J Mol Evol (1989) 28:185–190

Journal of Molecular Evolution

Characterization of a Polymorphism in the 3′ Part of the Chicken Vitellogenin Gene

J.N.J. Philipsen, J.E. de Vries, J. Samallo, C. van Dijk, A.C. Arnberg, and G. AB

Laboratory of Biochemistry, Nijenborgh 16, 9747AG Groningen, The Netherlands

Summary. An allele giving rise to a polymorphism within the 3′ part of the chicken vitellogenin gene was cloned, sequenced, and compared to the previously cloned allele. The polymorphism is formed by a perfect copy of 343 bp from intron 32 in tandem array with a perfect copy of 244 bp from intron 33; this 587-bp element is inserted in a head-to-tail arrangement in intron 33. We propose a mechanism in which an unequal crossing-over resulted in a vitellogenin gene with two exons 33, one of which was subsequently deleted. Thus, intron 33 was enlarged by the tandem repeats without affecting the protein-encoding sequence of the gene. At the boundaries of the repeated elements, two short direct repeats are found that resemble the recombination signals of immunoglobulin genes. They may have had a key role in the formation of the new allele.

Key words: *Gallus domesticus* — Vitellogenin gene — Allelic polymorphism — Insertion-deletion model

Introduction

The faithful passage of genetic information from one generation to the next is based on the accuracy of the replication process. It prevents species from loosing valuable information contained in their DNA including the amino acid sequences encoded in their genes and the sequences involved in regulation of gene expression. On the other hand, many examples of changes in eukaryotic genomes have been documented, ranging from large translocations and deletions that are visible at the chromosomal level (e.g., Nowell and Hungerford 1960) to point mutations that become apparent upon DNA sequence analysis (e.g., Valerio et al. 1986). A detailed analysis of these changes may reveal the molecular mechanisms underlying them.

In this paper, we describe the cloning and characterization of an allelic variant in the 3′ part of the chicken vitellogenin (Vtg) gene. Vtg is a yolk precursor protein that is synthesized in the liver of oviparous vertebrates under the influence of estradiol. Recently, we have cloned (Arnberg et al. 1981) and sequenced (Van het Schip et al. 1987a) a chicken Vtg gene encoding the major yolk precursor protein, VtgII (Wang et al. 1983). During a study on the methylation status of the Vtg gene by Southern blot analysis (Philipsen et al. 1985), it appeared that in all animals examined the restriction map of the gene was identical except for the 3′ part. The new allele contains an insertion in intron 33 that is characterized here; its possible origin is discussed.

Materials and Methods

Animals and DNA Analysis. White Leghorn chickens (*Gallus domesticus*) were obtained from Poultry Hatchery Van der Sterren, Venray, The Netherlands. DNA was isolated from liver, erythrocytes, and oviduct as described (Philipsen et al. 1985), digested with EcoRI, and analyzed for the presence of the polymorphism by Southern blotting.

Molecular Cloning of the Longer Allele—Partial Genomic Bank. Liver DNA (100 μg) from a chicken that was homozygous for the longer allele was digested to completion with BamHI. After size fractionation on a 0.75% agarose gel, the region containing the 12–14-kb fragments was cut out and the DNA recovered by

