Promoter variation in the ribosomal RNA genes in *Drosophila melanogaster* strains

Alec Mian and Gabriel A.Dover
Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK

Received April 10, 1990; Revised and Accepted May 30, 1990

ABSTRACT

The sequences of thirty *D. melanogaster* ribosomal DNA promoter regions have been determined. Fifteen of these were isolated from a wild population recently captured in North Wootten, England. The other fifteen were isolated from an inbred laboratory strain. The overall level of variation is almost twice as high in the North Wootten strain as in the inbred laboratory strain. Two mutations at nucleotides −17 and −21 relative to transcription start, fall directly within a region known to be transcriptionally important. The sequences are also compared to eight previously published sequences from another *D. melanogaster* strain, Oregon R. Two of these eight clones have a −17 mutation identical to the one found in this study, suggesting that this polymorphism is widespread. Strikingly, all eight of these clones carry two single base pair changes not found in any of the other thirty clones, indicating the extent with which promoter variants can be homogenized and fixed in a population. Polymorphisms show different levels of homogenization within the rDNA unit spacer repeats or between different arrays depending on the location of the polymorphism. This has implications for the evolution of the observed species-specific transcription of ribosomal RNA genes.

INTRODUCTION

In eukaryotic organisms ribosomal RNA genes (rDNA) comprise a large multigene family that is found tandemly arrayed on two or more chromosomes (1). As in many multigene families, the rDNA genes are subject to a variety of turnover mechanisms that lead to a phenomenon termed concerted evolution. This is the observation that individual members of a multigene family do not evolve independantly of each other but in a concerted manner so that all members within a species share species-specific mutations (2,3,4). Several different mechanisms are thought to be responsible for homogenizing mutations within the gene family and ultimately fixing them throughout the population. One is unequal crossingover (5), both at the level of the entire rDNA unit and at the level of the tandemly arrayed subrepeats within the intergenic spacer (IGS). This can occur both within a chromosome (6) and between homologous and nonhomologous chromosomes (7,8,9).

One result of concerted evolution is the lack of promoter sequence conservation between closely related species, for example between the two sibling species *Drosophila melanogaster* and *D. virilis* (10, 11). This lack of conservation is functionaly relevant, since rDNA transcription with respect to these species is species-specific (12). Such divergence in not easily explained in traditional evolutionary terms which invoke only the action of selection. Promoter regions interact with RNA polymerase I (pol I) and transcriptional cofactors which play a critical role in rDNA transcription regulation and initiation and might be expected to be prime candidates for a high degree of conservation. It has been proposed that transcriptional functionality can be maintained in the face of promoter sequence divergence by the coevolution of pol I and/or it's cofactors (2,13). A species may initially tolerate the spread of new and possibly non-functional promoter variants because of the continued presence of functional promoters within the family. With time, the presence of these variants may persist or they may possibly be eliminated. More rareiy, the gene(s) for the transcriptional cofactor(s) may undergo mutation resulting in functional compatibilty with the new promoter variants. Hence, the new variant may continue spreading through the population, ultimately leading to species-specific transcription.

Given this possibility, it is important to assess the level of variation around the promoter region within a species. Previously the sequences of eight rDNA promoter regions from the Oregon R. strain of *D. melanogaster* were determined (14,15). However, Oregon R. is an inbred laboratory strain which may show only limited genetic variation relative to a wild population of *D. melanogaster*. Nevertheless, two of the eight clones were polymorphic at position −17 relative to transcription start. This position falls well within the promoter region known to be critical for transcripton (16,17).

To further assess the extent of the polymorphism at position −17 and other positions, and to investigate the possibility of strain-specific variation that may not have been apparent from the sequences of just one strain, we have cloned and sequenced fifteen promoter regions from an inbred laboratory strain and an additional fifteen from a recently captured wild strain of *D. melanogaster*.

MATERIALS AND METHODS

Drosophila Strains and Sources of Clones

The North Wootten (NW) wild strain was isolated from a North Wooten vineyard in Somerset, England by Dr. Donald Black (ICRF, London). Two hundred flies were isolated from various

locations throughout the vineyard. Strains were kept in a laboratory at 20°C. Less than 3 months (ca. 5 generations) elapsed from the time the NW flies were isolated to the time when DNA was extracted.

The inbred laboratory strain was obtained from a stock kept by the Department of Genetics (since 1977) and is referred to as UK.

The Oregon R. strain from which previous promoter clones were derived (14, 15) is referred to as OR. The names for these clones are λDmr 275, 241, 231, 214, 326, 317, 312, and 290. For simplification these clones are renamed in this study as OR 1 to 8, respectively. The Oregon R. strain was isolated in Oregon, U.S.A. in 1925 (18).

