Genomic clones spanning the entire cow $\beta$-globin gene locus have been isolated and characterized. These clones demonstrate that the linkage of embryonic-like ($e$) genes and pseudogenes ($w$) to the previously described fetal ($\gamma$) and adult ($\beta$) genes is as follows: $5'$-$e^1$-$e^4$-$w^3$-$\beta^3$-$e^2$-$w^4$-$\gamma$-$3'$. Present data indicate that, like that of the goat, the fetal and adult genes arose via block duplication of an ancestral four-gene set: $e$-$e$-$w$-$\beta$. This duplication event preceded the divergence of cows and goats, which occurred $\geqslant 18$-$20$ Myr ago. However, cows do not have the additional four-gene block containing a preadult/stress globin gene ($\beta^c$). Furthermore, the cow fetal cluster contains an extra $\beta$-like pseudogene, which apparently arose by a small-scale duplication. The fixation of this duplication may indicate a possible evolutionary role for pseudogenes.

**Introduction**

The ruminants are ideal for studying globin gene evolution and developmental regulation. Cows (*Bos taurus*), like humans, switch from embryonic to fetal to adult forms of hemoglobin. Goats (*Capra hircus*), which shared a common ancestor with cows as recently as $18$-$20$ Myr ago (Romero-Herrera et al. 1973; Gentry 1978; Reisner and Bucholtz 1983), have an additional switch from a fetal to an intermediate preadult globin before finally producing the adult form at 3 months after birth. This final switch is reversible when the animal is subjected to anemic stress, hypoxia, or erythropoietin (Huisman et al. 1969; Thurman et al. 1970; Tucker 1971). Sheep, which diverged from goats $\sim 5$-$7$ Myr ago (Novacek 1982; Li and Gojobori 1983), do not have this additional developmental switch but do produce a stress hemoglobin analogous to that in the goat (Benz et al. 1978; Nienhuis et al. 1979). We have chosen to study the bovine $\beta$-globin genes in order to investigate the molecular basis for these recently evolved differences in globin regulation. Here we describe the cloning, characterization, and linkage of the bovine $\beta$-globin genes. Comparison of these data with those from the goat system show that whereas goats have a triplicated four-gene set containing fetal, preadult, and adult clusters, cows have a duplicated set composed of a fetal and adult cluster. Furthermore, the bovine fetal cluster includes two related $\beta$-like pseudoglobin genes, which apparently arose by duplication of a primordial pseudogene. These features indicate a possible evolutionary role for pseudogenes.

**Material and Methods**

All enzymes except calf intestinal phosphatase (obtained from Boehringer Mannheim) were purchased from New England Biolabs. Nitrocellulose filters were obtained

1. Key words: gene duplication, pseudogenes, gene conversion, molecular evolution.

2. Current address: Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544.

Address for correspondence and reprints: Dr. Craig H. Duncan, Children's Hospital Research Foundation, IDR Building, Room 727, Elland and Bethesda Avenues, Cincinnati, Ohio 45229.

© 1985 by The University of Chicago. All rights reserved.
0737-4038/85/0206-0666$02.00

514
from Millipore. *Escherichia coli* strain 490A and the pTL5 cosmid vector with a tetracycline-resistance marker, which were described by Steinmetz et al. (1982), were provided by Dr. Peter Stambrook. The M13 vectors mp18 and mp19 were from J. Messing. The Holstein calf whose DNA was used for genomic libraries was purchased from a farmer in Hamilton, Ohio.