Offprint requests to: G. AB

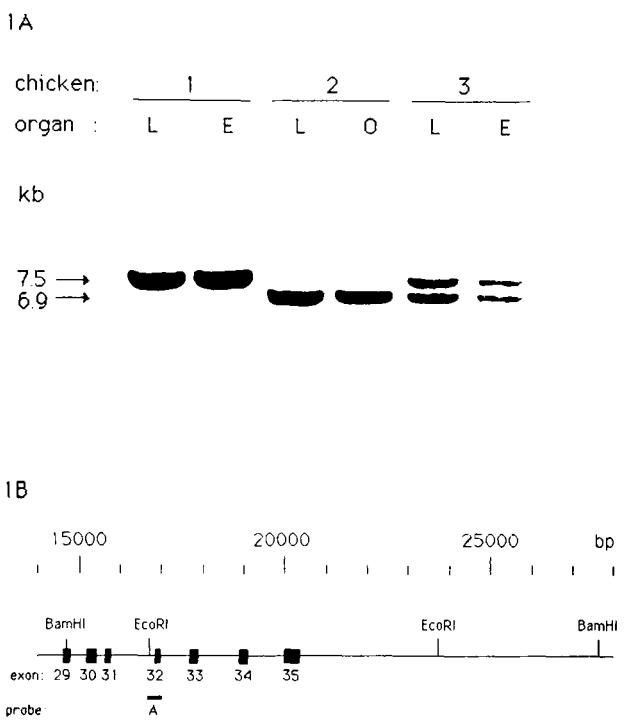

Fig. 1. A polymorphism within the 3'-end region of the chicken vitellogenin gene. **A** A Southern blot of EcoRI-digested DNAs from three different chickens (1, 2, and 3), hybridized to probe A (see **B**). The DNA was prepared from liver (L), erythrocytes (E), or oviduct (O). **B** A schematic map of the 3'-end region of the Vtg gene, with the positions of the relevant restriction sites and the site of probe A. The scale measures the distance in bp from the cap site (Van het Schip et al. 1987a).

electroelution (Maniatis et al. 1982). The yield was approximately 10 μg. The purified DNA (150 ng) was ligated to 0.5 μg EcoRI–BamHI-cut λ-EMBL3 DNA (Frischauf et al. 1983). In vitro packaging of the recombinant phage DNA and plating onto *Escherichia coli* NM359 yielded 80,000 plaques. In situ plaque hybridization (Maniatis et al. 1982) resulted in the identification of two clones containing the desired insert.

Subclones containing the insertion region were obtained by cloning the 7.5-kb EcoRI and the 2.0-kb PstI fragments into pBR329 (Covarrubias and Bolivar 1982). The similar 6.9-kb EcoRI fragment of the previously isolated allele (Arnberg et al. 1981) was also cloned into pBR329.

Heteroduplexing. Heteroduplexes between the purified 6.9-kb and 7.5-kb EcoRI fragments were made as described in Lenstra et al. (1986); further treatments and electron microscopy were done according to Arnberg et al. (1980). Length measurements were made with a Tulip PC coupled to an Ahrin GTCO digitizer DP5A-2436A (HA).

Sequencing and Sequence Analysis. The pBR329 plasmid containing the 2.0-kb PstI fragment was used as a DNA source for cloning into M13mp18/19 and pEMBL 18/19. Sequencing reactions were done according to Sanger et al. (1977) or Zagursky et al. (1985). Sequence analyses were performed on an AppleII microcomputer linked to a VAX computer using the Staden program (Staden 1982).

Suppliers. Enzymes were purchased from Boehringer Mannheim, FRG. Radiochemicals were from The Radiochemical Centre Amersham, UK, and blotting membranes from Schleicher and Schuell, Dassel, FRG.

