Volume 12 Number 22 1984 Nucleic Acids Research

Evolution of vitellogenin genes: comparative analysis of the nucleotide sequences downstream of the transcription initiation site of four *Xenopus laevis* and one chicken gene

Jacques-Edouard Germond, Philippe Walker, Béatrice ten Heggeler, Marianne Brown-Luedi, Eric de Bony and Walter Wahli

Institut de Biologie animale, Université de Lausanne, Bâtiment de Biologie, CH-1015 Lausanne, Switzerland

Received 6 September 1984; Revised and Accepted 24 October 1984

ABSTRACT

Electron microscopic analysis of heteroduplexes between the most distantly related Xenopus vitellogenin genes (A genes x B genes) has revealed the distribution of homologous regions that have been preferentially conserved after the duplication events that gave rise to the multigene family in Xenopus laevis. DNA sequence analysis was limited to the region downstream of the transcription initiation site of the Xenopus genes A1, B1 and B2 and a comparison with the Xenopus A2 and the major chicken vitellogenin gene is presented. Within the coding regions of the first three exons, nucleotide substitutions resulting in amino acid changes accumulate at a rate similar to that observed in globin genes. This suggests that the duplication event which led to the formation of the A and B ancestral genes in Xenopus laevis occurred about 150 million years ago. Homologous exons of the A1-A2 and B1-B2 gene pairs, which formed about 30 million years ago, show a quite similar sequence divergence. In contrast, A1-A2 homologous introns seem to have evolved much faster than their B1-B2 counterparts.

INTRODUCTION

Vitellogenin genes provide a good example of developmentally and hormonally controlled genes. The gene product whose synthesis is controlled by estrogen in the liver of mature females is the precursor of the major constituents of the egg-yolk in oviparous vertebrates (1). Among these vertebrates vitellogenesis is best characterized in amphibians and birds (2, 3, 4). In Xenopus laevis, vitellogenin is encoded by at least four related genes (5, 6, 7, 8). There is also evidence from protein analysis for multiple vitellogenin genes in chicken (9, 10) but so far only one gene has been isolated (11, 12, 13). Apart from its relevance to the molecular basis of hormone controlled gene expression, the vitellogenin multigene family is potentially useful for studies on gene evolution. In a few cases a direct comparison of the arrangement and structure of diverging related genes

within a species or between different species has offered some insight into how gene families were formed and how selective pressure controls their evolution (14, 15). Based on general analyses of the vitellogenin genes in Xenopus laevis (5, 8, 7, 16,) we proposed that the gene family arose by two successive duplication events probably well separated in time (7). Here we extend the analysis of the evolution of these genes in two ways. We define the distribution of conserved sequences between the most distant Xenopus laevis vitellogenin gene relatives and we compare the nucleotide sequences of the region spanning the first three exons of four Xenopus genes and one chicken gene.

MATERIALS AND METHODS

Heteroduplex analysis in the electron microscope

Heteroduplexes between A and B cloned genomic DNAs were formed as described (7).

DNA sequencing

The 5' end region of the genes B1 and B2 was sequenced by the dideoxy-M13 chain termination method and associated techniques (17, 18, 19, 20). With a few exceptions both strands of all the DNA fragments analyzed in this paper were sequenced. The sequences were handled with different analysis programs prepared and kindly provided by Carolyn Tolstoshev (Laboratoire de Génétique Moléculaire des Eucaryotes, Strasbourg).

RESULTS

Electron microscopic analysis of the relatedness between A and B vitellogenin genes

In Xenopus laevis, vitellogenin is encoded by four estrogen-controlled genes, named A1, A2, B1 and B2. Analysis of the thermal stability of heteroduplexes between cDNA clones representing the 3' half of the four mRNAs revealed that A and B coding sequences differ in approximately 20 percent of their nucleotides. The divergence between the A1 and A2 mRNAs and between the B1 and B2 mRNAs is similar and represents approximately 5 percent (5).

The four genes have recently been isolated from DNA libraries and the relatedness of the genomic sequences within the A group and the B group has been studied by analysis of heteroduplexes in the electron microscope (8,

16). General information on the distribution of the most conserved regions between very long stretches of DNA can be obtained quickly using this technique. This is especially important for the vitellogenin gene loci since the isolated genomic clones containing the four genes and their flanking regions encompass about 150 kb of DNA. As stated above, the divergence between A and B sequences is approximately 20 percent in the coding sequences and we expect it to be much higher in introns (8, 16). To analyze the distribution of the most conserved regions between A and B genes we used different genomic clones covering the four vitellogenin genes (see Fig. 1A and ref. 6, 7, 8) to form A1 x B1 (ref. 7), A1 x B2, A2 x B1, and A2 x B2 heteroduplexes under low stringency conditions followed by electron microscopic analysis. Heteroduplex molecules from the different gene pairs were measured and the regions of homology mapped. The results are shown schematically in Figure 1B. Homologies are localized almost exclusively within the genes. A comparative analysis between the different gene pairs shows that particular regions (A to G) are preferentially conserved and probably represent exons coding for the best conserved protein domains which appear to be scattered throughout the whole transcribed regions.

To analyze the relatedness between the isolated vitellogenin genes in more detail we first concentrated on the first three exons and their corresponding introns.