Clones A56 and Y22 are also isolated from Oregon R. (19, 20) but possibly not the same laboratory population as the above OR's, therefore they are dicussed separately from the OR clones. We are unsure of the population of *D. melanogaster* from which Dmr238 (Dr. David Glover, University of Dundee) came from. The sequences of Y22 and A56 have been previously determined (14,15). The sequence of Dmr238 has also been reported (11) and entered into the EMBL library. The published Dmr238 sequence differed at position −17 from the EMBL sequence. We resequenced this clone in this study and determined that the EMBL sequence, and not the published one (11), to be correct.

DNA Isolation

Approximately 250 flies were used for DNA extraction. The flies were first ground to a fine powder in liquid nitrogen. 4 mls of HTE (0.1M TRIS-HCL pH 7.6, 0.1M EDTA pH 7.6), 4 mls of 2% SDS and 20 μl of proteinase K (10 μg/ml) were added and incubated at 65°C for 30 minutes. The mixture became viscous indicating lysis. The lysate was extracted once with 8 mls of phenol:chloroform:isoamyl alcohol (24:24:1). The supernatant was transfered to another tube and 5 μl of RNAse A (10 μg/ml) was added. This was incubated at room temperature for 10 minutes. 20 μl of proteinase K was added and incubation continued for another 10 minutes, followed by another phenol:chloroform extraction and a chloroform extraction. DNA was percipitated by the addition of 16 mls of ethanol and spooled out with a pasteur pipette. The DNA spool was rinsed in 70% ethanol and resuspended in TE (10mM TRIS-HCL pH 7.6, 1mM EDTA pH 7.6).

rDNA Minilibrary Construction

Genomic DNA was digested using *Hin*dIII (Boeringer Manheim) and electrophoresed on a 1% agarose gel containing 5 μg/ml ethidium bromide. The gel was visualized with 300nm U.V. light, and a gel slice corresponding to 2.5 to 8.0 kb of the digested DNA was cut out with a razor blade and purified using a Geneclean kit (Bio 101, La Jolla). The purified *Hin*dIII DNA was then ligated into phosphatised Bluescript (Stratagene, San Diego) vector and transformed (21) into a TGI (Toby Gibson, MRC) strain of *E. coli*.

Library Screening

Minilibraries were colony lifted and screened using standard proceedures (22) and probed with nick translated rDNA spacer (from Dmr238).

Promoter Fragment Subcloning and Sequencing

Positive colonies were grown overnight and recombinant plasmid DNA isolated by standard alkaline lysis proceedure. The relevant promoter fragment was contained on a 468 bp *Dde* I fragment which was subcloned into phosphatised *Sma* I cut Bluescript vector. The promoter subclones were sequenced by dideoxy chain termination (23) and standard double stranded sequencing techniques and Sequenase (United States Biochemicals). Using universal and reverse primers both strands were sequenced.

RESULTS

A schematic representation of the *Drosophila melanogaster* rDNA unit is shown in figure 1. These units are known to be tandemly arrayed numbering approximately 250 units on each of the X and Y chromosomes. The IGS is polymorphic in size due to variant numbers of the 240bp and 330 subrepeats (8). The 240 bp repeats contain a 51 bp region termed spacer promoter (24, 26, 35), of high homology to the region surrounding the true promoter and are known to support transcription *in vitro* (25) and *in vivo* (26).

Two conserved *Hin*dIII sites in the 28S and 18S genes were used to isolate the IGS from the rest of the unit. To ensure true representation of all IGS size variants in the minilibraries, Southern blot analysis of *Hin*dIII genomic digests was performed prior to cloning. This revealed that all size variants for the North Wootten (NW) and inbred laboratory (UK) strain were greater than 2.5 kb but less than 8 kb (data not shown), hence this was the range of the size selection for the minilibraries (see Materials and Methods section).

For sequencing, a 468 bp promoter fragment resulting from a *Dde* I digest was chosen (Fig. 1). This spanned a region −358 nucleotides upstream of the transcription start site to +110 nucleotides downstream, and contained the first upstream spacer promoter (nucleotides −289 to −231, see top of Fig. 3). The sequence data from the 15 NW and 15 UK promoter clones is shown and compared to the sequence data from the previously sequenced OR (OR 1−8) clones and three other clones (A56, Y22, Dmr238), (Fig. 2).