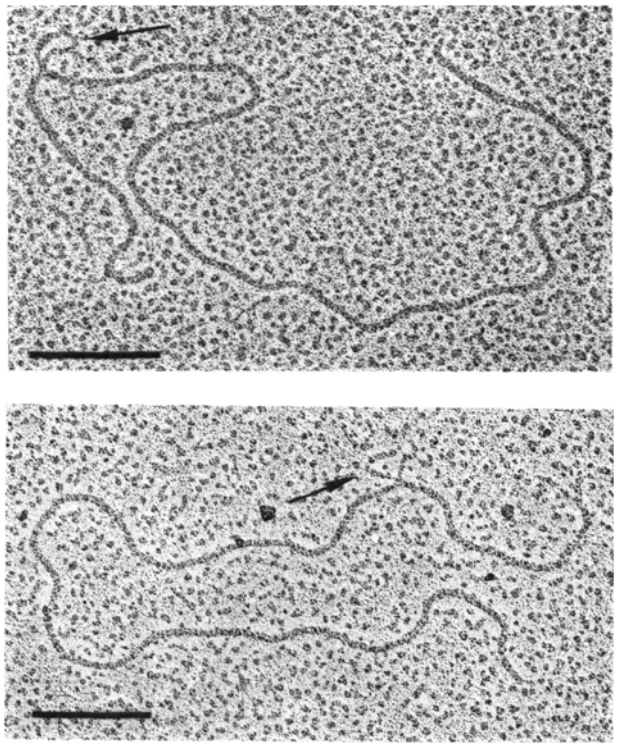

Fig. 2. Heteroduplex structure formed between the 6.9-kb and 7.5-kb EcoRI fragments. A nonhomologous region is visible as a single-stranded loop (arrows). The bars represent 0.2 μm.

Results

Localization of an Allelic Variant within the Chicken Vitellogenin Gene

A Southern blot of genomic DNA from different chickens demonstrating the polymorphism in the 3'-moiety of the Vtg gene is shown in Fig. 1. DNAs from chickens 1 and 2 contain one hybridizing EcoRI fragment of 7.5 kb and 6.9 kb, respectively, characteristic for animals having a homozygous gene pair. Chicken 3 is heterozygous for the polymorphism. In the population examined (17 animals); the allele frequency of the longer allele was 0.73. Earlier, a Vtg gene was cloned (Arnberg et al. 1981) and sequenced (Van het Schip et al. 1987a) in our laboratory. Because this gene represents the shorter allele (data not shown), we set out to clone and sequence the corresponding region from the longer allele. To this end, a 13-kb BamHI fragment containing the 7.5-kb EcoRI fragment was cloned from DNA of chicken 1.

To localize the difference between the 7.5-kb and 6.9-kb EcoRI fragments, we performed a heteroduplexing experiment and comparative restriction enzyme analysis. Electron micrographs of the heteroduplex (Fig. 2) clearly show that the fragments are seemingly similar over their entire length, except for a small region apparent as a single-stranded loop