Structural organization of the 5' end region of five vitellogenin genes

The nucleotide sequence of the 5' end region of the Xenopus genes A1, B1 and B2 was determined. Sequence data from the corresponding regions of the Xenopus A2 and of the cloned vitellogenin II chicken gene have already been published (21, 22, 23). Figure 2 indicates the sequenced gene segments, their structural organization (exon-intron map) and the region compared in this paper. For the A1, B1 and B2 gene segments, the schemes are based on the data presented below and in the accompanying paper.

The compared region starts at the 5' end of the vitellogenin mRNAs defined as position +1 (see accompanying paper) and stops within Intron 3. The nucleotide sequence of this region of the five genes is shown in Figure 3. The exon-intron boundaries of the A1, B1 and B2 gene segments were deduced from sequence and structural comparisons with the well-analyzed

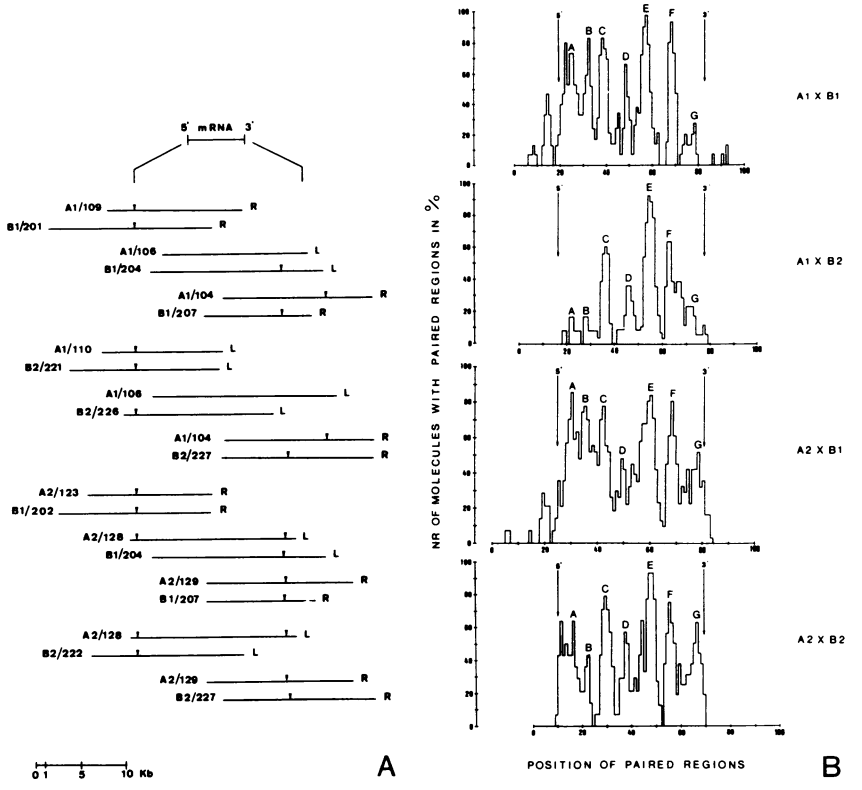


Figure 1. A) Overlapping pattern of the cloned gene fragments used to form heteroduplexes. Only the A1, A2, B1 and B2 cloned genomic DNAs used in this study are shown. Together the genomic clones cover the entire length of each of the four Xenopus genes and some flanking sequences. Heteroduplexes were formed between the DNAs as arranged in the Figure : A1/109 x B1/201; A1/106 x B1/204; A1/104 x B1/207; ...; ... Arrow heads indicate the 5' end and tentative 3' end of the genes. L and R refer to the left and right arm of the λ phage Charon 4, giving the orientation of the cloned DNA fragment in the vector. The vitellogenin mRNA is shown at the top and is drawn at the same scale as the genomic DNA fragments.

B) Diagram showing the regions of homology between different pairs of the most distant Xenopus vitellogenin genes. Pairs of genomic clones which together cover the genes with their flanking regions were heteroduplexed. The individual maps of 45 heteroduplex molecules for each pair of clones were drawn and summarized. The total length of the genomic DNA of a gene pair was normalized (100%) and the positions of the duplexed regions were plotted. For comparison, the different gene pairs (A1 x B1; A1 x B2; A2 x B1; A2 x B2) were aligned on their 5' and 3' ends which are indicated by arrows. The letters A to G indicate the most common prominent regions of homology. The diagram for the A1 x B1 gene pair is adapted from Wahli et al. (7).