The total number of mutations for all 41 NW, UK, OR and Dmr238, Y22 and A56 clones is 40. There are no deletions, but only substitutions and additions, of which there are 15 transversions, 11 transitions and 11 additions. We are assuming additions rather than deletions are the recent events, for they occur in a minority of clones.

Patterns of Mutant Distribution

Since populations evolve in a concerted manner, we may assume that the rate of mutation is less than the rate of spreading the variant throughout a multigene family. Hence, we may consider shared mutations as the result of spreading a single muational event, rather than the coincidental occurrence of the same mutational event in different units. This becomes apparent in the schematization of the distribution of patterns of sharing of mutations in the promoter regions as shown in figure 3. Significantly, only 6 of the 40 changes are single events, unique to a clone. This reduces the number of independently occurring mutations to 4 transversions and 7 transitions. This is significantly ($p < 0.05$) below the expected 2:1 transversion to transition ratio for noncoding sequences in *Drosophila* (27), on the assumption of equiprobability of mutations from any nucleotide to any other, and that all have an equal chance of being fixed.

The variation within a strain breaks down as follows: in NW, 10 individual events can be characterised into 2 types of shared events and 4 unique events. In UK, the 4 individual mutations

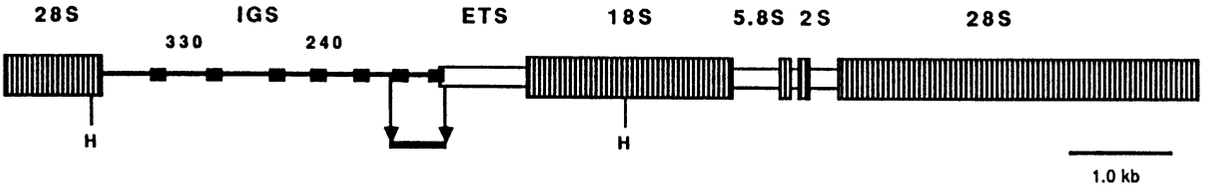

Figure 1. The rDNA unit in *D. melanogaster*. The intergenic spacer (IGS) may be partitioned into regions of 330 and 240 bp repeats which contain a 51 bp region (indicated by the small black box) that is homologous to a region in the true promoter. Transcription start is from the beginning of the externally transcribed spacer (ETS). The ETS and the internally transcribed regions are indicated by open boxes. The rRNA genes are represented by hatched boxes. The two conserved *Hind* III (H) sites used in the cloning strategy are also shown. The arrows and thick line indicate the *Dde* I promoter fragment that was sequenced.

