) the intron length class from which it derives (in parentheses) and iii) the terminal region of the intron (3' or 5') compared. D. virilis sequences are from Reference 10 and P.M.M. Rae personal communication. D. melanogaster sequences from References 11, 13 and 14.

b) The number of residues of each sequence present in homologies of the type described above was determined and a percentage homology was then calculated.

c) Additional residues of the D. virilis intron were used to include a homology localized slightly further from the junction.

d) Sequences from References 13 and 14 for this intron length class from D. melanogaster were both used for this comparison. The sequence used here (which shows higher homology using these criteria) is from Reference 13.

bp in length, the percentage of each sequence pair which could be matched at greater than 80% homology using homologies showing no greater than a 25 bp positional difference, was calculated (see Legend, Table 1 for further details). This parameter was chosen since it selects for high homology whilst permitting positional drift due to divergence by additions and deletions. The results of this homology quantitation are shown in Table 1. The level of homology seen between sequences not predicted to have a common evolutionary origin (such as the 5' intron sequences of D. virilis and the shortened C. erythrocephala intron) suggests that homology up to at least 25% as judged by this parameter is fortuitous and thus that only homology at considerably greater than this level should be considered indicative of sequence similarity and conservation. As can be seen, all sequence pairs examined show homology at or below this 25% level with the exception of the 3' sequences of C. erythrocephala and D. virilis which show much greater homology (57%).

In addition to these paired comparisons, which probed for overall sequence homology, a search for small stretches of sequence conservation similarly positioned in the 5' regions of all three forms of shortened introns (the two short type 1 intron classes of D. melanogaster and the 2.8 kb intron of C. erythrocephala) was performed. If found such sequences might implicate a common site-specific deletional process in the generation of these structures - but no convincing small homologies of this type were detected.

DISCUSSION

1. Variable region +298 - +322 of the Dipteran 28S rRNA coding regions

Variable region +298 - +322 and its flanking inverted repeats correspond to a domain identified within Dictyostelium discoideum and Xenopus laevis 28S rRNA as the eukaryotic equivalent of the prokaryotic L1 ribosomal protein binding site (30). Although slightly different secondary structures were initially proposed for this region in X. laevis (30) and Saccharomyces carlsbergensis (yeast) (32) rRNA, current consensus (S. Gerbi, personal communication) is that the yeast structure, which is stabilized by an additional 5 bp stem (see Figure 6), is the more plausible model. Figure 6 shows this region of the C. erythrocephala 28S rRNA arranged in a secondary structure based on that derived for X. laevis, with the addition of the 5 bp stem proposed in the yeast structure. Variable region +298 - +322 is not single-stranded, as we initially expected, but rather forms a terminal stem struc-

Figure 6. Possible secondary structure for the analogue of the L1 binding site in Dipteran 28S rRNA. Residues +261 - +356 (see Figure 2) of the *C. erythrocephala* 28S rRNA arranged in a secondary structure based on that proposed for *X. laevis* (see Figure 6, Reference 27) with the addition of a central 5 bp stem (**AGUGG/UCACC**) proposed within this region for *S. carlsbergensis* rRNA (see Figure 4D, residues 2433 - 2510, Reference 29). The lower 17/18 bp stem corresponds to the most perfectly matched regions of the inverted repeats detected by us (residues +261 - +277 and +339 - +356, see Figure 2). Variable region +298 - +322 (see text) lies within the upper stem (bold type). The stem structures generated by the *D. melanogaster* and *D. virilis* versions of this region are also shown. The arrow indicates a highly conserved site known to be T1 ribonuclease sensitive and therefore single-stranded in *E. coli* 23S rRNA (33,34).

ture. This terminal stem region has already been demonstrated to vary considerably in length in different species and to be very truncated in *D. discoideum* and *S. carlsbergensis* rRNA (30,32). As is shown in Figure 6, all three Dipteran versions of this region will form a reasonable stem at this position, despite being very different in sequence.

L1 is thought to be a regulatory ribosomal protein with the dual function of control of ribosome assembly and regulation of the synthesis of

several ribosomal proteins (30,31). Assuming a eukaryotic equivalent for this protein, this comparison of three Dipteran versions of its binding site within the rRNA emphasizes the lack of constraint on primary sequence within one region of this site. However, in demonstrating the tendency to preserve a base-paired stem structure at this location, this study perhaps suggests some function for this variable region.