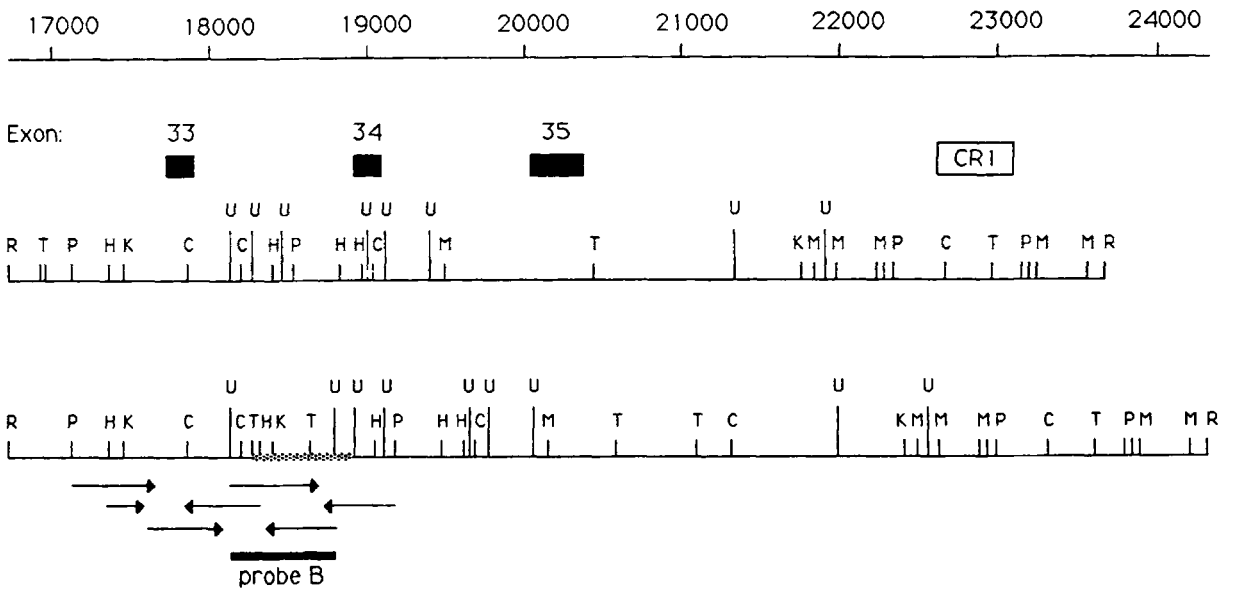

Fig. 3. Comparative restriction enzyme maps of the 6.9-kb and 7.5-kb EcoRI fragments. The upper map depicts the 6.9-kb EcoRI fragment and is in accordance with the nucleotide sequence of the Vtg gene cloned earlier (Van het Schip et al. 1987a). The three last exons and a repetitive CR1 element, present in the 3'-flanking region (Van het Schip et al. 1987b), are indicated. The scale measures the distance in bp from the cap site. The lower map represents the 7.5-kb EcoRI fragment, with the insertion indicated as a stippled bar. The sequencing strategy of the 2.0-kb PstI fragment containing the insertion is depicted by arrows indicating the length and the direction of the sequences determined. The site of probe B, which was used in the experiment shown in Fig. 6A, is given. C = CfoI; H = HindIII; K = KpnI; M = MspI; P = PstI; R = EcoRI; T = TaqI; U = PvuII.

of approximately 0.6 kb, located about 1.4 kb from the proximal EcoRI site.

The same conclusion can be drawn from the restriction maps of both fragments (Fig. 3), which are similar except for an additional sequence located around exon 33 in the longer allele. This insertion appears to be large enough to account for the observed difference in EcoRI restriction fragment sizes.

Sequence of the Polymorphic Region

To analyze the nature of the insertion in more detail, we have sequenced the relevant region that is contained in the 2.0-kb PstI fragment. The sequencing strategy is outlined in Fig. 3 and the complete sequence is presented in Fig. 4. The PstI fragment is 1985 bp long, which is 584 bp more than the corresponding fragment in the shorter allele and contains one perfect copy of exon 33. Analysis of the entire sequence reveals that the larger fragment size is due to repetition of two intron sequences, one extending from positions 17,350 to 17,692 in intron 32 and another from positions 17,940 to 18,183 in intron 33 (numbering according to Van het Schip et al. 1987a). Apparently, the longer allele has a structure that is schematically presented in Fig. 5. This structure may be due to a tandem duplication of the 17,350–18,183 region followed by the deletion of an internal segment, including exon 33, from the duplicate. The duplicated segments are not positioned contiguously, but are separated by a T, that is possibly a remnant of the duplication event. Elsewhere in the compared sequences, we find 11 additional differences between both alleles, most of which are single base substitutions. These differences, as well as those between the duplicated segments, are indicated in the sequence (Fig. 4) and the schematic representation (Fig. 5). In the corresponding sequences, the interallelic differences appear to outnumber the intraallelic differences.

Repetitive Nature of the Duplicated Elements

Concerning the mechanism by which the observed polymorphism has arisen, the following observation might be relevant. When a Southern blot of EcoRI-digested chicken DNAs was hybridized to a probe specific for the duplicated region (Fig. 3), a number of discrete bands, superimposed on a faint smear of DNA, appeared in addition to the expected 7.5-kb and 6.9-kb fragments (Fig. 6A). When the same blot was hybridized for comparison to a probe specific for single-copy DNA, only the expected fragments appeared (Fig. 6B). Thus, the duplicated sequences must be repeated within the genome to some extent. The degree of repetition is probably low, as can be concluded from Southern blots of cloned DNA hybridized to nick-labeled genomic chicken DNA (Shen and Maniatis 1980). No signal is found with fragments containing the duplicated sequences (data not shown), whereas strong hybridization is observed with fragments from the 3'-flanking region of the Vtg gene that contains the repetitive CR1 element