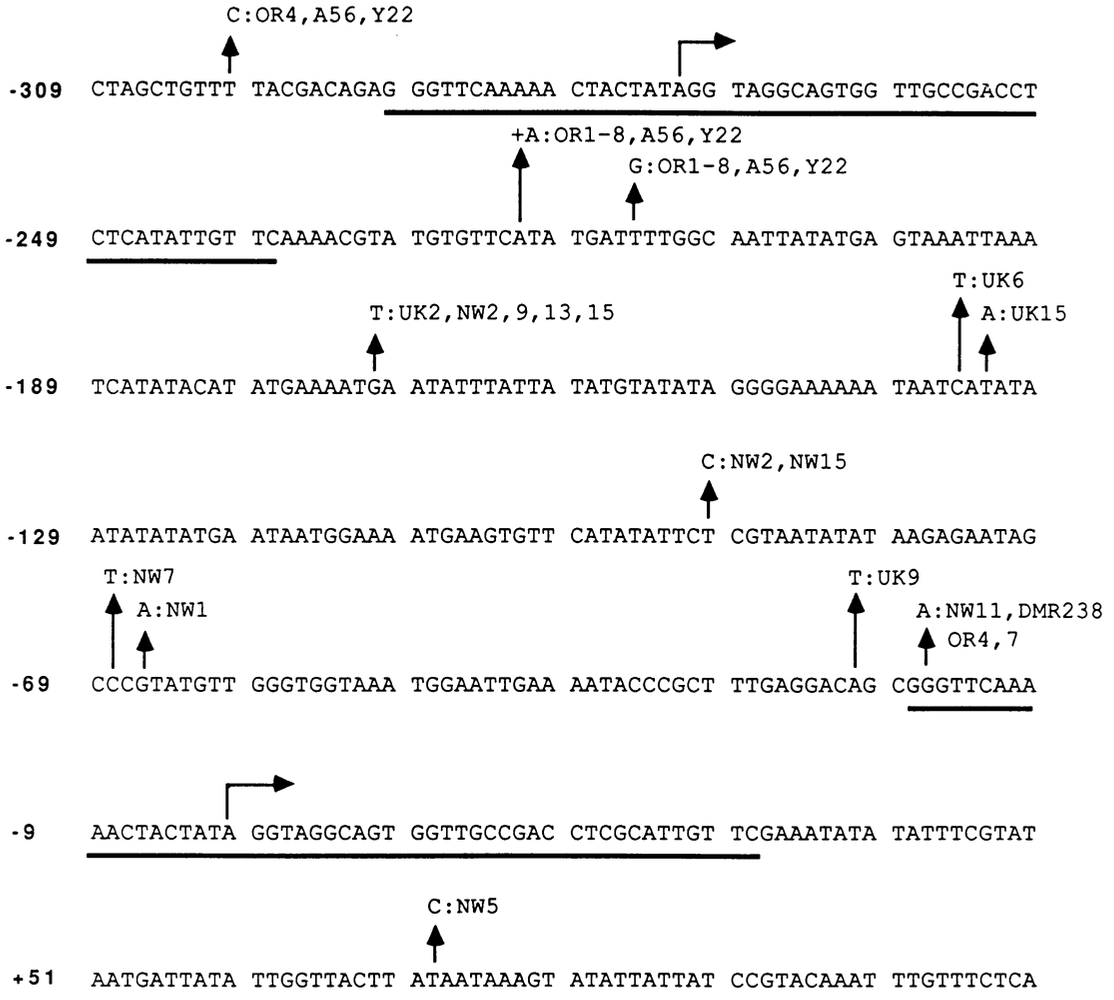

Figure 2. Nucleotide sequence within the *DdeI* fragment is shown. Only sequences from position −309 to +110 are shown since there were no changes outside of this region. The horizontal arrow indicates the transcription start site. The 51 bp region around the true promoter that is repeated in the first upstream spacer promoter is underlined. All UK, NW, OR, A56, Y22 and Dmr 238 clones are compared and the nucleotides changes indicated with vertical arrows and the relevant clones in which the change were found.

are all unique, however one mutation (G→T at −171) is shared among several NW clones. In OR, 19 individual mutations can be categorized into 3 types of shared events and one unique event. No mutations that are shared within a strain are shared by all three strains. Two mutations are shared by two of the three strains: UK and NW share G→T at −171, NW and OR share G→A at −17. The two inbred laboratory strains OR and UK are the only strains that do not share any mutations with each other.

All of the OR clones share two mutations within the strain, an additional A at position −222 and a T→G at position −215. The absence of these mutations from any of the NW or UK clones suggests that they are strain specific.

The three clones of uncertain origin (Dmr238, A56, Y22) have 7 mutations of 4 types, 3 shared and one unique. Both Y22 and A56 have a T→C substitution at position −300. This mutation is also present in OR4. Y22 and A56 also share the two strain specific mutations (an additional A at position −222 and a T→G