2. The type 1 intron family and the origin of short type 1 rDNA introns

The sequence data presented here demonstrate that the rDNA introns of C. erythrocephala are of the type 1 sequence family already characterized in D. melanogaster and D. virilis. In each of these genomes a major length class of these intron sequences (5-6 kb) predominates within the rDNA. Minor length variants in D. virilis appear to be mainly perfect tandem duplications of the intron sequences which could be generated by unequal crossing-over (35), and in C. erythrocephala there is also evidence for a small percentage of tandem duplications of the rDNA introns (8,16). However, the single cloned length variant isolated from C. erythrocephala is shorter than the major genome class and appears to have a similar structure to the only type of length variants isolated from D. melanogaster.

D. virilis is more closely related to the primitive progenitor of the genus *Drosophila* than D. melanogaster (see below). In D. virilis, the boundary structure for the major genome length class of type 1 sequences is very reminiscent of that seen for translocatable elements both in eukaryotes and prokaryotes (10), in that the intron sequences are flanked by a duplication of 14 bp of the adjacent rRNA coding sequence (which we will refer to as the junction sequence). In D. melanogaster, the junction sequence is missing at the 5' boundary of the 5 kb major length class of intron sequences together with an additional 9 bp of the rRNA (11). Viewed as a whole these findings suggest therefore that the type 1 intron sequences were introduced into the Dipteran lineage by a transposition event which resulted in duplication of the host DNA at the integration site and further, that these sequences have been modified during subsequent evolution by several mechanisms in the different Dipteran genomes resulting in duplications and both small and large deletions of the intron sequences.

In D. melanogaster the 0.5 kb shortened type 1 introns are bordered by the perfect duplication of the junction sequence seen in D. virilis and the 1.0 kb type 1 class are flanked by a duplication of a slightly modified form of this sequence. The data presented for the short intron length class of C. erythrocephala establishes that its boundary structure is identical to

that of the 0.5 kb type 1 introns of D. melanogaster in that the 3' intron-coding sequence junction of the major genome length class is perfectly maintained and the 14 bp junction sequence is perfectly duplicated at the 5' boundary. Considered in terms of the structure and evolution of the type 1 intron sequences suggested above, the retention of the junction sequence at the 5' boundaries of these deleted introns in these two different species suggests that a similar mechanism for generating these deleted forms is operating in the two species and that this mechanism involves the nucleotide residues of the 14 bp junction sequence.

Rae (12) and Roiha and Glover (13) have examined the points at which the 1.0 kb and 0.5 kb short introns terminate within the 5 kb (major length class) type 1 introns of D. melanogaster and have found that short homologies to the 14 bp junction sequence exist at these positions. Both laboratories have therefore postulated mechanisms (reviewed in Reference 8) for the generation of the short introns which involve recombination between the junction sequence and these short homologies. Each of the models proposed has the problem that a class of rDNA introns composed of the 5' sequences missing from the 1.0 kb and 0.5 kb introns should be at least as common within the rDNA as these 3' sequence introns themselves, whereas no such intron classes have ever been detected within D. melanogaster rDNA.

3. Type 1 rDNA intron sequences within the Diptera

rDNA introns have been detected in two species from the most primitive of the three Dipteran suborders, the Nemotocera, (Sciara coprophila (5) and Rhynchosciara americana (7)) and in five species of the most recently evolved suborder, the Cyclorrapha (D. virilis (3), D. melanogaster (1), D. hydei (2), C. erythrocephala (4), and Sarcophaga bullata (6)). One Nematocerous species (S. coprophila (5)) and one Cyclorraphous species (D. melanogaster (9)) have been shown to possess two sequence classes of rDNA introns.

Barnett and Rae (3) hybridized cloned type 1 intron sequences from D. melanogaster to genomic DNA of Diptera from all three suborders and found cross-hybridization only to those species within the suborder Cyclorrapha. Species from both the calyptrate (Musca domestica, S. bullata) and acalyptrate (D. virilis, D. hydei) sections of the Cyclorrapha were found to possess type 1 homologous sequences. They therefore suggested that the type 1 intron sequence family arose within a primitive Dipteran genome around the time of the evolution of the Cyclorrapha and before the appearance of the second antennal segment groove diagnostic of the calyptrates.