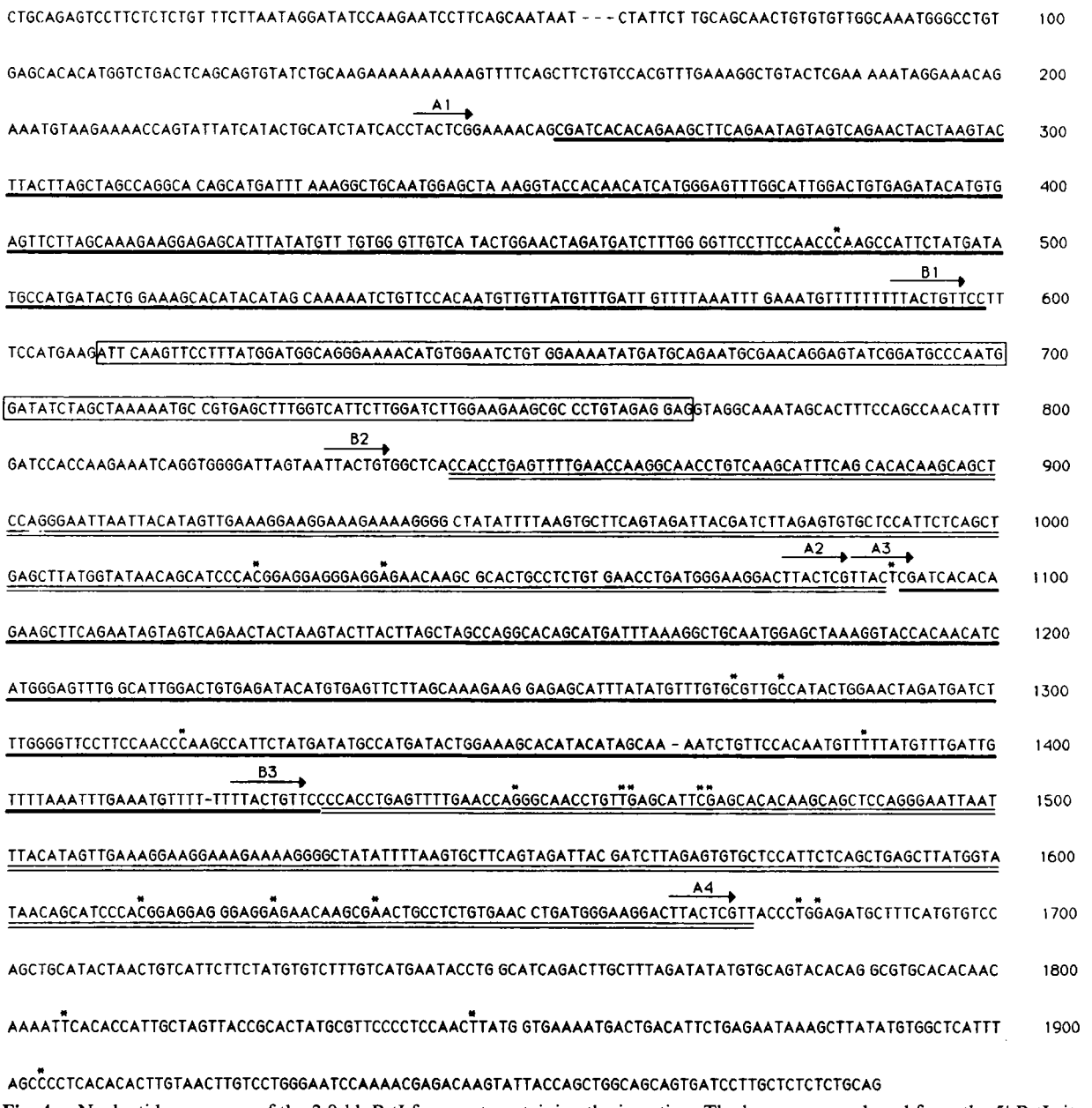

Fig. 4. Nucleotide sequence of the 2.0-kb PstI fragment containing the insertion. The bases are numbered from the 5′ PstI site (see Fig. 3). Bases differing from the previously sequenced allele are indicated (*); "—" indicates a deleted base (see Fig. 5). Element I and (duplicated) element I′ are singly underlined; element II and (duplicated) element II′ doubly (see Fig. 5). Exon 33 is boxed. Short direct repeats are indicated by arrows above the sequence.