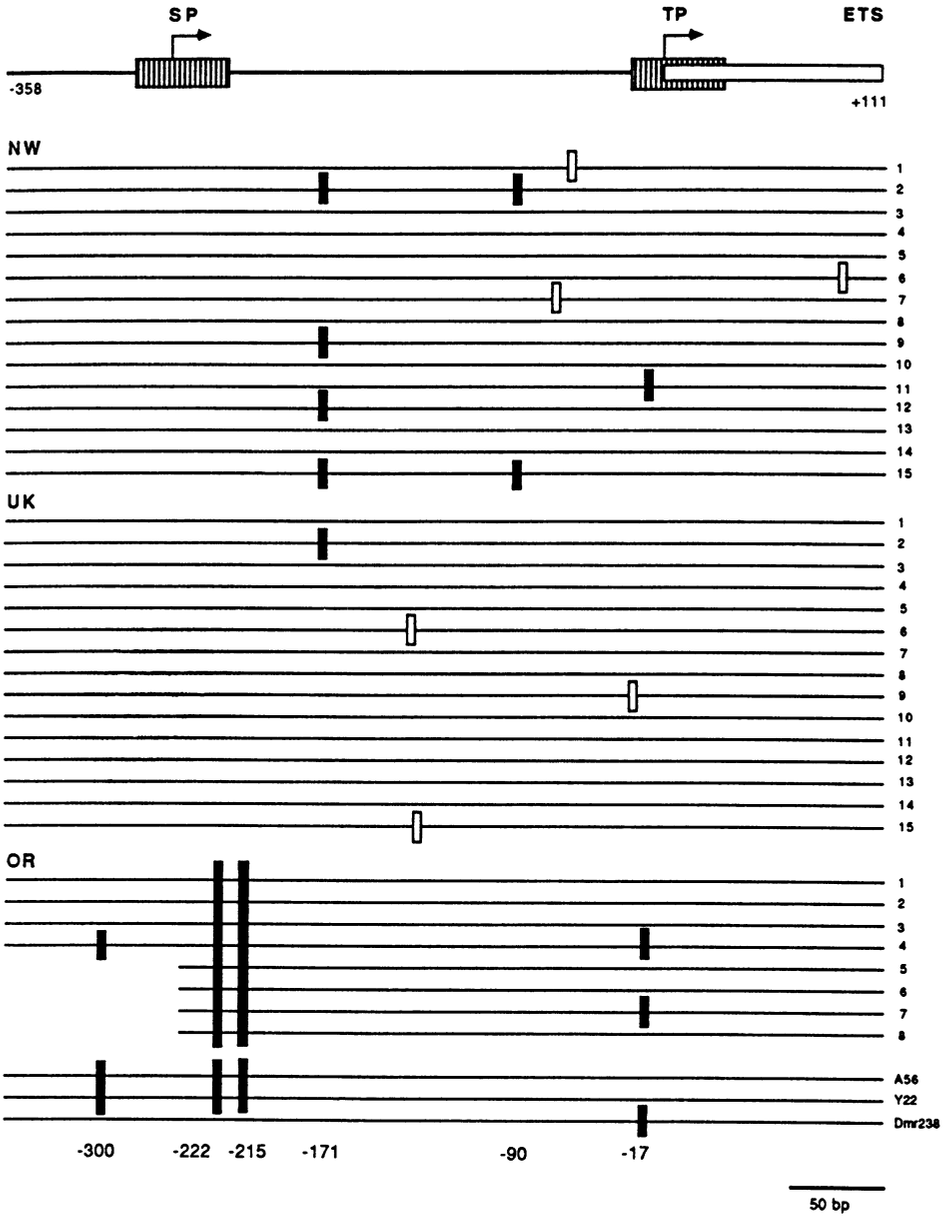

Figure 3. Schematization of promoter regions and the nucleotide changes that were found in them. At the top, a diagram of the promoter fragment indicates the 51 bp region of homology in the true promoter (TP) and the spacer promoter (SP) with hatched boxes. Arrows indicate transcription start. The external transcribed spacer (ETS) is indicated by an open box. Below, all of the clones are compared. Open boxes indicate nucleotide changes found in only one clone and black boxes indicate nucleotide changes found in two or more clones. The nucleotide positions of the shared mutations are numbered at the bottom.

at position −215) which are also found in all of the OR clones suggesting that they are indeed from Oregon R.. Dmr238 has the −17 G→A substitution that is found in one NW11 clone as well as OR4 and 7.

Intrestingly, although four G→A substitutions at position −17 fall within the true promoter (TP), which is repeated in the upstream spacer promoter (SP), the mutations themselves are not repeated.

DISCUSSION

Intraspecies Variation

In this study we attempt to approach the question of how species-specific rDNA promoters may have evolved in *D. melanogaster* by using comparative sequence analysis. The spread of variants throughout a multigene family has been well studied and is thought to be primarily due to unequal crossingover. Here we determine whether the multicopy nature of the rDNA gene family may allow a high level of variation in the promoter region, and how this variation may spread between true promters in other units and between the upstream spacer promoters within the same unit.

In Table 1 we compare the level of sequence variation in NW, UK and OR to that found between other coding and non-coding sequences. The highest level of variation in the *D. melanogaster* strains is 1.26% as found in the wild strain, NW. Both inbred laboratory strains UK and OR, have only 0.84% of polymorphic sites. Also, neither of these inbred strains share any mutations,

Table 1. Comparison of levels and type of variation found in this study with that in other sequences. NW, UK and OR columns show the level of promoter region variation found in this study. In the *D. melanogaster* and *D. orena* interspecies comparison, data from 300 bp from the 5′ end of the 18S genes and a region around the true promoter are shown (11). The range of variation in satellite DNA is from multiple clones of 500 bp satellite DNA from each of five species in the *D. melanogaster* subgroup (27).

	Drosophila melanogaster intrastrain variation				*D. orena vs D. mel.* interspecies variation		
	NW	UK	OR	Total	18S	promoter	satellite DNA
% polymorphic sites	1.26	0.84	0.84	2.30	0.00	11.0	0.70–3.67
transvers./transit.	0.20	0.33	0.50	0.57	0.00	2.08	1.33–2.05
% add. or delet.	0.00	0.00	0.21	0.21	0.00	3.24	0.15–0.90
sample size	15	15	8	38	2	2	9–15

although both share mutations with the wild type strain. This is indicative of the stochastic nature of homogenization in the separate inbred strains from the mutations in the wild.