C. erythrocephala is a calyptrate, Cyclorraphous Dipteran (see Figure

Figure 7. The evolutionary relationships of *D. virilis*, *D. melanogaster*, and *C. erythrocephala* and the homology in the extreme 3' sequences of their type 1 rDNA introns. The Genera *Drosophila* and *Calliphora* have arisen from two distinct evolutionary pathways within the Division Schizophora, one (*Drosophila*) from the ancient acalyptrate section and one (*Calliphora*) from the more recently evolved calyptrate branch. *D. virilis* is one of the most primitive species of the genus *Drosophila* and therefore can be considered somewhat closer to *C. erythrocephala* in evolutionary distance than *D. melanogaster*. Percentage homologies within the pairs of 3' rDNA intron sequences are derived from Table 1.

7) and thus the sequence data that we present here confirms Barnett and Rae's finding that these sequences are prevalent in the genomes of Cyclorraphous Diptera. However, our results also indicate that their prediction concerning the absence of type 1 sequences in lower Dipteran genomes may need reappraisal. We have found previously that cloned DNA fragments representing the entire type 1 rDNA introns of *D. melanogaster* do not hybridize under standard Southern hybridization conditions to either genomic DNA of *C. erythrocephala* or cloned fragments of the *C. erythrocephala* rDNA introns. Similarly our sequence data show no homology between the intron sequences proper of these two species and our recognition of their common evolutionary origin rests upon the shared intron-coding sequence boundary structure and the obvious homology between the *C. erythrocephala* intron sequences and those of *D. virilis*. Thus the absence of this type 1 sequence family from a

particular genome cannot be inferred from homology studies of the type performed by Barnett and Rae, particularly when type 1 sequences representative of only a single species are used for cross-hybridizations. It is possible therefore, that the type 1 intron family entered the Dipteran lineage at a much earlier evolutionary time and may still be present in the rDNA of such Nematoceros species as S. coprophila and R. americana.

Figure 7 delineates the evolutionary relationships between the three Diptera considered in this study and summarizes our findings concerning sequence homologies within the extreme 3' regions of their type 1 family rDNA introns. The 3', as opposed to the 5', intron-coding sequence junction appears to be stable in all three of these species (see above) and therefore the intron sequences in this region are more likely to have derived from the same initial sequence. The much greater homology seen between the 3' intron sequences of C. erythrocephala and D. virilis than between those of D. virilis and D. melanogaster is surprising since far greater evolutionary distance exists between the calyptrates and the acalyptrates than between two species of the same genus (see Figure 7). However, it is not known whether the entire rDNA introns of C. erythrocephala and D. virilis are more homologous than those of D. virilis and D. melanogaster since cross-hybridizations between the cloned rDNA introns of D. virilis and C. erythrocephala have not yet been performed. The greater homology between the 3' intron sequences of C. erythrocephala and D. virilis as compared to C. erythrocephala and D. melanogaster is less surprising in that the species of the virilis subgroup are amongst the most primitive of the genus Drosophila (36) and therefore can be presumed to be closer to the common ancestor of the calyptrate and acalyptrate Cyclorraphous Diptera.

ACKNOWLEDGEMENTS

We are grateful to Drs. W. Beatty and L. Rothblum for technical hints and generous gifts of ³²P-labelled deoxynucleotides; Drs. P. Rae and S. Gerbi for communication of their unpublished data and helpful discussion and Dr. C. Lawrence for the generous use of his computer facilities and DNA sequence programs. We thank Ann Valverde for her careful preparation of the manuscript and figures. This work was supported by NIH grants GM 28379 and HD 17688 and grant C-848 from the Welch Foundation.

