Volume 11 Number 3 1983 Nucleic Acids Research

Multiple heterogeneities in the transcribed spacers of ribosomal DNA from *Xenopus laevis*

Monica A.Stewart, Lucinda M.C.Hall and B.Edward H.Maden

Department of Biochemistry, University of Glasgow, Glasgow G12 8QQ, UK

Received 14 December 1982; Accepted 11 January 1983

ABSTRACT

Ribosomal DNA (rDNA) from Xenopus laevis contains several heterogeneities in all three transcribed spacers, as revealed by analysis of cloned and uncloned amplified rDNA from oocytes and cloned chromosomal rDNA from erythrocytes. Heterogeneities include single base changes and length variants of one to several nucleotides. Sites of variation are widely but non-uniformly distributed, some occurring only a short distance outside the boundaries of the rRNA coding regions. No two transcription units that we have yet examined are identical throughout their transcribed spacer regions.

INTRODUCTION

Previous work from this laboratory has established the nucleotide sequence of the 18S-28S intergene region (1), the 18S gene (2) and the external transcribed spacer (3) of a cloned ribosomal transcription unit from Xenopus laevis. A further study has established that the 18S gene sequence in X. laevis is highly homogeneous by the criteria of sequence analysis on additional clones and partial sequence analysis of a large preparation of uncloned rDNA (4). The source of rDNA for these studies was amplified rDNA from X. laevis oocytes (4). In this paper we describe the results of an examination of the transcribed spacers of X. laevis rDNA for heterogeneities. We have examined the same clones and uncloned rDNA as were used in the 18S gene analysis (4), together with further clones of amplified rDNA and newly derived clones of chromosomal rDNA. Our findings reveal considerable microheterogeneity in all three transcribed spacers, in marked contrast to the findings (4) for the 18S gene sequence.

OUTLINE OF EXPERIMENTS

Our search for variants covered part of the external transcribed spacer (ETS) and the whole of both internal transcribed spacers (ITS 1

© IRL Press Limited, Oxford, England.

and ITS 2). The reference sequence for all these regions was the previously determined sequence in the rDNA clone pXlr101 (1, 3). In the present work we first examined the various sources of amplified oocyte rDNA. For each region studied we determined the complete sequence in either one or two other clones, and carried out partial sequence analysis on further clones and on a large preparation of uncloned amplified rDNA. In the partial analyses we took advantage of the fact that the transcribed spacers are extremely rich in G plus C, and screened for sites of variability by carrying out comparative G cleavage or C cleavage experiments. Confirmatory or additional data were then obtained by carrying out appropriate full sequencing runs. Finally, we examined appropriate regions of the chromosomal rDNA clones by one or more of the following procedures: partial G cleavage, partial C cleavage or full sequencing runs.

MATERIALS AND METHODS

Clones of amplified oocyte rDNA

The clones of oocyte rDNA that were used are summarized in figure 1. pXlr101, 102 and 103 contain complete units of rDNA: the ETS and ITS occur in a single transcription unit in each of these clones. pXlr14 and pXl108 and are clones of a segment of the rDNA repeating structure that includes the NTS and ETS, but not the ITS. pXlr11 and pXl212 contain the ITS regions but not the ETS. The derivations of most of the clones have been described (4). pXlr102 and 103 were derived from the same experiment as pXlr101 (4). Parts of several of the clones were subcloned into pBR322 prior to most of the sequence analyses. Selected control experiments showed no variation between the subclones and the parent clones.

Uncloned amplified rDNA

The origin of this preparation has been described (4). In brief, it was isolated from the ovaries of a large number of frogs. Both the uncloned and the cloned amplified rDNA originated from groups of frogs which had been bred at the Carnegie Institution of Washington, Department of Embryology, between 1973-76, after descent by one or a few generations from the wild.

Cloned chromosomal rDNA

A sample of purified chromosomal rDNA was donated by Dr. Adrian Bird. The rDNA was isolated from the erythrocytes of several frogs and

was purified by banding in CsCl. The frogs had been purchased from Xenopus Ltd. and originated from the wild. Restriction of a small aliquot of the rDNA with Eco RI yielded the expected bands (ref. 5). Another small aliquot was shown to be highly resistant to Hpa II. This resistance is characteristic of erythrocyte chromosomal rDNA, in contrast to amplified or cloned rDNA, and is due to methylation of the Hpa II sites (6). About 0.8 μg of the DNA was then digested with Hind III and the products were cloned into pAT153. Five rDNA clones were obtained, each containing a complete rDNA unit bounded by Hind III sites, thus corresponding in configuration to the oocyte rDNA clones pXlr101-103 (figure 1). The chromosomal clones were designated pXlcr1-5, where "Xlcr" stands for "X. laevis chromosomal ribosomal".

DNA sequence analysis

The general procedures were those of Maxam and Gilbert (7), with 5' labelling of restriction fragments followed by secondary restriction to obtain fragments uniquely labelled at one end. Chemical modification reactions were for G, A plus G (using pyridine formate. pH 2), T plus C and C.