The promoter region divergence between *D. melanogaster* and the sibling species *D. orena* reveals that 11.0% of sites have changed (11). This is much higher than the intraspecies level of variation, as would be expected from concerted evolution. Interestingly, the transversion to transition ratio of 0.57 found within *D. melanogaster* is not reflected in this interspecies comparison, where the ratio is 2.08. The transversion to transition ratios found in multiple clones of a 500 bp satellite DNA from several species in the *D. melanogaster* subgroup is also nearer to the expected 2:1 ratio (27). The low ratio within *D. melanogaster* may be an indication of different selective and/or genomic constraints operating within a species than between species.

It is difficult to make anything but very crude comparisons between functionally diverse sequences due to differing selective forces that may be operating on them, as well as differing sample sizes. Satellite DNA variation in species from the *D. melanogaster* species subgroup ranges widely from 0.70 to 3.65% (27). It is difficult to explain this range of intraspecies variation through selective constraints only. Should satellite DNA have any role at all, it is unlikely that it would vary so dramatically between such sibling species. Thus, species-specificity in the rates and degrees of constraints on turnover mechanisms responsible for spreading mutations in the satellite DNA families are more likely to be variable.

No sequence variation was detectable between *D. orena* and *D. melanogaster* in the 18S rRNA gene (11). This high level of similarity is expected for the rRNA genes, although only part of the 300 bp compared is thought to be functionally critical. Nevertheless, it is difficult to draw meaningful conclusions from all the above comparisons, other than that inbred *D. melanogaster* strains have a relatively low level of promoter region variation, comparable to the minimum level of variation found in satellite DNA. Laboratory practices leading to inbreeding would reduce heterozygosity in the population but would not necessarily affect the level of variability amongst repeats of a multigene family. The level of variation in a wild strain, however, is twice that found in the inbred strains, and falls towards the upper level of variation found in satellite DNA. Since we do not know which are the specific nucleotides involved in the promotion of rDNA transcription, we can only conclude that the promoter region displays a significantly higher level of variation than other regions in the rDNA unit, such as in the 18S, and that this suggests either tolerance of functionally neutral mutations or a coevolutionary adjustment in the pol I cofactor genes to mutations that have affected promoter functions.

Homogenization of Polymorphisms

The extent to which multigene families can homogenize a polymorphism within a population is revealed by the presence of the +A:−215 and T→G:−222 mutations in all of the OR clones, but none of the UK or NW clones. It is probable that A56 and Y22 are from the same population as the ORs since they have both of these mutations and another one also found in OR 4 (T→C:−300). The +A:−215 and T→G:−222 mutations probably exist in the rDNA arrays of the X as well as the Y chromosome since these clones are known to come from both chromosomes. These are not found in Dmr238 suggesting that this clone is derived from a different population. However, it should be noted that an increase in the number of chromosomes with rDNA arrays in some other species, such as humans, may affect the homogenization rates between chromosomes and hence may affect the patterns of distribution of variant promoters (28).

The G→A:−17 polymorphism is shared by one NW, two OR and the Dmr238 clones. These mutations, as well as the A→T:−21 in UK9 fall within the −43 to +4 region known to be critical to *in vitro* transcription (16, 17). We are currently using *in vitro* and *in vivo* transcription assays to compare the ability of these variants to support transcription. Evidently, the G→A:−17 is a stable and old polymorphism since it is found in both OR (American) as well as UK (British) strains. It is also found in the Dmr238 clone and therefore, possibly, in a third population. However, unlike the +A:−215 and T→:G−222 mutations it has not spread to all of the units within a population. It is worth noting that the neighbouring G at −16 is conserved in many diverse species, and that point mutation of this G in the mouse promoter results in a 95% reduction in *in vitro* transcription (29). The fact that the G→A:−17 mutation has not spread to all units within a population could reflect functional constraints, in that any further spreading might lead to a reduction in transcription. It is such a polymorphic situation that is thought to play a vital role in the evolution of species-specific promoters, since there now exists the small possibility that a mutation in the cofactor gene(s) will improve promoters carrying the G→A:−17 variant to be more functional, thereby allowing this variant to spread through the family.