REFERENCES

1. Glover, D. M., White, R. L., Finnegan, D. J. and Hogness, D. S. (1975) Cell 5, 149-157.

2. Renkawitz-Pohl, R., Glatzer, K. H. and Kunz, W. (1980) *Nucleic Acids Res.* 8, 4593-4611.
3. Barnett, T. and Rae, P. M. M. (1979) *Cell* 16, 763-775.
4. Beckingham, K. and White, R. L. (1980) *J. Mol. Biol.* 137, 349-373.
5. Renkawitz-Pohl, R., Matsumoto, L. and Gerbi, S. (1981) *Nucleic Acids Res.* 9, 3747-3764.
6. French, C. K., Fouts, D. L. and Manning, J. E. (1981) *Nucleic Acids Res.* 9, 2563-2576.
7. Zaha, A., Leoncini, O., Hollenberg, C. P. and Lara, F. J. S. (1982) *Chromosoma* 87, 103-116.
8. Beckingham, K. (1982) in *The Cell Nucleus*, Busch, H. and Rothblum, L., Eds., Vol. X part A, pp. 205-269, Academic Press, New York.
9. Wellauer, P. K., Dawid, I. B. and Tartof, K. D. (1978) *Cell* 14, 269-278.
10. Rae, P. M. M., Kohorn, B. D. and Wade, R. P. (1980) *Nucleic Acids Res.* 8, 3491-3504.
11. Roiha, H., Miller, J. R., Woods, L. C. and Glover, D. M. (1981) *Nature* 290, 749-753.
12. Rae, P. M. M. (1981) *Nucleic Acids Res.* 9, 4997-5011.
13. Roiha, H. and Glover, D. M. (1981) *Nucleic Acids Res.* 9, 5521-5532.
14. Dawid, I. B. and Rebbert, M. L. (1981) *Nucleic Acids Res.* 9, 5011-5022.
15. Beckingham, K. and Thompson, N. (1982) *Chromosoma* 87, 177-196.
16. Beckingham, K. (1981) *J. Mol. Biol.* 149, 141-169.
17. Beckingham, K. (1980) *Plasmid* 4, 354-356.
18. Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. H. and Falkow, S. (1977) *Gene* 2, 95-113.
19. Smith, H. O. and Birnstiel, M. L. (1976) *Nucleic Acids Res.* 3, 2387-2399.
20. Munjaal, R. P., Chandra, T., Woo, S. L. C., Dedman, J. R. and Means, A. R. (1981) *Proc. Natl. Acad. Sci. USA* 78, 2330-2334.
21. Thuring, R. W. J., Sanders, J. P. M. and Borst, P. (1975) *Anal. Biochem.* 66, 213-220.
22. Tracy, S. (1981) *Prep. Bioc.* 11, 251-268.
23. Maxam, A. M. and Gilbert, W. (1980) in *Methods in Enzymology*, Grossman, L. and Moldave, K. Eds., Vol. 65 pp. 499-560, Academic Press, New York.
24. Friedmann, T. and Brown, D. M. (1978) *Nucleic Acids Res.* 5, 615-622.
25. Maizel, J. V., Jr. and Lenk, R. P. (1981) *Proc. Natl. Acad. Sci. USA* 78, 7665-7669.
26. Pustell, J. and Kafatos, F. C. (1982) *Nucleic Acids Res.* 10, 4765-4782.
27. Gourse, R. L. and Gerbi, S. A. (1980) *Nucleic Acids Res.* 8, 3623-3638.
28. Kan, N. C. and Gall, J. G. (1982) *Nucleic Acids Res.* 9, 2809-2822.
29. Cech, T. R., Tanner, N. K., Tinoco, I., Weir, B. R., Zuker, M., and Perlman, P. S. (1983) *Proc. Natl. Acad. Sci. USA* 80, 3903-3907.
30. Gourse, R. L., Thurlow, D. L., Gerbi, S. A. and Zimmerman, R. A. (1981) *Proc. Natl. Acad. Sci. USA* 78, 2722-2726.
31. Branlant, C., Krol, A., Machatt, A. and Ebel, J.-P. (1981) *Nucleic Acids Res.* 9, 293-307.
32. Veldman, G. M., Klootwijk, J., de Regt, V. C. H. F. and Planta, R. J. (1981) *Nucleic Acids Res.* 9, 6935-6952.
33. Branlant, C., Krol, A., Machatt, M. A., Pouyet, J., Ebel, J.-P., Edwards, K. and Kössel, H. (1981) *Nucleic Acids Res.* 9, 4303-4324.
34. Noller, H. F., Kop, J., Wheaton, V., Brosius, J., Gutell, R. R., Kopylov, A. M., Dohme, F., Herr, W., Stahl, D. A., Gupta, R. and Woese, C. R. (1981) *Nucleic Acids Res.* 9, 6167-6189.
35. Rae, P. M. M. (1982) *Nature* 296, 579-581.
36. Patterson, J. T. and Stone, W. S. (1952) *Evolution in the Genus Drosophila*, pp. 45-111, MacMillan, New York.