Restriction and sequence analysis of uncloned amplified rDNA

On the basis of known restriction and sequence data from clones (8-10, 1) a restriction strategy was devised to make the maximum use of a limited amount of uncloned rDNA (about 20 micrograms). The strategy involved combined digestion with Bam HI and Xba I (figure 2) and yielded material for this work and that described in ref.4. Aliquots of several

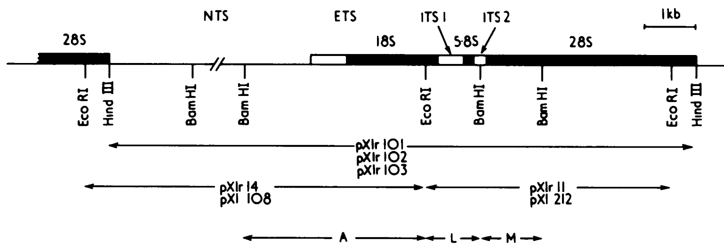
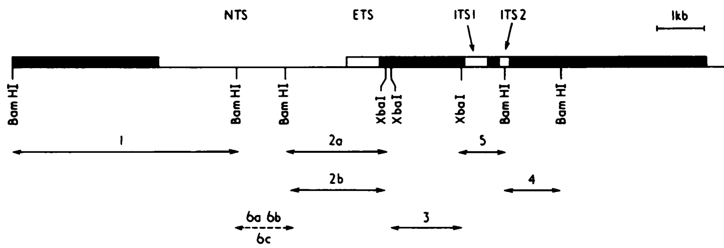


Figure 1. Structure of rDNA from X.laevis, showing slightly more than one unit of the repeating structure, the restriction sites used for cloning and the amplified oocyte rDNA clones used in this work. NTS, ETS, and ITS denote non-transcribed, external and internal transcribed spacers. The numbers and spacing of Bam HI sites in the NTS differ between different clones (ref.8). Subclones containing regions A, L and M from some of the clones were used in this work. The chromosomal rDNA clones, pXlcr1-5, are homologous in configuration to pXlr101-103.



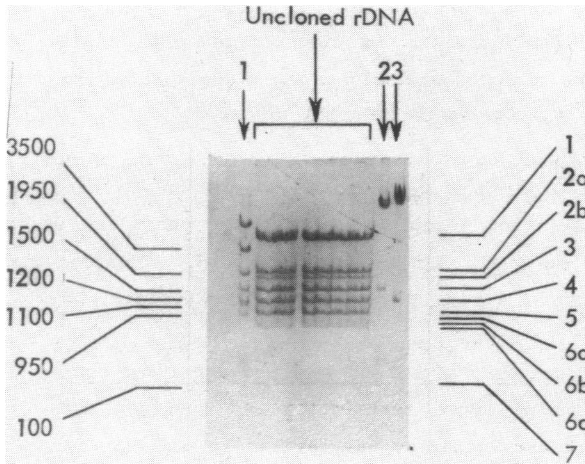


Figure 2. Restriction of uncloned amplified rDNA.
Upper section. Map of Bam HI and Xba I sites in rDNA, established from data from various clones (8-10,1). Lines with arrowheads represent fragments obtained from complete digestion of uncloned rDNA with Bam HI and Xba I, numbered according to size (see lower section). Fragments 2a and 2b are variants due to heterogeneity in the location of Bam HI sites in the NTS. (See the Results, ETS section). Fragments 6a, b and c are also variants due to Bam HI site heterogeneity in the NTS.
Lower section. The wide central section shows the preparative Bam HI plus Xba I digest of uncloned rDNA. The relationship of the fragments to the rDNA repeating structure was confirmed by markers in lanes 1-3. Lane 1, complete digestion of pXlr101 with Bam HI and Xba I. Lane 2, digestion of pXlr101 with Xba I. Lane 3, digestion of pXl212 with Bam HI. The top two fragments in lane 1, and the top fragment in each of lanes 2 and 3, contain some material from the respective vector linked to rDNA. The more rapidly migrating fragments contain only rDNA. The sizes of the rDNA fragments are indicated on the left (including a small Xba I fragment 7 from the left hand region of the 18S gene, not numbered in the upper section). The photograph is a negative print of an ethidium bromide-stained 1% agarose gel.

of the fragments were digested with further enzymes to give access to additional sites for sequencing (see figures 3 and 6 in the Results Section). The various fragments were 5' end-labelled and subjected to final, secondary restriction on the basis of sites which had been mapped in clones, yielding the expected fragments usually in just sufficient quantity for visualization in polyacrylamide gels after staining with ethidium bromide. As expected from previous work (6, 11, 4), no methylated cytosines were encountered in this uncloned amplified rDNA, and this facilitated the restriction and sequence analysis.

Comparative partial chemical cleavage

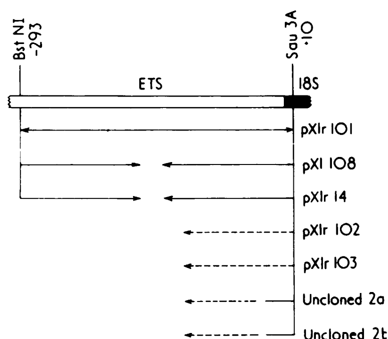
Corresponding restriction fragments from the various clones and from uncloned rDNA, after 5' terminal labelling and secondary restriction, were subjected in parallel to partial G modification or partial C modification, followed by chain cleavage with piperidine, under the same conditions as used for full sequencing runs (7). The choice between the G or C reactions was made so as to maximise the amount of useful information that could be obtained, based on data from the reference clone and on any preliminary data from other clones. Where suitable, G cleavage was preferred because hydrazine (C specific) can cross react weakly with G giving weak extra bands in the cleavage pattern. The products were analyzed in parallel on an 8% sequencing gel. This procedure was applied in separate experiments to material derived from amplified and from chromosomal rDNA. An example is shown in figure 4, Results Section.