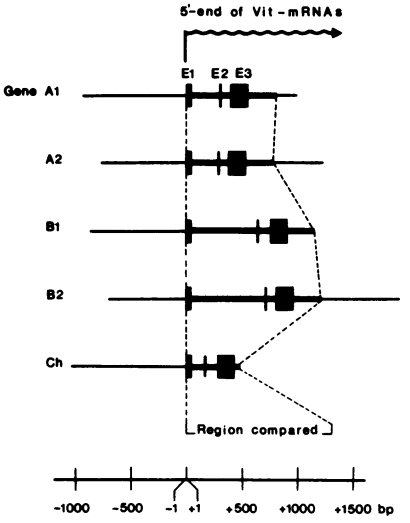
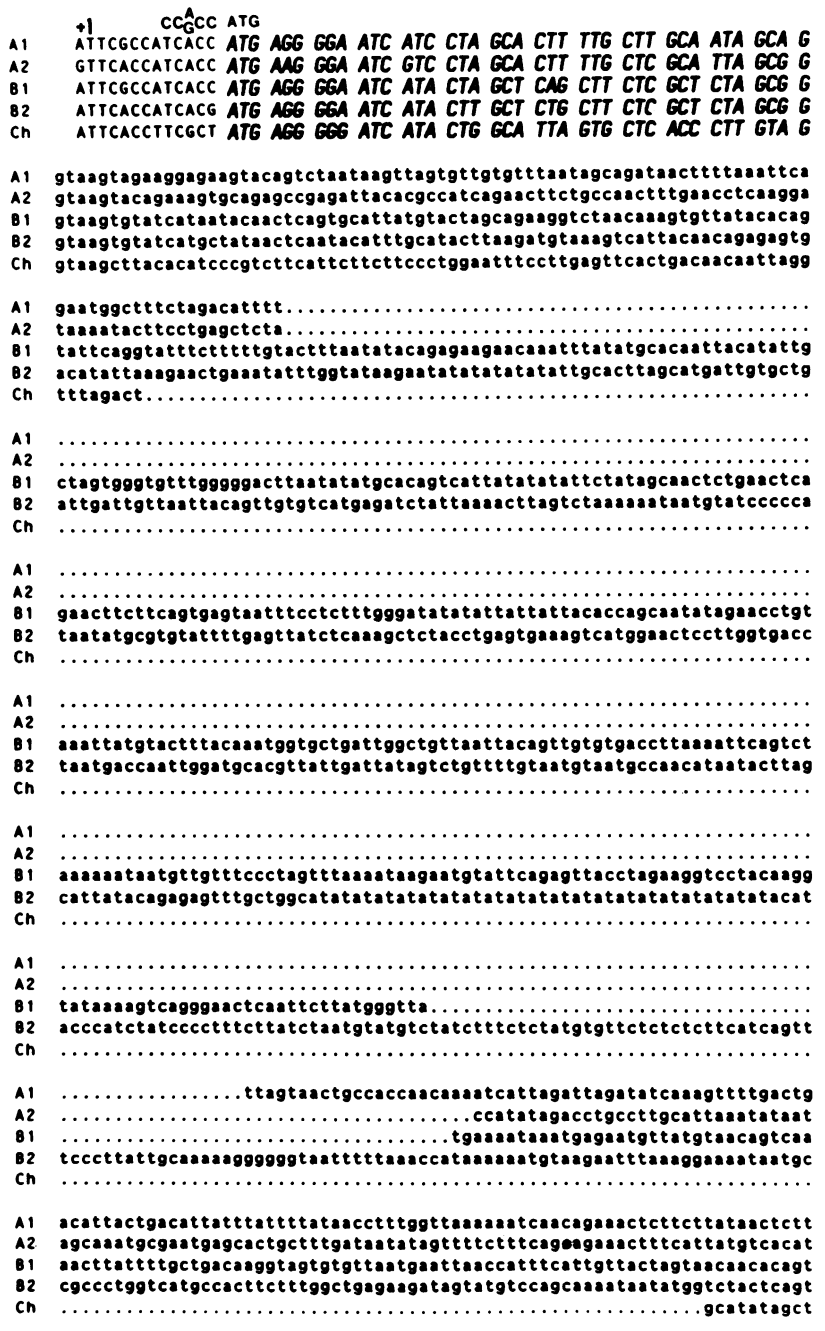


Figure 2. Structural organization of the 5' end region of five vitellogenin genes. The diagram indicates the region of the four Xenopus (A1, A2, B1, B2) and of the chicken gene (Ch) which were sequenced (A1, B1 and B2) or for which sequence data were available (A2 and Chicken, see ref. 21, 22, 23). Boxes E1, E2 and E3 represent exons which have the same length in all five genes compared, namely 53, 21 and 152 bp, respectively. The length of the introns in genes A1, A2, B1, B2 and Ch is for Intron 1, 239, 213, 578, 649 and 115 bp, and for Intron 2, 88, 80, 103, 91 and 100 bp, respectively. The region compared is given between the dashed lines.

Xenopus A2 and chicken gene (21, 22, 23). The presence of characteristic splice junctions (24) exclusively at these positions supports the arrangement proposed. An open reading frame of 71 amino acids is observed through the exons of all genes if the first AUG downstream of the transcription initiation site is taken as the translation initiation codon as described below.

Comparison of introns

The length variation of the first intron, the larger of the two analyzed, is much greater than for the second one (Fig. 2). To explore the relatedness between analogous introns, we searched for blocks of homology with two arbitrarily defined parameters : namely a minimum length of 30 nucleotides and a degree of homology of at least 65%. Using these criteria no homology was detected between analogous chicken and Xenopus introns, but our analysis