The presence of repeats in the *D. melanogaster* rDNA spacer is characteristic of the rDNA spacers of many species. In *Xenopus laevis*, *D. virilis* and *D. melanogaster* the promoter homology

Table 2. The distribution of polymorphisms found in spacers of A56 and Dmr238 from within the promoter region (−19 to +33), the first 240 bp upstream repeat, the middle 240 bp repeats or the 330 bp repeats is tabulated. NA = not applicable in the case where the nucleotide position is not within the repeated area of homology. (−) signifies absence of addition.

	A56 spacer				Dmr 238 spacer			
polymorphism	promoter	first 240	other 240's	330's	promoter	first 240	other 240's	330's
G→A:−17	G	G	G	NA	A	G	G	NA
G→T:−171	NA	G	T or A	T	NA	G	T or A	T
T→G:−215	NA	G	T	T	NA	T	T	T
+A:−222	NA	A	−	−	NA	−	−	−
T→C:−300	NA	C	C	NA	NA	T	T or C	NA

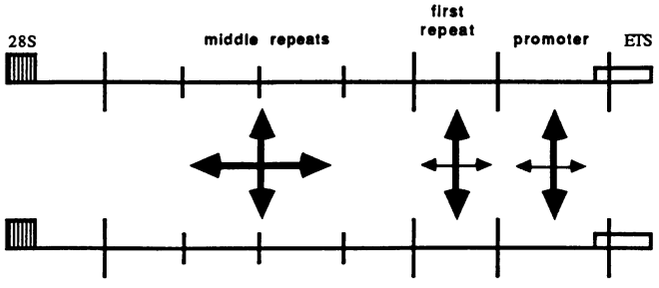

Figure 4. Diagram showing proposed rates of homogenization between or within an rDNA unit. The size of the arrow indicates relative rates. The rDNA spacer is divided into three parts: the promoter region (−140 to transcription start), the first upstream spacer repeat (−380 to −140) and the middle spacer repeats. Vertical arrows indicate rate of homogenization of a variant between rDNA units. Horizontal arrows indicate the rate of homogenization within an rDNA unit's array of repeats.

found in the repeats is known to support transcription *in vivo* (26, 30) and *in vitro* (24, 25). Furthermore, the presence of these upstream repeats enhances transcription at the true promoter and thus are thought to have functional importance (31, 32, 33, 34). In *D. melanogaster* the 3′ end of the first 240 bp upstream spacer repeat is −140 bp relative to transcription start (15,24,35). This is preceded by several more tandem copies which are in turn preceded by a variable number of 330 bp repeats. The sequences of the repeats are highly homologous to each other (>97%). The variable numbers of repeats between rDNA spacers within a population is thought to be the result of the same unequal crossingover events that are responsible for their sequence homogenization (8).

The positions of five of the eleven polymorphisms can be found in the spacer repeats. Since the entire spacer sequences are known for clones A56 and Dmr238 (11,15,35), it is instructive to check their distribution. This is shown in table 2.

It is immediately clear that the presence or absence of a mutation in the promoter region (−140 to +1) is not indicative of it's presence in any of the repeats. For example, the G→A:−17 mutation is found in the true promoter in four different clones, but never in the spacer promoter, suggesting that the primary unit of turnover (unequal crossingover) is the entire rDNA unit rather than the spacer repeats. In other words, a variant whole unit can spread through the family, without the mutation in question spreading to the repeats within the variant rDNA unit itself (see Fig. 4).

Surprisingly, the first upstream 240 bp repeat also seems polymorphically independent of the repeats upstream to it. The G→T:−171 found in the first 240 bp repeat of four NW and one UK clone, is not present in the first repeat of either A56 or Dmr238. However, in the latter two clones this site is found to be hypervariable (T or A) in the upstream 240 bp repeats, but only as a T in the 330 bp repeats. Similarly, the +A:−222 present in the first 240 of all the OR's as well as Y22 and A56, is never found in any of the upsrteam 240's or 330's.

These observations can be rationalized by supposing that the mechanisms that homogenize variants in the first repeat and the true promoter operate predominantly at the level of the entire rDNA unit. This is in contrast to the middle repeats, where the homgenization of variants to neighbouring repeats occurs much more rapidly. This is schematicized in Figure 4.

This lack of homogenization between the first repeat and true promoter with the rest of the repeats could be because they are at the end of the array of repeats. It has been shown within wheat and maize rDNA that the repeat at the ends of arrays is much more sequence divergent and less homogeneous with those in the middle (36). This observation extends to interspecies data as well: in *D. melaongaster* and *D. orena*, the 240 bp repeats display a higher similarity to each other than do the promoters of these species (11).