(Fig. 3) (Van het Schip et al. 1987b). The CR1 element is repeated about 7000 times in the haploid genome (Stumph et al. 1981). We estimate that the degree of repetition of the duplicated intron 32/33 sequences is at least one order of magnitude less than that of CR1.

Individual chickens from our flock showed different patterns of additional bands (Fig. 6A) that probably represent allelic length polymorphisms. Between the individual animals, the segregation of these bands is different from the 6.9/7.5-kb fragments, which shows that the repeats are not strongly linked to the Vtg gene region, but are located elsewhere in the genome.

Discussion

It is reasonable to assume that the longer allelic variant of the chicken Vtg gene has arisen from the shorter one. The mechanism by which the rearrangement has taken place is far from clear; a likely possibility is an unequal crossing-over followed by a deletion, as suggested by the schematic representation in Fig. 5. Analysis of the sequence has revealed a number of small direct repeats that might have played a part in the presumed rearrangements (Fig. 4). One repeat, TACTCG, is found at a short distance in front of the supposed crossing-over points (Fig. 5) at positions 17,336 and 18,173 of the shorter

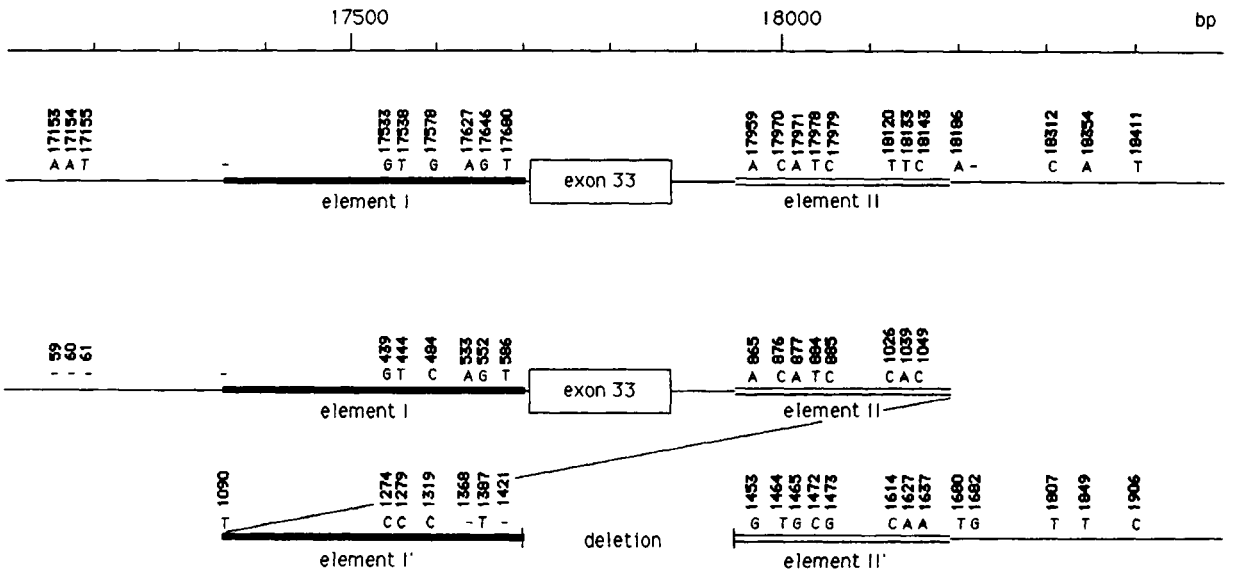

Fig. 5. Schematic representation of the two allelic variants. The upper part represents the shorter allele; the numbering is according to Van het Schip et al. (1987a). In the lower part, the longer allele is shown; numbers are relative to the 5′ PstI site (Fig. 4). Inter- and intraallelic differences are indicated with their relative positions above them. Element I and (duplicated) element I′ are depicted by black bars; element II and (duplicated) element II′ by double lines.