RESULTS

Amplified rDNA

External transcribed spacer (ETS)

This is the longest of the three transcribed spacers in X. laevis rDNA (approximately 712 nucleotides in pXlr101). Parts of the reference sequence were technically difficult to establish, especially in the central region (3). In preliminary comparative studies the region near the 18S gene revealed some sites of variation between clones. This region was therefore chosen for detailed study. Figures 3-5 show, respectively, the analyses that were carried out, an illustrative experiment, and the results obtained. The illustrative experiment (figure 4) shows the G cleavage patterns of several rDNA clones and of uncloned amplified rDNA, extending leftwards from a Sau 3A site in the



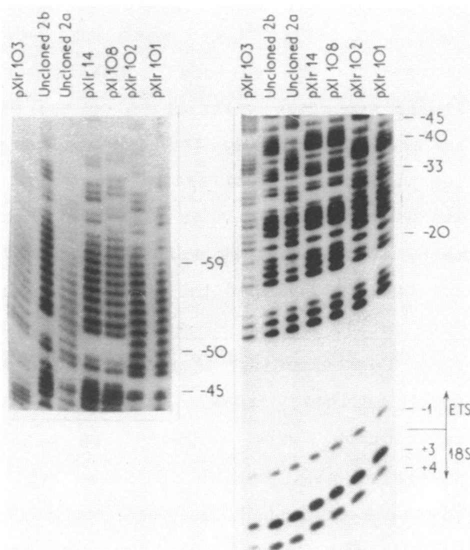


Figure 3 (above left). Diagram of sequencing runs on the ETS (oocyte rDNA). pXlr101 was subjected to complete sequence analysis throughout the ETS as detailed in ref.3. pXl108 and pXlr14 were subjected to full sequencing runs (continuous arrows) from the restriction sites shown. (Nucleotide numbers are from the complete pXlr101 sequence). pXlr102 and pXlr103 were subjected to G cleavage runs (interrupted arrows) from the indicated Sau 3A site. Uncloned rDNA fragments 2a and 2b were subjected to complete sequencing runs from the Xba I site at nucleotide +158 in the 18S gene, permitting reading into the first 25 nucleotides of the ETS, and to G cleavage runs (interrupted arrows) from the indicated Sau 3A site.

Figure 4 (above right). G cleavage patterns of fragments from rDNA clones and uncloned rDNA, extending leftwards from the Sau 3A site at nucleotide +10 in the 18S gene. The figure shows the relevant areas from two overlapping loadings from the same 40 cm sequencing gel. Guanine bands in this figure correspond to cytosine on the rightwards strand in figure 5. Nucleotides are numbered for pXlr101 according to ref.3. (The pXlr101-type sequence gave some spacing anomalies between nucleotides -20 and -33 on the leftwards strand. The full sequence was established from both strands in pXlr101 itself (ref.3). It was thereby possible to identify the 101-type sequence unambiguously in other sources from data on the leftwards strand).

18S coding region into the ETS. Bands signifying G in this strand correspond to C in the rightwards, RNA-like strand in figure 5. Several aspects of transcribed spacer variability are exemplified by this experiment.

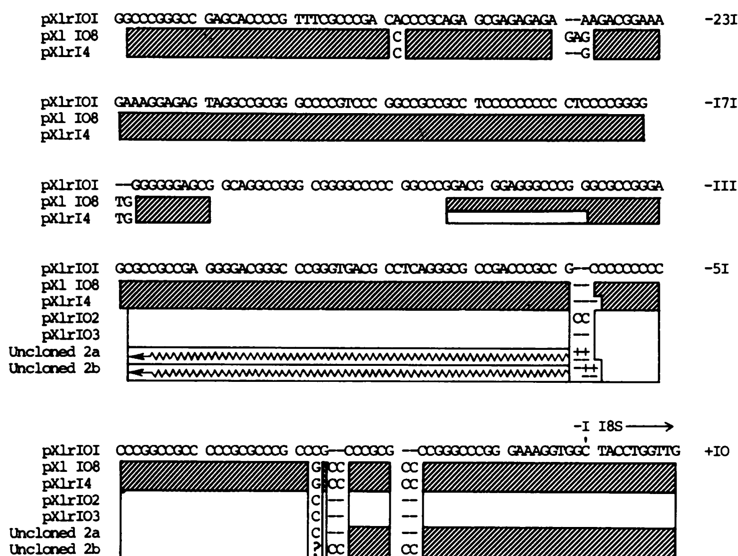


Figure 5. Summary of comparative sequence data in the ETS (rightwards, RNA-like strand). The figure shows the sequence in pXlrl01 and the variations between amplified oocyte rDNA sources. Dashes indicate deletions with respect to the longest sequence examined. Plus signs for uncloned rDNA mean that part of the population contains extra bases at this point. A shaded box signifies that the indicated rDNA sources were subjected to complete sequencing runs, with identical findings to pXlrl01 in the boxed region. An open box signifies that G cleavage runs were carried out, with no differences evident from pXlrl01 in the boxed region. The zig-zag lines for uncloned rDNA indicate partly unresolved cleavage patterns due to length heterogeneity at the origins of the lines. The question mark at position -28 in uncloned fragment 2b means that there was no G in the leftwards strand at this point (figure 4). Thus the actual nucleotide cannot be specified with certainty. However, by analogy with sequence data from pXl108 and pXlrl4 a cytosine is almost certainly present, and hence guanine in the rightwards strand. The region of pXl108 that was previously sequenced (ref.12) terminates at nucleotide -168 according to the pXlrl01 numbering system shown. To the left of this point, our sequence for pXl108 agrees with that of ref.12 except at two points. At nucleotides -209 to -206 we find four cytosines whereas Moss et al reported three. At nucleotides -200 and -199 we find two guanines whereas Moss et al reported one. At both of these points there was some degree of compression of nucleotides in gels of the rightwards strand in all three clones examined. However resolution was sufficient to indicate the sequence shown here.

First, in the region between nucleotides -20 and -33 there is a small cluster of sites of variation between clones: compare pXlrl01, 102 and 103 with pXlrl4 and pXl108 (figure 4). Full sequencing runs on three of the clones showed that pXlrl4 and pXl108 differ from pXlrl01 by

the presence of two mini-insertions and a substitution (figure 5). After nucleotide -33 the two types of pattern are similar but out of register by four nucleotides due to the mini-insertions in the pXlr14-type sequence.

Second, in the left hand section of figure 4 all the clones show a long run of guanines corresponding to the C tract at nucleotide -50 in the rightwards strand (figure 5). The length of this homopolymeric tract varies from 11 to 14 in different clones.