Construction of a Bovine Cosmid Genomic Library

High-molecular-weight DNA was prepared from the thymus of a 3-day-old male Holstein calf, essentially according to the procedure described by Blin and Stafford (1976). After partial *MboI* digestion, 35–50-kb DNA was size-fractionated on NaCl gradients (Grosveld et al. 1982), ligated to a 5–10-fold molar excess of phosphatase-treated pTL5 cosmid arms (see Steinmetz et al. [1982] for the cloning strategy involving this cosmid vector), and packaged in vitro using extracts from the *E. coli* strains BHB2688 and BHB2690 (Scalenghe et al. 1981). Approximately 150,000–200,000 packaged cosmids were obtained per microgram of cow DNA. The packaged cosmids were mixed with an equal volume of late log phase *E. coli* 490A for 15 min at 23 C, diluted in five volumes of L-Broth (10 g/liter Difco bacto-tryptone, 5 g/liter Difco yeast extract, 5 g/liter NaCl), and shaken for 45 min at 37 C. The cells were pelleted, resuspended in 7.5 ml L-Broth, and 500 kl (15,000 cosmids) was plated onto each of 15 moist, sterile nitrocellulose filters layered on 150-mm petri dishes containing L-broth plus 10 μg/ml tetracycline. The plates were incubated overnight at 37 C until colonies grew to a diameter of 0.1–0.2 mm; two replicas were made from each master, the cells on the replicas were lysed, and the DNA was fixed as described (Grosveld et al. 1981). After prehybridization overnight at 63 C in 5 X Denhardt's solution (Denhardt 1966), 5 X SSC (1 X SSC = 150 mM NaCl, 15 mM NaCitrate, pH = 7.7), and 50 mM Na phosphate buffer, pH = 7.0, the DNAs on the filters were hybridized in 100 ml of an identical solution plus 200 pg of HaeIII-cut pBR329 (Covarrubias and Bolivar 1982). Primary screening was performed using 4 μg of nick-translated, purified goat fetal gene probe, Gγ5, kindly supplied by P. Liberator and J. Lingrel. This probe contains the first and second exons of the goat fetal (βγ) gene. Later, after we had constructed probes from cloned cow sequences, a second screening was performed using a mixture of Gγ5 and probes from the bovine e2 and ψ3 genes. Filters were washed five times for 15 min at room temperature in 2 X SSC and 0.1% SDS, then twice for 2 h in 0.1 X SSC and 0.1% sodium dodecyl sulfate in a 50 C shaking water bath. The filters were then exposed to Kodak XAR-2 film for 12 h at −70 C against Dupont Cronex Lightning Plus intensifying screens.

Construction of a Bacteriophage Genomic Library

Thymus DNA from the same calf as used in the procedure described above was partially digested with *EcoRI*, and fragments in the 15–20-kb range were purified on a preparative, low-melting-temperature agarose gel. This DNA was mixed in a 2:1 molar ratio with purified Charon 4 phage arms, ligated, and packaged as above. Without amplification, a total of 1,200,000 recombinant phage were plated out onto 155-mm petri dishes at a density of 25,000 pfu per plate. Globin-containing phages were detected using the same hybridization conditions and probes as described above. A total of 37 positive clones were obtained.

Southern Blotting

Thymus DNA from the same calf used to construct the genomic libraries was digested completely with *EcoRI*, run out on a 0.8% agarose gel, and transferred to
nitrocellulose paper as described by Southern (1975). The filters were hybridized and washed according to the method of Thomas (1980).

**DNA Preparation from Clones**

Cosmid containing bacteria were grown either in liquid culture (L-broth plus 10 µg/ml tetracycline) or, if the particular clone showed instability during growth in liquid culture, on solid medium in large baking dishes. DNA was isolated by alkaline extraction (Birnboim and Doly 1979).

Phage DNA was isolated by a miniprep method described by Leder et al. (1977), except that the DNA was further treated with 30 µg/ml RNase for 1 h at 37 C, and chromatographed on a 1-ml Sephadex G-25 centrifugal column.

**DNA Sequencing**

DNA to be sequenced was subcloned into the M13 vectors mp18 or mp19 and propagated on the host strain JM105 (Norandor et al. 1983). Sequences were determined by the dideoxy-chain termination method (Sanger et al. 1977).

**Results**

Because the structures of members of other ruminant β-globin gene families are large (Haynes et al. 1980; Kretschmer et al. 1981; Lingrel et al. 1981; Schon et al. 1981), we employed both cosmid- and bacteriophage-cloning strategies to isolate the bovine globin genes. Initial and secondary screenings of our unamplified cosmid and phage libraries resulted in the isolation of a set of overlapping clones (fig. 1). These

---

**Fig. 1.**—Linkage map of the bovine β-globin locus. Charon 4 phage clones are numbered; cosmid clones are preceded by cos. EcoRI sites are indicated as small vertical marks below the linkage map. The curled ends of the cosmid clones signify that the precise end points of the insert DNA are not known. Underlined clones represent the allele of the chromosome indicated in the map. The region defined by a horizontal bar below the linkage map contains the polymorphic EcoRI sites in that allele.
clones span ~90 kb of DNA, and genomic Southern blotting data (see fig. 2) indicate that all the β-like globin genes are contained in these clones.

Two cosmid clones (not shown) were highly unstable, thus prohibiting a reliable characterization. Both apparently mapped to the region encompassing ε3 and β. Clones cos1c and cos8c, which extend approximately halfway into this region, were also somewhat unstable when grown in liquid culture, but rearrangement was minimized by growing them on solid medium. None of the clones between ε1 and γ was unstable. We believe that the ε3-β region contains sequences that are either highly recombinogenic (such as repetitive elements or the genes themselves) or deleterious to *E. coli*. The Charon 4 clones did not show signs of instability.