+1
CCACC ATG
A1 ATTCGCCATCACC ATG AGG GGA ATC ATC CTA GCA CTT TTG CTT GCA ATA GCA G
A2 GTTCACCATCACC ATG AAG GGA ATC GTC CTA GCA CTT TTG CTC GCA TTA GCG G
B1 ATTCGCCATCACC ATG AGG GGA ATC ATA CTA GCT CAG CTT CTC GCT CTA GCG G
B2 ATTCACCATCACG ATG AGG GGA ATC ATA CTT GCT CTG CTT CTC GCT CTA GCG G
Ch ATTCACCTTCGCT ATG AGG GGG ATC ATA CTG GCA TTA GTG CTC ACC CTT GTA G
A1 gtaagtagaaggagaagtacagtctaataagttagtgttgtgtttaatagcagataacttttaaattca
A2 gtaagtacagaaagtgcagagccgagattacacgccatcagaacttctgccaactttgaacctcaagga
B1 gtaagtgtatcataatacaactcagtgcattatgtactagcagaaggtctaacaaagtgttatacacag
B2 gtaagtgtatcatgctataactcaatacatttgcatacttaagatgtaaagtcattacaacagagagtg
Ch gtaagcttacacatcccgtcttcattcttcttccctggaatttccttgagttcactgacaacaattagg
A1 gaatggctttctagacatttt.....
A2 taaaatacttcctgagctcta.....
B1 tattcaggtatttctttttgtactttaatatacagagaagaacaaatttatatgcacaattacatattg
B2 acatattaaagaactgaaatatttggtataagaatatatatatattgcacttagcatgattgtgctg
Ch tttagact.....
A1 .....
A2 .....
B1 ctagtgggtgtttgggggacttaatatatgcacagtcattatatattctatagcaactctgaactca
B2 attgattgttaattacagttgtgtcatgagatctattaaaacttagtctaaaaaataatgtatccccca
Ch .....
A1 .....
A2 .....
B1 gaacttcttcagtgagtaatttcctctttgggatatatattattattacaccagcaatatagaacctgt
B2 taatatgcgtgtattttgagttatctcaaagctctacctgagtgaaagtcatggaactccttggtgacc
Ch .....
A1 .....
A2 .....
B1 aaattatgtactttacaaatggtgctgattggctgttaattacagttgtgtgaccttaaaattcagtct
B2 taatgaccaattggatgcacgttattgattatagtctgttttgtaatgtaatgccaacataatacttag
Ch .....
A1 .....
A2 .....
B1 aaaaaataatgttgtttccctagtttaaaataagaatgtattcagagttacctagaaggtcctacaagg
B2 cattatacagagagtttgctggcatatatatatatatatatatatatatatatatatatatatatacat
Ch .....
A1 .....
A2 .....
B1 tataaaagtcagggaactcaattcttatgggtta.....
B2 acccatctatccccttttcttatctaatgtatgtctatctttctctatgtgttctctctcttcatcagtt
Ch .....
A1 .....ttagtaactgccaccaacaaaatcattagattagatatcaaagttttgactg
A2 .....ccatatagacctgccttgcattaaatataat
B1 .....tgaaaataaatgagaatgttatgtaacagtcaa
B2 tcccttattgcaaaaaggggggtaatttttaaaccataaaaaatgtaagaatttaaaggaaaataatgc
Ch .....
A1 acattactgacattatttatttttataacctttggttaaaaaatcaacagaaactcttcttataactctt
A2 agcaaatgcgaatgagcactgctttgataatatagttttctttcagcagaaactttcattatgtcacat
B1 aacttattttgctgacaaggtagtgtgttaatgaattaaccatttcattgttactagtaacaacacagt
B2 cgccctggtcatgccacttctttggctgagaagatagtatgtccagcaaaataatatggtctactcagt
Ch .....gcatatagct

```
A1  cttctttgcttcttgtgtttttttccag GC TCT GAA AGA ACT CAA ATA G gtaggtttt
A2  tttgcttttt.....ctttccttcccag GC TCT GAA AGA ACT CAC ATA G gtaagtgct
B1  ctaaccattgtttatttattattcacag GA AGT GAA AAG TCA CAA TAT G gtaagtata
B2  ttaacagttgtttatttattattcacag GA TGC GAA AAA TCA CAA TAT G gtaagtaca
Ch  catgtggtttttctatctctttttgtag GC AGC CAG AAG TTT GAC ATT G gtaagtaca

A1  gtgctaagatcaccaatacaagttt..............cataatagaaatataacatgacactga
A2  tgctttagttaccaaat......................gcaattttaaattctgcatggtactga
B1  tttttgaatttatgtttttgcatttaaaaatatgcaatattttattgttatatttatacatacacat
B2  ttttaaaagtttatgtttgcataagaga............tatatatatatatatacaaacacacat
Ch  tttctacctataaacttggtgactttt...gttatgatgactattcattagaatatgcttacagcttc

A1  ctgttcttgtgcatctgtttatttttag  AG CCT GTG TTC AGT GAA AGC AAG ACA TCT
A2  ccattcttgtgcatttgtctattttag  AG CCT GTG TTC AGT GAA AGC AAG ATA TCT
B1  acacctttatcatctatgtattacag  AA CCT TTT TTC AGT GAG AGC AAG CCA TAT
B2  aaacttttaccatctgtatattacag  AA CCG TTT TTC AGT GAG AGC AAG ACA TAT
Ch  tatgtaaatggctgttattccccacag  AC CCA GGA TTC AAT AGC AGA AGG AGT TAC

A1  GTC TAT AAC TAT GAA GCT GTT ATC TTA AAT GGA TTT CCT GAA AGT GGT TTG
A2  GTG TAT AAC TAT GAA GCT GTC ATA CTG AAT GGA TTT CCT GAA AGT GGT TTG
B1  GTG TAC AAT TAC GAA GGC ATT ATT CTT AAT GGA ATC CCA GAA AAT GGT TTG
B2  GTG TAC AAT TAT GAA GGC ATT ATT CTT AAT GGA ATC CCA GAA AAT GGT TTG
Ch  CTG TAC AAC TAT GAA GGT TCT ATG TTG AAT GGG CTT CAA GAC AGA AGT TTG

A1  TCT CGG GCT GGT ATT AAA ATT AAC TGC AAG GTT GAG ATC AGC GCC TAT GCC
A2  TCC CGG GCT GGT ATT AAA ATT AAC TGC AAA GTT GAG ATC AGC GCC TAT GCT
B1  GCC CGG TCT GGT ATT AAA CTG AAC TGC AAG GCT GAG ATC AGT GGC TAT GCC
B2  GCC CGG TCT GGT ATT AAA TTG AAC TGC AAG GTT GAG CTC AGT GGC TAT GCG
Ch  GGC AAA GCT GGT GTG CGC TTG AGC AGC AAG CTA GAG ATC AGT GGG CTA CCA

A1  CAG AGG TCC TAC TTT CTA AAG gtaagtaccacttgcgtttgcttctgtttttaaaataaa
A2  CAG AGG TCA TAC TTC CTA AAG gtaagtaccatttgcccctattttaaaaaagaaaaaaaa
B1  CAG AGG TCC TAC ATG CTA AAG gtaagccataaaagagacaatctctttgaaaatgagact
B2  CAG AGG TCC TAC ATG CTA AAG gtaagacataaaagagacagtcactttaagaataaaact
Ch  GAG AAT GCT TAC CTC CTC AAG gtactggccatgtcttgttccaaacgcaccaaccaacac

A1  gccatgttcagaattgaagatattacagaataactcagaaaatgttaaaacagtaacaccttaattt
A2  aaaccacatggtcacccctaattagtgatgtaaacacttaaaattgattagtctgttcgttattcat
B1  acagatatgactgtgctcactgtttatatttctggtaacagcttcaaactgaaaaagggatgttga
B2  gcagatgtggctgcactggtgctcagtatttagattcctagtaaaagcttaaaaatgaaaaggactg
Ch  tgaattc

A1  tatatagttacagagttcaacatagctttgatactctgagatgagaataatgcaggaatacaaaaaa
A2  atgatagcaccatatatctgaacatacaaaaaggcagtgcagcacagctacaatgtatgaagccga
B1  ttaattttcaaccaaaataaatcagtactcttttattatttacatctgatgccagtagaaaaatatg
B2  cacattaattttcaagtaatataaatcaatggctttgtactctcttatcatttacatacaatgccaa

A1  aataaaaaaaaagtgagattaattaaggacatatcagtataatcctaaaatgtacccatattaaattt
A2  tattttaatttcagttaacatgaaatactaggcaactgggcacatatacagtaattgtttatattta
B1  aggaagatcccaagaaatgtttctttaccagtaccccactagtgtaccctgcagtatatatatacag
B2  tagaaaaatatcaaggggatcccaaataatgtttaataatgtttttataccagtaccccactagtgc
```