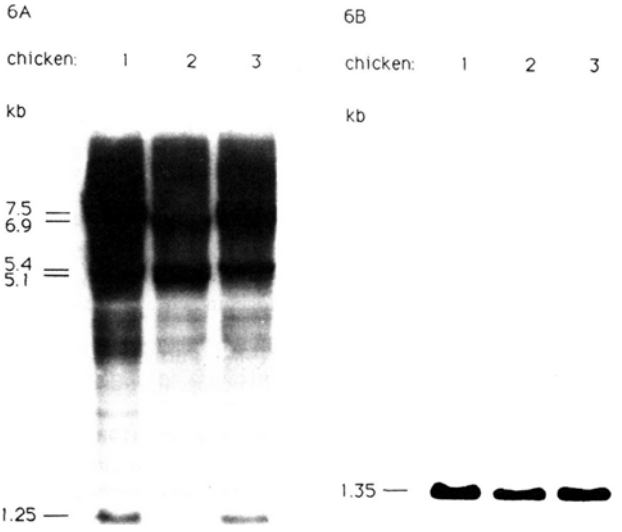

Fig. 6. The polymorphic region contains low-repetitive DNA. Liver DNA (10 μg) from chickens 1, 2, and 3 (see Fig. 1) was digested with EcoRI, electrophoresed on a 0.8% agarose gel, and blotted. **A** The blot was hybridized to probe B (Fig. 3), a PvuII fragment containing exactly all the duplicated sequences. **B** The same blot was rehybridized to an EcoRI fragment from the Vtg gene promoter, which should give rise to a 1.35-kb band.

allele (A1 and A2 in Fig. 4). As a consequence of the rearrangement and through the insertion of a T between the duplicated segments, an additional repeat of this kind is formed at the boundary of the duplicated segments (A3 in Fig. 4). Another repeat, TTACTGT, is found a few nucleotides upstream from the boundaries of the deleted segment (Fig. 5) at positions 17,683 and 17,927 of the shorter allele (B1 and B2 in Fig. 4).

These repeats show similarity to the heptamer recombination signals found in Ig genes (Tonegawa 1983; Reynaud et al. 1987), which may point to their involvement in the supposed rearrangements. Moreover, these repeats may not only have caused the duplication/deletion events, but also have facilitated the spread of the duplicated sequences through the genome (Fig. 6A). Generally, such a process is mediated via transposable genetic elements that are characterized by short inverted repeats at their boundaries. Our sequences however do not have these typical transposon-like features, and consequently must have been spread via a different mechanism.

Whatever the origin of the longer allele may be, its altered configuration will have no effect on the final gene product. Although sequences on both sides of exon 33 are duplicated, the exon itself is not, and remains unaffected. Moreover, as far as one can deduce from the sequence, the splicing signals are left intact as well.

Involvement of CG Residues in Point Mutations

The number of nucleotide differences between the shorter and the longer allele is relatively small, and restricted to the introns. The duplicated elements I′ and II′ appear to be less well conserved (Fig. 5). However, the strong conservation of these elements in general suggests that the duplication event occurred very recently on an evolutionary time scale. CG doublets within the sequenced region have a notably high mutation frequency. The dinucleotide 5′-CpG-3′ is the main site where in DNA from higher

vertebrates the methylated C-residue 5-methylcytosine is found (see Bird 1987, for a recent review). In the region sequenced, 21 5′-CpG-3′ dinucleotides are present versus 102 5′-GpC-3′ dinucleotides. This is in accordance with the strong underrepresentation of the 5′-CpG-3′ dinucleotide in DNA of higher vertebrates (Bird 1987). Nevertheless, 5′-CpG-3′ dinucleotides are three times involved in the observed nucleotide changes, whereas 5′-GpC-3′ dinucleotides only twice. This observation is in favor of the model that explains the systematic underrepresentation of 5′-CpG-3′ by the high mutability of the methylated C-residue in this dinucleotide (Bird 1987).