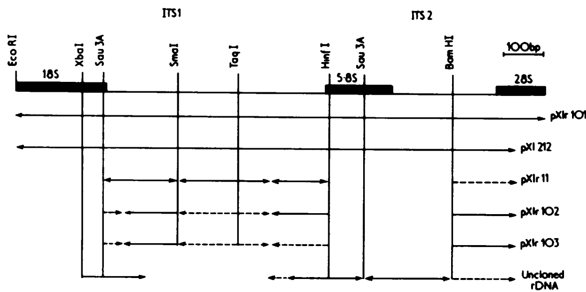
Third, the findings for uncloned, amplified rDNA correspond to those for the clones. Fragments 2a and 2b are alternative fragments containing part of the NTS, the entire ETS and part of the 18S gene (figure 2). Between nucleotides -20 and -33 the longer fragment 2a contains predominantly the pXlr101-type sequence as revealed by G cleavage (figure 4) whereas the shorter fragment 2b contains predominantly the pXlr14-type sequence. Continuing upstream the cleavage patterns remain distinct until the long run of guanines corresponding to the oligo C tract in the rightwards strand. Beyond this the patterns show loss of internal register: there are contributions from "out of phase" patterns, indicating that length heterogeneity occurs at this point in both fragments, in correspondence with the length variants at this site in the clones.

(Fragments 2a and 2b differ in length from each other by some 150 nucleotides. Known features of NTS sequence organization (8, 9) implied that the main source of this length difference would be at the NTS end of the respective fragments. This was confirmed by a restriction test which showed that the distance from the last Alu I site in the NTS to the first Xba I site in the 18S gene was similar to within a few nucleotides in the two fragments: unpublished data obtained during work for ref. 4).

Finally in the ETS, analysis of three clones in the region extending rightwards from nucleotide -293 revealed four further sites of variation: two with mini-insertions and two with base substitutions (figure 5). pXl108 had been sequenced previously in this region (12). The sequence shown here differs at two points from that which was previously reported, as noted in the legend to figure 5.

<u>First internal transcribed spacer (ITS 1)</u>

Figure 6 depicts analyses that were carried out on the 18S-28S intergene region; figure 7 summarises the findings for ITS 1. Note that we have replaced the preliminary numbering system of ref. 1 for the



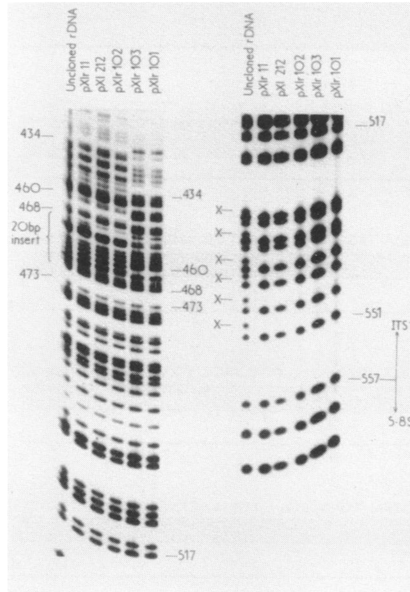
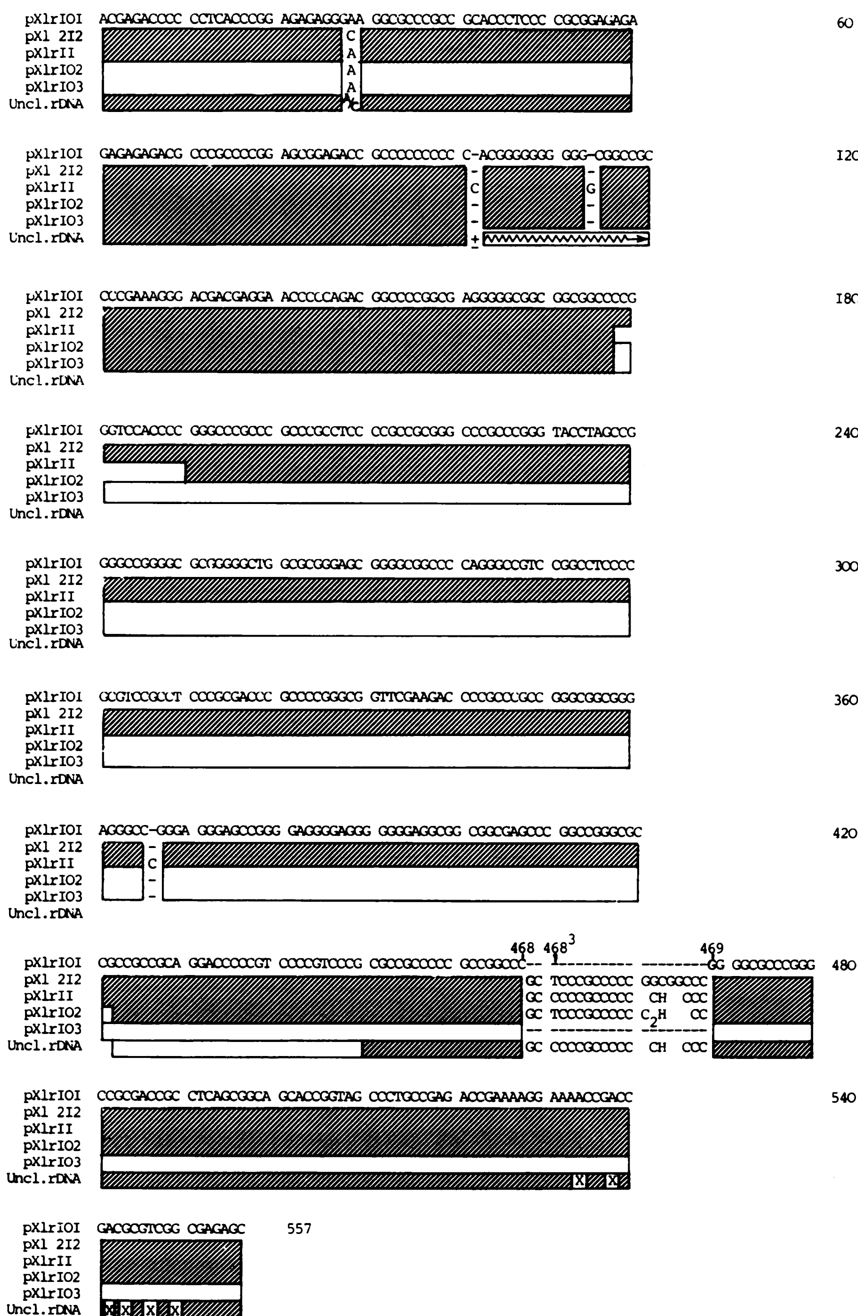