Figure 3. Display of the nucleotide sequences of the 5' end transcribed region (see Fig. 2) of the Xenopus laevis genes A1, A2, B1, B2 and of the chicken gene (Ch). The nucleotide sequence of the non-coding (mRNA strand) is given in the 5' to 3' direction. Exon sequences are in capital letters arranged in triplets for the coding regions; introns are in lower case letters. Dots were inserted to achieve optimal alignment of the exons. The consensus sequence for translation initiation sites proposed by Kozak (26) is given. The A2 gene sequence is taken from Walker et al. (21) and completed and the Ch gene sequence is taken from ref. 21 and 23.

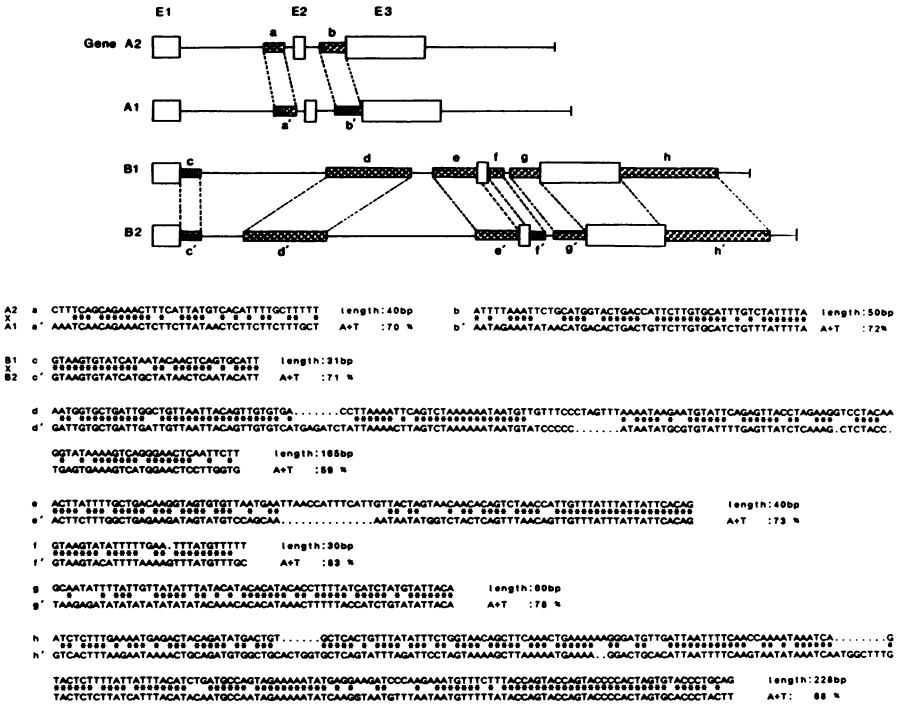


Figure 4. Homologies between introns of the Xenopus genes. The computer search for homology blocks was performed with the following criteria : minimal block length of 30 nucleotides with 65% homology. The different regions found were lettered from a/a' to h/h'. Open boxes in the drawing are exons (E1, E2, E3). The DNA sequence of the homology blocks is shown.