Acknowledgments. We thank Drs. P. Terpstra and O. Bakker for help with the computer analysis, and F. Heering, B. Hennis, R. van Nues, and A. Wolthuis for experimental help. We also thank Prof. Dr. M. Gruber for critically reading the manuscript and Dr. G.J.B. van Ommen for stimulating discussions and helpful suggestions. This work was supported by grants from the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands organization for scientific research (NWO).

References

Arnberg AC, Van Ommen GJB, Grivell LA, Van Bruggen EFJ, Borst P (1980) Some yeast mitochondrial RNAs are circular. Cell 19:313–319

Arnberg AC, Meijlink FCPW, Mulder J, Van Bruggen EFJ, Gruber M, AB G (1981) Isolation and characterization of genomic clones covering the chicken vitellogenin gene region. Nucleic Acids Res 9:3271–3286

Bird AP (1987) CpG islands as gene markers in the vertebrate nucleus. Trends Genet 3:342–347

Covarrubias L, Bolivar F (1982) Construction and characterization of new cloning vehicles VI. Plasmid pBR329, a new derivative of pBR328 lacking the 482-base-pair inverted duplication. Gene 17:79–89

Frischauf AM, Lehrach H, Poustka A, Murray N (1983) Lambda replacement vectors carrying polylinker sequences. J Mol Biol 170:827–842

Lenstra JA, Van Vliet A, Arnberg AC, Van Hemert FJ, Möller W (1986) Genes coding for the elongation factor EF-1_α in *Artemia*. Eur J Biochem 155:475–483

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning. Cold Spring Harbor Press, Cold Spring Harbor NY, pp 164–165, 256–293, and 310–328

Nowell PC, Hungerford DA (1960) A minute chromosome in human chronic granulocytic leukemia. Science 132:1497

Philipsen JNJ, Gruber M, AB G (1985) Expression-linked demethylation of 5-methylcytosines in the chicken vitellogenin gene region. Biochim Biophys Acta 826:186–194

Reynaud C-A, Anquez V, Grimal H, Weill J-C (1987) A hyperconversion mechanism generates the chicken light chain preimmune repertoire. Cell 48:379–388

Sanger F, Nicklen S, Coulson AS (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467

Shen C-KJ, Maniatis T (1980) The organization of repetitive sequences in a cluster of rabbit β-like globin genes. Cell 19: 379–391

Staden R (1982) An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. Nucleic Acids Res 9:2951–2961

Stumph WE, Kristo P, Tsai M-J, O'Malley BW (1981) A chicken middle-repetitive DNA sequence which shares homology with mammalian ubiquitous repeats. Nucleic Acids Res 9: 5383–5397

Tonegawa S (1983) Somatic generation of antibody diversity. Nature 302:575–581

Valerio D, Dekker BMM, Duyvesteyn MGC, Van der Voorn L, Berkvens TM, Van Ormondt H, Van der Eb AJ (1986) One adenosine deaminase allele in a patient with severe combined immunodeficiency contains a point mutation abolishing enzyme activity. EMBO J 5:113–119

Van het Schip AD, Samallo J, Broos J, Ophuis J, Mojet M, Gruber M, AB G (1987a) Nucleotide sequence of a chicken vitellogenin gene and derived amino acid sequence of the encoded yolk precursor protein. J Mol Biol 196:245–260

Van het Schip AD, Samallo J, Meijlink FCPW, Gruber M, AB G (1987b) A new repetitive element of the CR1 family downstream of the chicken vitellogenin gene. Nucleic Acids Res 15:4193–4202

Wang SY, Smith DE, Williams DL (1983) Purification of avian vitellogenin III: comparison with vitellogenin I and II. Biochemistry 22:6206–6212

Zagursky RJ, Baumeister K, Lomax N, Berman ML (1985) Rapid and easy sequencing of large linear double-stranded DNA and supercoiled plasmid DNA. Gene Anal Techniques 2: 89–94

Received February 2, 1988/Revised April 25, 1988