Figure 6. Upper section: sequencing runs on the 18S-28S intergene region. pXlr101 was fully sequenced in this region as detailed in ref.1. pXl212 was also fully sequenced in this region: see the legend to figure 7. Other sources of oocyte rDNA were subjected either to sequencing runs (continuous arrows) or G or C cleavage (interrupted arrows) where shown.

Lower section G cleavage patterns of fragments from rDNA clones and uncloned rDNA extending leftwards from the Hinf I site just after the start of the 5.8S gene. The figure shows 2 overlapping loadings from the same 40 cm sequencing gel. Guanine bands correspond to cytosine on the rightwards strand in figure 7. Nucleotides are numbered from pXlr101 according to the system in figure 7. A region with extra G bands in uncloned rDNA (marked "X") is almost certainly due to a mixture of the "standard" sequence with a "phase shifted" variant: see under "chromosomal rDNA clones". The region with the 20 bp insert is also seen. Uncloned rDNA was subjected to a complete sequencing run through the region shown in this figure, and the main findings were confirmed and extended (see figure 7).



pXlrIOI ACGAGACCCC CCTCACCCGG AGAGAGGGAA GGCGCCCGCC GCACCCTCCC CGCGGAGAGA 60
pXl 2I2
pXlrII
pXlrIO2
pXlrIO3
Uncl.rDNA
C A A A
pXlrIOI GAGAGAGACG CCCGCCCCGG AGCGGAGACC GCCCCCCCCC C-ACGGGGGG GGG-CGCCCGC I20
C G
pXlrIOI CCCGAAAGGG ACGACGAGGA ACCCCCAGAC GGCCCCGGCG AGGGGGCGGC GGCGGCCCCG I80
pXlrIOI GGTCCACCCC GGGCCCGCCC GCCCGCCTCC CCGCCGCGGG CCCGCCCGGG TACCTAGCCG 240
pXlrIOI GGGCCGGGGC GCGGGGGCTG GCGCGGGAGC GGGGCGGCCC CAGGGCCGTC CGGCCTCCCC 300
pXlrIOI GCGTCCGCCT CCCGCGACCC GCCCCGGGCG GTTCGAAGAC CCCGCCCGCC GGGCGGCGGG 360
pXlrIOI AGGGCC-GGGA GGGAGCCGGG GAGGGGAGGG GGGGAGGCGG CGGCGAGCCC GGCCGGGCGC 420
C
pXlrIOI CGCCGCCGCA GGACCCCCGT CCCCGTCCCG CGCCGCCCCC GCCGGCCC-- ---------- --------GG GGCGCCCGGG 480
468 468³ 469
GC TCCCGCCCCC GGCGGCCC
GC CCCCGCCCCC CH CCC
GC TCCCGCCCCC C₂H CC
GC CCCCGCCCCC CH CCC
pXlrIOI CCGCGACCGC CTCAGCGGCA GCACCGGTAG CCCTGCCGAG ACCGAAAAGG AAAACCGACC 540
X X
pXlrIOI GACGCGTCGG CGAGAGC 557
X X X X

18S-28S intergene region by separate systems for ITS 1 and ITS 2. There are several sites of variability in ITS 1.

At nucleotide 29 pX1212 contains C whereas other clones contain A. Next, in the vicinity of nucleotide 100, most clones contain a tract of 10 cytosines followed after two nucleotides by 10 guanines, whereas pXlr11 contains 11 of each of these nucleotides. Then, pXlr11 contains an extra cytosine following C366 of the reference sequence. (This gives rise to an extra Sma I site which was experimentally demonstrated during this work).

The most striking site of variation in ITS 1 occurs immediately following nucleotide 468 in the reference sequence (figure 7). At this point three out of the five clones possess an extra block of 20 nucleotides. Complete sequencing runs showed variation within variation: at position 468.3 two clones possess T and a third possesses C (figure 7).

Uncloned amplified rDNA was analysed inwards from both ends of ITS 1 by full sequencing runs. At nucleotide 29 there was a single, clearly defined A/C heterogeneity corresponding to the site of A/C variation between clones. Otherwise the first part of ITS 1 appeared homogeneous until the vicinity of nucleotide 100, where there was evidence of length heterogeneity corresponding to the site of length variation between clones in this region. The sequencing run from the right hand end of ITS 1 in uncloned rDNA showed that the predominant sequence contained the extra 20 bp segment next to nucleotide 468. A sizing experiment with an

Figure 7. Summary of comparative data from amplified oocyte rDNA for ITS 1, shown for the rightwards, RNA-like strand. The figure shows the complete sequence in pXlr101 and the variations between rDNA sources. For explanation of "boxing" see figure 5. Unambiguous sequence determination of the 20 base pair insert was achieved only in pX1212. 'H' in the remaining clones and uncloned rDNA signifies two or more guanine residues, and the order of nucleotides between positions 468^{13} and 468^{17} was not resolved due to compression of nucleotides in the gels obtained. 'X' in uncloned rDNA indicates heterogeneities: see under "chromosomal rDNA clones". Boseley et al (10) previously reported the sequence of 5.8S rDNA and the flanking regions in ITS 1 and ITS 2 in the clone pX1212. In the sequenced regions of the transcribed spacers there appeared to be several sites of variation from pXlr101 (summarised for ITS 1 in ref.1). However, repetition of the pX1212 sequence determination in this work, with sequencing of both strands and thin gels for electrophoresis, revealed an identical sequence to pXlr101 in the respective regions (the last 80 nucleotides of ITS 1 and the first part of ITS 2). Since there are no apparent ambiguities in our data we believe that the present version of the sequence is correct.