revealed several homologies between corresponding introns of the closely related Xenopus gene pairs (Fig. 3 and 4). The length of the blocks varies between 30 and 230 base pairs. Most are close to or directly flank exons and show a higher A+T content than the surrounding sequences (Fig. 4). The d/d' and h/h' blocks differ from the others by their length of 160 and 228 base pairs and their relatively low A+T content of 59 and 68%, respectively. The main observation is that there are more than three times as many homologous sequences, as defined above, between B1 and B2 introns (about 50% of the intron sequences analyzed) than between A1 and A2 introns. These results are in good agreement with previous electron microscopic analyses (8, 16).

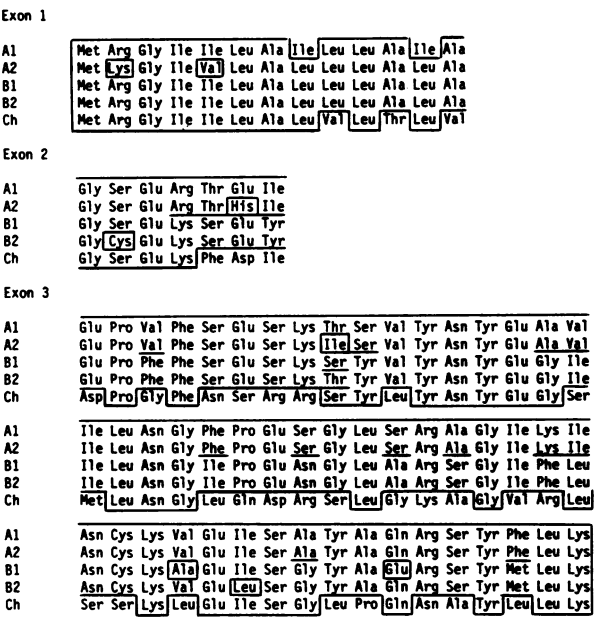


Figure 5. The amino acids encoded in the first three exons of the compared genes. Positions with two or more identical amino acids are enclosed in boxes.

Comparison of exons

The first three homologous exons have the same length in all five genes compared, namely 53, 21 and 152 nucleotides (Fig. 3). There are only 13 nucleotides from the 5' end of the mRNAs to the first AUG. This in all probability serves as the translation initiation codon for the following reasons:

a) it is the first ATG from the transcription initiation site and the first codon of an open reading frame through the three exons analyzed in all five genes. In a compilation of 211 messenger RNAs, Kozak has shown that in 95% of the cases the first AUG serves as the initator codon (25);

b) as pointed out by Kreil (26) most signal peptides of secreted proteins contain, besides the amino terminal methionine, a charged amino acid near the amino end, frequently an arginine or lysine. Figure 5 shows that in four of the five coding sequences there is an arginine as second amino acid and

Table I : Percent divergence between vitellogenin genes

	Observed percentages of changes				Corrected percentages of changes (Perler et al., ref.14)[a]		
	Exon 1	Exon 2	Exon 3	Exon 1-3	Silent changes	Replac. changes	Silent ch. Replac.ch.
A1 x A2	13.2	4.8	5.3	7.1	28.8	3.4	8.6
x B1	18.9	47.6	19.7	22.1	74.1	17.5	4.2
x B2	22.6	42.9	19.7	22.6	84.4	15.4	5.5
x Ch.	30.2	57.1	43.4	41.6	128.3	45.0	2.9
A2 x B1	22.6	52.4	21.1	24.3	81.9	19.1	4.3
x B2	22.6	47.6	20.4	23.5	82.8	16.9	4.9
x Ch.	34.0	52.4	40.1	39.8	105.7	46.2	2.3
B1 x B2	7.5	14.3	4.6	6.2	22.2	3.3	6.8
x Ch.	30.2	47.6	40.8	38.9	75.1	42.2	1.8
B2 x Ch.	28.3	52.4	38.8	37.6	67.7	41.0	1.6

a) The calculations include all codons of the three first exons of the five genes

lysine in the fifth one (gene A2). Then follows a stretch of ten or more non-polar amino acids;

c) Figure 3 reveals a predominance of C upstream from the ATG and a purine (A in the Xenopus genes; G in the chicken gene) three nucleotides upstream of the initiator codon. Thus, the sequences at the putative translation initiation site of the 5 vitellogenin mRNAs agree well, within the variability observed for other mRNAs with the consensus sequence $CC\begin{matrix}A\\G\end{matrix}CCAUG$ proposed by Kozak (26).

Table I gives the divergence between the homologous exons. The observed means of divergence of the three exons are approximately 6-7% between the Xenopus A1 and A2 or B1 and B2 genes, 22-25% between the A and B genes and 37-41% between Xenopus and chicken genes. The short Exon 2 of 21 bp is the least conserved in every pair except for A1 x A2. In the A x B pairs for

instance, the differences between Exon 2 are twice as great as between Exon 1 or Exon 3. The corrected divergence percentage between the homologous coding regions was calculated according to Perler et al. (14) taking possible multiple changes into account. It is presented as silent substitutions (S) and replacement substitutions (R) in Table I. The ratio "silent substitutions" over "replacement substitutions" is drastically reduced with evolutionary time.