In summary, mutations within the promoter spread much more rapidly to other rDNA units, rather than to the upstream repeats. Taken with the interspecies data from *D. melaongaster* and *D. orena* (11), this suggests that the promoter is diverging more rapidly and relatively independently from the upstream repeats. This situation underlines the necessity of pol I cofactors to coevolve in order to keep up with these changes.

REFERENCES

1. Long, E.O., and I.B. Dawid, (1980) Repeated Genes in Eukaryotes. Annu. Rev. Biochem. **49**, 727−764.
2. Arnheim, N. (1983) In *Evolution of Genes and Proteins* (Nei, M. & Koehn, R.K., eds), pp.38−61, Sinauer, Sunderland, U.S.A.
3. Dover, G.A. (1982) Nature, **299**, 111−117.
4. Fedoroff, N.V. (1979) Cell, **16**, 697−710.
5. Smith, G.P. (1976) Science **191**, 528−535.
6. Tartof, K, (1974) Cold Spring Harb. Symp. quant. Biol. **38**, 491−500.
7. Arnheim, N., Krystal, M., Schmickel, R., Wilson, G., Ryder, O. and Zimmer, E. (1980) Proc. Nat. Acad. Sci. U.S.A., **77**, 7323−7327.
8. Coen, E.S., Thoday, J. M. and Dover, G.A. (1982) Nature **295**, 564−568.
9. Coen, E.S. and Dover, G.A. (1983) Cell **33**, 849−855.
10. Murtif, V.L. and Rae, P.M.M. (1985) Nucl. Acids Res. **13**, 3221−3239.
11. Tautz, D., Tautz, C., Webb, D.A. and Dover, G.A. (1987) J. Mol. Bio. **195**, 525−542.

12. Kohorn, B.D. and Rae, P.M.M. (1983) Proc Nat. Acad. Sci. U.S.A., **79**, 1501–1505.
13. Dover, G.A. and Flavell, R.B. (1984) Cell, **38**, 623–624.
14. Long, E.O., Rebbert, N.L. and Dawid, I.B. (1981) Proc. Nat. Acad. Sci. U.S.A., **78**, 1513–1517.
15. Simeone, A., La Volpe, A. and Boncinelli, E. (1985) Nuc. Acids. Res. **13**, 1089–1101.
16. Kohorn, B.D. and Rae, P.M.M. (1983) Proc. Nat. Acad. Sci. U.S.A., **80**, 3265–3268.
17. Kohorn, B.D. and Rae, P.M.M. (1983) Nature, **304**, 179–181.
18. Lindsley, D.L. and Grell, E.H. (1968) Genetic Variations of the *Drosophila melanogaster*, Carnegie Institution.
19. Dawid, I.B., Wellauer, P.K. and Long, E.O. (1978) J. Mol. Biol. **126**, 749–768.
20. Tartof, K.D. and Dawid, I.B. (1976) Nature, **263**, 27–30.
21. Chung, C.T. and Miller, R.H. (1988) Nuc. Acids Res. **16**, 3580.
22. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Nat. Acad. Sci. U.S.A., **74**, 5463–5467.
24. Coen, E.S. and Dover, G.A. (1982) Nuc. Acids Res. **10**, 7017–7026.
25. Kohorn, B.D. and Rae, P.M.M. (1982) Nuc. Acids Res. **10**, 6879–6886.
26. Miller, J.R., Hayward, D.C. and Glover, D.M. (1983) Nuc. Acids Res.,**11**, 11–19
27. Strachan, T., Webb, D. and Dover, G.A. (1985) EMBO J. **4**, 1701–1708.
28. Dover, G.A. (1989) Genetics, **122**, 249–252.
29. Skinner,J.A., Ohrlein, A. and Grummt, I. (1984) Proc. Natl. Acad. Sci. USA, **83**, 2137–2141.
30. Murtif, V. L. and Rae, P.M.M. (1985) Nucl. Acids Res. **13**, 3221–3239.
31. Labhart, P. and Reeder, R.H. (1983) Cell, **37**, 285–289.
32. Moss, T. (1983) Nature, **302**, 223–228.
33. Grimaldi, G. and Di Nocera, P.P. (1986) Nuc. Acids Res., **14**, 6417–6432.
34. Hayward, D.C. and Glover, D.M. (1988) Nuc. Acids Res., **16**, 4253–4268.
35. Simeone, A., La Volpe, A. and Boncinelli, E. (1982) Nuc. Acids. Res. **10**, 8263–8272.
36. Barker, R.F., Harberd, N.P., Jarvis, M.G. and Flavell, R.B. (1988) J. Mol. Biol. **201**, 1–17.