aliquot of a labelled restriction fragment encompassing this region showed that only about 2-5% of the material lacked this 20 bp segment. Finally uncloned rDNA revealed a block of ambiguous nucleotides near the start of the 5.8S gene (figure 6, lower section). The explanation for these ambiguities was revealed when further clones were analyzed: see under "chromosomal rDNA clones", below.

Second internal transcribed spacer (ITS 2)

This, the region between the 5.8S and 28S genes, is the shortest transcribed spacer in X. laevis rDNA. Analysis was facilitated by a Bam HI site near the middle of the region (figure 6). To the left of this site the two clones that were examined were identical and the uncloned preparation showed no detectable heterogeneities. The region to the right of the Bam HI site in the clones showed two sites of base substitution followed by two sites of length variation in tracts of cytosines (figure 8). Uncloned rDNA was analyzed on the rightwards

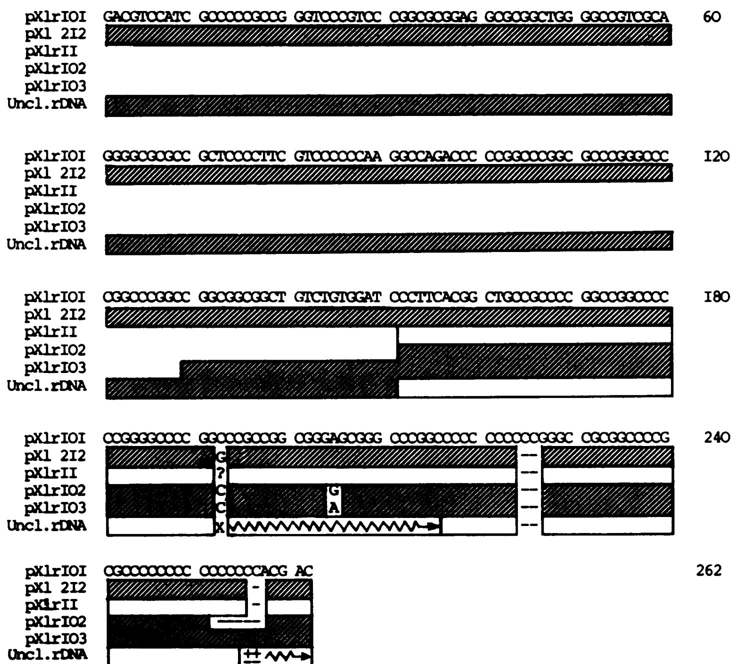


Figure 8. Summary of comparative data from amplified oocyte rDNA for ITS 2 shown for the rightwards, RNA-like strand. The complete sequence in pXlr101 and the variation between rDNA sources are shown. "Boxing" is as explained in figure 5.

strand and showed heterogeneity in the second oligo C tract. The first tract appeared to be largely homogeneous. There was also a complicated pattern for about 20 nucleotides following the first site of variation between clones (at nucleotide 193). This was probably a result of one or more point heterogeneities in uncloned rDNA with consequent effects on spacing downstream.

Chromosomal rDNA clones

The chromosomal rDNA clones were subjected to various complete sequencing runs or to G or C cleavage runs in selected regions of the transcribed spacers. The findings corresponded well to those for amplified rDNA, with minor variations (figure 9). One variant in amplified rDNA for which no corresponding variant has yet been found in the sequenced parts of the chromosomal clones is in ITS 1 following nucleotide 468. Here none of the chromosomal clones resembles pXlr101: all of them contain the extra 20 bp segment, three with T and two with C at position 468.3.

It was mentioned above that uncloned amplified rDNA revealed a block of ambiguous nucleotides near the start of the 5.8S gene. Two of the chromosomal clones possess variant sequences in this region, with a single base deletion at position 525 in ITS 1 followed by a single base insertion following position 551. (There is also some variation within this region: figure 9, inset). Since the region between nucleotides 525 and 551 is "phase shifted" by one base in the variant sequence with respect to the reference sequence, we interpret the ambiguities in uncloned amplified rDNA as being due to superposition of the reference and phase shifted sequences. Thus the variant occurs in amplified rDNA, although not in any clones of amplified rDNA that we have so far studied.

Chromosomal rDNA also showed a new variant in ITS 2, a G$\rightarrow$A change in pXlcr4 at position 192 (figure 9), next to the site of C$\rightarrow$G variation at 193 in pX1212.

rRNA coding regions

As already mentioned, no sites of sequence variation were found in an extensive survey of the 18S coding region, carried out in parallel with this work (4). No variants have been found in the 5.8S coding region in the following sources: pXlr101 (ref.1), pX1212 (ref.10 and this work), pXlcr2 and a partial analysis on uncloned amplified rDNA (this work).