DISCUSSION

Relatedness between vitellogenin genes

Previous heteroduplex and R-loop analyses of pairs of the closely related Xenopus genes (A1 x A2; B1 x B2) showed that sufficient homology to form stable duplexes under the given conditions was limited to exon sequences for the A1 x A2 pair (16). In contrast, paired regions between the other closely related genes (B1 x B2) involved, in addition to exon sequences, about 50% of the intron sequences (8) despite the fact that exons have diverged to the same extent in the two pairs. Here we extend this analysis and screen for homologies between less related gene pairs (A x B genes). At low stringency, short regions of homology were found in each pair at similar positions; these reveal regions which have been preferentially conserved after the gene duplication events. This broad analysis is extended by a comparison of the DNA sequences at the 5' end of four Xenopus and one chicken vitellogenin gene. Interesting features concerning intron as well as exon sequences are revealed. The length of the first three exons of the five compared genes has been strictly conserved. The presence of the very short second exon of such a long gene (16-22 kb) in different species is most likely due to positive selection. A possible role for the second exon is to encode part or all of the signal peptidase recognition site. This would allow the removal of the hydrophobic N-terminal signal sequence whose general characteristics can be found in the amino acids encoded by the first exon. Analysis of the last amino acids of Exon 1 and the first few of Exon 2 agrees with the non-random amino acid utilization around cleavage sites as defined by von Heijne (27).

Two introns were sequenced in their entirety. Length variation is much

greater in the first and larger intron compared to the second whose length is relatively constant. Similar observations have been made with globin and preproinsulin genes (14, 28). Homologies within otherwise rapidly diverging intron sequences might not necessarily be the result only of selective pressure. If the A1-A2 and B1-B2 gene duplications occurred simultaneously as suggested by the divergence in coding regions (ref. 9 and below) it is rather surprising that A introns have diverged more rapidly than B introns (see also ref. 8). At this stage of our analysis of the vitellogenin gene family, it is not possible to formulate a convincing interpretation of this observation. Local exchange events or position effects might have contributed to the apparently higher conservation of the B introns.

Evolution of vitellogenin genes

By comparing the amino acid sequence encoded in the first three exons of the 5 genes we see various degrees of homology between the different genes. According to the evolutionary clock hypothesis this reflects the divergence time of the analyzed sequences (29). We compared coding regions using the method of Perler et al. (14) which considers both the base changes leading to an amino acid change (replacement) and the changes at silent sites producing synonymous codons. The calculation provides the percent corrected divergence because it considers multiple events. It is thought that a whole genome duplication took place in Xenopus about 30 million years ago (30). This event would have produced the A1/A2 and B1/B2 pairs of vitellogenin genes. In addition, since the amphibian/reptile-bird divergence occurred about 330-350 million years ago (31) it is possible to plot the accumulation of changes as a function of divergence time. As shown in Figure 6, replacement changes accumulate linearly with time thus providing a good evolutionary clock. One percent replacement changes are fixed every 8.8 million years. Using this clock we can estimate when the original Xenopus vitellogenin gene was first duplicated to form the ancestral A and B genes. As seen in Figure 6 the A-B divergence rate suggests that this first duplication event in Xenopus occured about 150 million years ago.

Changes at silent sites accumulate much more rapidly than at the replacement sites and large amount of scatter is observed in the rates (Fig. 6). In contrast to the accumulation of replacement changes accumulation of silent

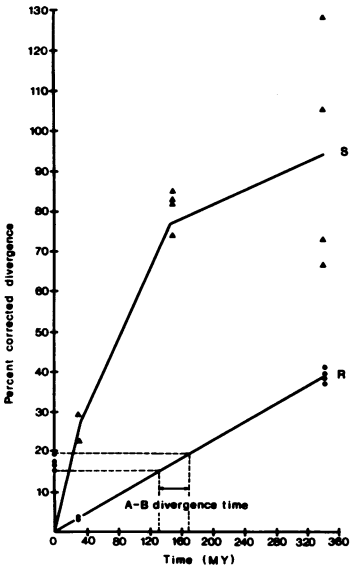


Figure 6. Sequence divergence at silent (S) and replacement sites (R) in function of divergence time. Percent corrected changes at silent (▲) and replacement (●) sites in the codons of the three first exons of the five analyzed genes were calculated according to Perler et al. (14). Based on the percent corrected divergence at silent sites between A and B gene sequences the predicted divergence time given by the dashed lines is between 130 and 170 million years. The percent corrected divergence at silent sites between A and B gene sequences was therefore plotted against a divergence time of 150 million years (mean between 130 and 170).

substitutions is not linear and its rate diminishes with time. At first (see Table I), it is seven to nine times more rapid than at the replacement sites (A1-A2; B1-B2), then four to six times more rapid (A-B) and finally only about two to three times more rapid (Xenopus-chick). Thus, there is selective pressure on the part of the silent sites that is imposed by constraints which are still not well understood. These general observations agree with those made by Perler et al. (14) for the preproinsulin and globin genes. Comparison of accumulation rates reveals that the constraints on the N-proximal peptide region of the vitellogenin genes are just slightly lower than those imposed on globin (14). Further analysis will reveal if these constraints are similar for the entire coding domain of the genes or if, alternatively, regions encoding the different cleavage products (yolk proteins), each with its quite specific characteristics, have fixed mutations at different rates.

ACKNOWLEDGEMENTS

We wish to thank Carolyn Tolstoshev for the computer program used to analyze the nucleotide sequences, Roland Sahli for gift of material, Françoise Givel for excellent technical assistance, Bob Hipskind, John Knowland and Riccardo

Wittek for comments on the manuscript, Hannelore Pagel and Nelly Buchi for typing the manuscript. This work was supported by the Etat de Vaud and by the Swiss National Science Foundation.

REFERENCES

1) Wallace, R.A. (1978), In Vertebrate Ovary, Jones R.E., ed., pp. 469-502, Plenum New York.
2) Wahli, W.; Dawid, I.B.; Ryffel, G.U. and Weber, R. (1981), Science 212, 298-304.
3) Shapiro, D.J. (1982), CRC Critical Rev. Biochem. 12, 187-203.
4) Ryffel, G.U. and Wahli, W. (1983), In Eukaryotic genes : their structure, activity and regulation. Maclean, N; Gregory, S.P. and Flavell, R.A. eds, pp. 329-341, Butterworths London.
5) Wahli, W.; Dawid, I.B.; Wyler, T., Jaggi, R.B.; Weber, R. and Ryffel, G.U. (1979), Cell 16, 535-549.
6) Wahli, W. and Dawid I.B. (1980), Proc. Natl. Acad. Sci. USA, 77, 1437-1441.
7) Wahli, W.; Germond, J.E.; ten Heggeler, B. and May, F.E.B. (1982), Proc. Natl. Acad. Sci. USA, 79, 6832-6836.
8) Germond, J.-E.; ten Heggeler, B.; Schubiger, J.-L.; Walker, P.; Westley, B. and Wahli, W. (1983), Nucleic Acids Res., 11, 2979-2997.
9) Wang, S.-Y. and Williams, D.L. (1980), Biochemistry 19, 1557-1563.
10) Wang, S.-Y; Smith, D.E. and Williams, D.L. (1983), Biochemistry 22, 6206-6212.
11) Arnberg, A.C.; Meijlink, F.C.P.W.; Mulder, J.; van Bruggen, E.F.J.; Gruber, M. and AB, G. (1981), Nucleic Acids Res. 9, 3271-3286.
12) Wilks, A.; Cato, A.C.B.; Cozens, P.J.; Mattaj, I. and Jost, J.-P. (1981), Gene 16, 249-259.
13) Burch, J.B.E. and Weintraub, H. (1983), Cell 33, 65-76.
14) Perler, F.; Efstratiadis, A.; Lomedico, P.; Gilbert, W.; Kolodner, R. and Dodgson, J. (1980), Cell 20, 555-566.
15) Efstratiadis, A.; Posakony, J.W.; Maniatis, T.; Lawn, R.M.; O'Connell, C.; Spritz, R.A.; DeRiel, J.K.; Forget, B.G.; Weissman, S.M.; Slightom, J.L.; Blechl, A.E.; Smithies, O.; Baralle, F.E.; Shoulders, C.C. and Proudfoot, N.J. (1980), Cell 21, 653-668.
16) Wahli, W.; Dawid, I.B.; Wyler, T.; Weber, R. and Ryffel, G.U. (1980), Cell 20, 107-117.
17) Messing, J.; Crea, R. and Seeburg, P.H. (1981), Nucleic Acids Res. 9, 309-321.
18) Sanger, F.; Nicklen, S. and Coulson, A.R. (1977), Proc. Natl. Acad. Sci. USA 74, 5463-5467.
19) Barnes, W.M. and Bevan, M. (1983), Nucleic Acids Res. 11, 349-368.
20) Hong, G.F. (1982), Journal of Molecular Biology 158, 539-549.
21) Walker, P.; Brown-Luedi, M.; Germond, J.-E.; Wahli, W.; Meijlink, F.P.W.; van het Schip, A.D.; Roclink, H.; Gruber, M. and AB, G. (1983), The EMBO Journal 2, 2271-2279.
22) Geiser, M.; Mattaj, I.W.; Wilks, A.F.; Seldran, M. and Jost, J.P. (1983), J. Biol. Chem. 258, 9024-9030.
23) Burch, J.B.E. (1984), Nucleic Acids Res. 12, 1117-1148.
24) Mount, S.M. (1982), Nucleic Acids Res. 10, 459-472.

25) Kozak, M. (1984), Nucleic Acids Res. 12, 857-872.
26) Kreil, G. (1981), Ann. Rev. Biochem. 50, 317-348.
27) von Heijne, G. (1984), Journal of Molecular Biology 173, 243-251.
28) Blanchetot, A.; Wilson, V.; Wood, D. and Jeffreys, J. (1983), Nature 301, 732-734.
29) Wilson, A.C.; Carlson, S.S. and White, T.J. (1977), Ann. Rev. Biochem. 46, 573-639.
30) Bisbee, C.A.; Baker, M.A.; Wilson, A.C.; Hadzi-Azimi, J. and Fischberg, M. (1977), Science 195, 785-787.
31) Carroll, R.L. (1969). In Biology of the reptilia Gans, C.; Bellairs A. d'A. and Parsons T.S. eds, vol.1, pp.1-44, Academic Press, London and New-York.