In the short region of 28S rDNA so far examined one variant site

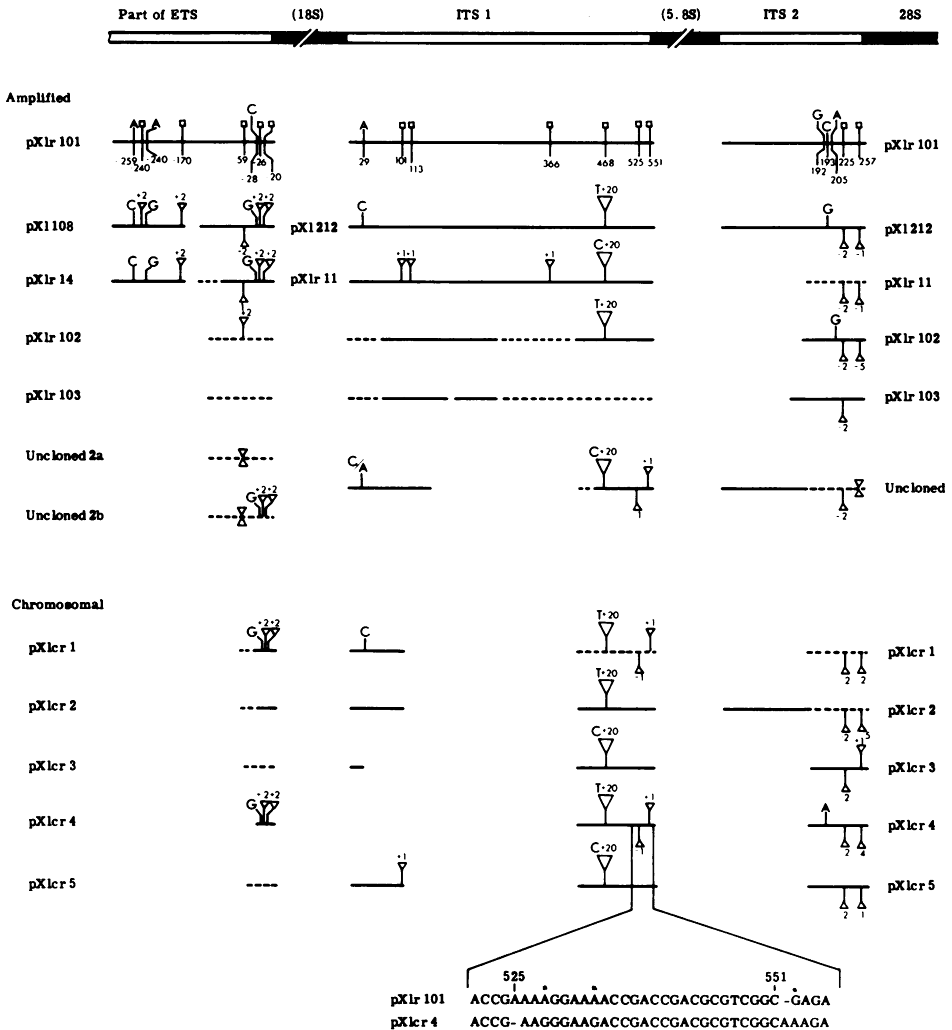


Figure 9. Summary of variants identified in the transcribed spacers of X.laevis rDNA. pXlr101 is the reference sequence: on this line is indicated every site where a transcribed spacer variant has been found in another rDNA source. Where substitutions were found in other sources, the base that occurs in pXlr101 is shown above the line, together with the nucleotide number below the line. For insertions or deletions in other sources a box is shown above the pXlr101 line, and the number of the adjacent nucleotide in pXlr101 below the line. For other rDNA sources a solid line denotes coverage by complete sequencing in the region indicated; an interrupted line indicates G cleavage or C cleavage only. A gap indicates that no data are yet available. Substitutions are denoted by showing the base that differs from that in pXlr101. Insertions with respect to pXlr101 are shown thus Y, with the number of

extra nucleotides, +1 etc. Deletions with respect to pXlr101 are shown thus ⅄ , with the number of deficient nucleotides, -1 etc. For amplified rDNA sources the nature of the additional or missing nucleotides is shown in figure 5, 7 or 8. For chromosomal rDNA the additional or deleted nucleotides were the same as in cloned, amplified rDNA except as follows:- For the region of ITS 1 between nucleotides 525 and 551, the inset shows the variant, phase shifted sequence of pXlcr4 (written for the rightwards strand). pXlcr1 showed a similar G cleavage pattern to pXlcr4 but was not fully sequenced. In ITS 2 of pXlcr4, the A residue indicated is at nucleotide 192.

has been found:- The amplified rDNA clone pXlr102 and the chromosomal clone pXlcr4 both contain an extra C residue adjacent to G40 in the sequence. (Because this variant induces a secondary structure effect it is at present uncertain whether the extra C precedes or follows G40).

DISCUSSION

The above experiments reveal considerable microheterogeneity in the transcribed spacers of X.laevis rDNA. The regions examined in DNA from various sources encompass 1,100 nucleotides of the rDNA unit structure. Some 20 sites at which variants occur have been identified. There is very good general correspondence between the findings for amplified oocyte rDNA and chromosomal rDNA. This indicates that the primary reservoir of variant sequences is genomic rDNA. The recurrence of many of the variants in rDNA from different groups of frogs with largely uncharacterized pedigrees (see Materials and Methods) suggests that these variants are fairly widely distributed in natural populations.

The findings contrast strikingly with those for the rRNA coding regions (ref.4 and this work). In the latter we have found only one site of variation in 2,100 nucleotides that we have so far searched by comparison of two or more rDNA sources. This occurs near the 5' end of the 28S coding region (see under rRNA coding regions, above).

The transcribed spacer variants comprise base substitutions, insertions and deletions of one or a few nucleotides, and a site in ITS 1 which varies by the presence or absence of a 20 bp insert (figure 7). The reference sequence pXlr101 lacks this insert, as also does pXlr103, but the form with the insert now appears to be the common one (figure 9).

Several of the variable sites occur quite close to the rRNA coding regions, implying that there is no unique relationship between sequence and function in these rRNA-flanking regions. There are variant sites within 30 nucleotides of the 5' and 3' ends of the 18S sequence and the

5' end of the 5.8S sequence (figures 5 and 7). A variable length C tract in ITS 2 terminates only a few nucleotides from the 28S boundary (figure 8).

This oligo C tract is representative of a number of "simple sequence" tracts, which include other oligo C tracts and oligopurine tracts (figures 5, 7, 8). The longer oligo C tracts and the long polypurine tract between nucleotides -248 and -221 in the ETS appear to constitute foci for length microheterogeneities.

No two transcription units that we have yet examined are identical throughout their transcribed spacer regions (figure 9). Diversity arises at two levels: different combinations of variants within a single transcribed spacer region (ETS, ITS 1 or ITS 2), and combinations of variants between the three transcribed spacers in a single transcription unit. (The two least different transcription units that we have yet found are pXlr101 and 103, with only a single difference so far identified, in ITS 2).

Notwithstanding this diversity, some linked groups of variants can be recognized. First, in the ETS between nucleotides -20 and -28 all clones so far analyzed conform to one or other of two alternative patterns, resembling pXlr101 or pXlr14 respectively (figures 5 and 9). The latter type possesses two mini-insertions with respect to the former (figures 4 and 5): no intermediate form with only one of the two insertions has yet been found. Secondly, in uncloned rDNA the pXlr101-type sequence in this region predominates in the longer fragment 2a whereas the pXlr14 type sequence predominates in the shorter fragment 2b (figure 4). This implies that in this rDNA preparation the pXlr14 type ETS variant is linked mainly to the shorter of two alternative variants at the NTS end of the fragment (8,9). However, not all of the clones were found to conform to this second linkage pattern (unpublished observations). Moreover, some other sites in the ETS vary independently of the sequence at nucleotides -20 to -28: in particular the length of the oligo C tract at -50, and also the oligopurine tract between -248 and -221 in the two pXlr14 type variants that were sequenced. This implies that change has occurred at these sites after the events that differentiated the pXlr14 type from the pXlr101 type patterns near the 18S end of the ETS. This is consistent with the interpretation that these simple sequence tracts are hot spots for variation. One further example of linkage between variants occurs in ITS 1: the deletion at

position 525 is evidently coupled to the insertion following 551.

The patterns of linked and independent variation must reflect the history of the appearance of mutants and their propagation through the rDNA multigene family. As yet we have no data on the diversity of transcribed spacers within single nucleolar organizers, but we anticipate that some degree of intranucleolar variation will be found, as for NTS length variation (5).

The transcribed spacer sequences in eukaryotic rDNA are phylogenetically highly variable (13-18, 1,3 and Furlong and Maden, in preparation), in contrast to the rRNA coding sequences (2,19). The degree of microgeterogeneity that we find in the transcribed spacers of X.laevis rDNA suggests a state of appreciable sequence flux which may underly this phylogenetic variability. Data on sequence variability at both intraspecies and interspecies levels should be taken into account in attempts to identify functions, if any, of the transcribed spacers during ribosome maturation.

ACKNOWLEDGEMENTS AND NOTES

We thank D.D.Brown and A.P.Bird for gifts of uncloned amplified and chromosomal rDNA respectively and R.Reeder and T.Moss for gifts of some of the clones used in this work. The work was supported by the Medical Research Council. Lucinda Hall's present address is:- Department de Biologie Moleculaire Sciences II, Quai Ernest Ansermet, CH 1211 Geneve 4, Switzerland.

REFERENCES

1. Hall, L.M.C. and Maden, B.E.H. (1980) Nucleic Acids Research 8, 5993-6005.
2. Salim, M. and Maden, B.E.H. (1981) Nature 291, 205-208.
3. Maden, B.E.H., Moss, M. and Salim, M. (1982) Nucleic Acids Research 10, 2387-2398.
4. Maden, B.E.H., Forbes, J.M., Stewart, M.A. and Eason, R. (1982) The EMBO Journal 1, 597-601.
5. Wellauer, P.K., Reeder, R.H., Dawid, I.B. and Brown, D.D. (1976) J.Mol.Biol. 105, 487-505.
6. Bird, A.P. and Southern, E.M. (1978) J.Mol.Biol. 118, 27-47.
7. Maxam, A.L. and Gilbert, W. (1980) in Grossman, L. and Moldave, K. (eds) Nucleic Acids, Part I, Methods in Enzymology, 65, Academic Press, N.Y. pp 499-560.
8. Botchan, P., Reeder, R.H. and Dawid, I.B. (1977) Cell 11, 599-607.
9. Boseley, P., Moss, T., Machler, M., Portmann, R. and Birnstiel, M. (1979) Cell 17, 19-31.
10. Boseley, P.G., Tuyns, A. and Birnstiel, M.L. (1978) Nucleic Acids Research 5, 1121-1137.
11. Dawid, I.B., Brown, D.D. and Reeder, R.H. (1970) J.Mol.Biol. 51, 341-360.

12. Moss, T., Boseley, P.G. and Birnstiel, M.L. (1980) Nucleic Acids Research 8, 467-485.
13. Forsheit, A.B., Davidson, N. and Brown, D.D. (1974) J.Mol.Biol. 90, 301-314.
14. Gourse, R.L. and Gerbi, S.A. (1980) J.Mol.Biol. 140, 321-339.
15. Skryabin, K.G., Krayev, A.S., Rubstov, P.M. and Bayev, A.A. (1979a) Dokl.Akad.Nauk. USSR 247, 1275-1277.
16. Veldman, G.M., Brand, R.C., Klootwijk, J. and Planta, R.J. (1980) Nucleic Acids Research 8, 2907-2920.
17. Veldman, G.M., Klootwijk, J., van Heerikhuizen, H. and Planta, R.J. (1981) Nucleic Acids Research 9, 4847-4862.
18. Subrahmanyam, C.S., Cassidy, B., Busch, H. and Rothblum, L.I. (1982) Nucleic Acids Research 10, 3667-3680.
19. Zwieb, C., Glotz, C. and Brimacombe, R. (1981) Nucleic Acids Research 9, 3621-3640.