Four *EcoRI* restriction fragment-length polymorphisms were detected between the sister chromosomes. One globin-hybridizing sequence, designated ε2, was contained in two different-sized *EcoRI* fragments (see bands ε2 and ε2a in fig. 2). The larger version, ε2a, was contained in clones cos1c, J11, and J8. DNA sequence analysis (data not shown) indicated a single base change obliterating an *EcoRI* site as the basis for this polymorphism. An allelic version of ψ2 (band ψ2a in fig. 2), linked downstream of ε2a, was also contained in one cosmid, cos10, and one phage clone, J32. Curiously, all four *EcoRI* restriction fragment-length polymorphisms between the sister chromosomes map to the ε2-ψ2 region, as indicated in figure 1.

Linkage of the Bovine Fetal and Adult β-Globin Genes

We have previously reported the complete nucleotide sequences of the bovine fetal (γ) and adult (β) globin genes (Schimenti and Duncan 1984). Figure 1 demonstrates the linkage order of these genes. Their transcription orientation is toward the right.

![Southern blot hybridization of EcoRI-calf thymus DNA to nick-translated β-globin probes. Blots were hybridized to either a purified fetal gene probe (Gγ5; see Material and Methods) or a purified probe containing the first intron, second exon, and approximately half of the second intron of the embryonic-like bovine ε2 gene. Marker lanes are indicated by M. Note that the alleles ψ2 and ψ2a light up with approximately half the intensity of the other bands. The ε1 fragment, which is ~1.1 kb in length, was run off the gel in order to obtain better separation of higher-molecular-weight fragments. The γ and β genes cross-hybridized with the ε probe.](image-url)
side of the cluster. γ is located 3' to β, and four other globin sequences lie between them. Unlike the situation in most other known β-globin gene families, cows, like goats, do not have globin genes situated on the chromosome in the order of developmental expression (Neinhuis and Stamatoyannopoulos 1981; Townes et al. 1984b).

The bovine γ and β genes are very similar (~90% identical) but are more similar to their orthologous goat genes (Schimenti and Duncan 1984). Townes et al. (1984b) have shown that the goat fetal (βF), and adult (βA) genes arose by block duplication of a four-gene set, ε-ε-ψ-β, to produce separate fetal and adult clusters. Comparison of both the fetal and adult genes to other mammalian globin genes indicated that they were β- or adult-like in nature, which prompted the renaming of the goat fetal gene from γ to βF (Schon et al. 1981; Townes et al. 1984a). As shown in figure 3, the fetal and adult genes of the cow and goat occupy comparable positions in the globin gene clusters.

Identification of Embryonic-Like Globin Genes

The gene designated ε2 hybridized very weakly to a fetal gene probe on blots of purified clone DNA, and not at all in a genomic Southern blot. A probe prepared from this gene detected five new EcoRI bands on genomic blots (fig. 2), each of which corresponds to a single gene (fig. 1). We have determined the entire nucleotide sequences of ε2, ε4 (see Schimenti and Duncan 1985, in this issue), and parts of ε1 and ε3 (data not shown). These pairs were more than 90% identical to the reported nucleotide sequences of the goat εII/εIV/εVI and εI/εIII/εV genes, respectively (Shapiro et al. 1983; Shapiro and Lingrel 1984; Townes et al. 1984a, 1984b). Based on these sequence similarities and the corresponding positions of the bovine genes and the goat genes (see fig. 3), we believe that the bovine ε3, ε4, ε1, and ε2 genes are orthologous to the goat εI/εIII, εII/εIV, εV, and εVI genes, respectively.

Although no amino acid sequences of any bovine embryonic globins are known, probably at least one of the four ε genes is expressed as an embryonic globin. As Shapiro et al. (1983) noted in the goat system and as our data indicate (see Schimenti and Duncan 1985, in this issue), these genes show strong similarities to other mammalian embryonic globins, such as the human ε and rabbit β4 (Baralle et al. 1980; Hardison 1983). Furthermore, it is known that cows express embryonic-specific globin; three different variants have been detected by polyacrylamide gel electrophoresis (Kleihauer and Stoffler 1968). Since we have cloned all of the β-like genes, at least one of these must be expressed as embryonic globin.

Identification of Three β-Like Pseudogenes

Four regions were identified, ψ1, ψ2 and its allele ψ2a, and ψ3, that hybridized to a β-like fetal gene probe but not to a probe prepared from the ε2 gene (fig. 2). Preliminary nucleotide sequence analysis (the complete sequences of which are to be published)
later) indicates that these globin-like sequences contain multiple defects that render them pseudogenes. In the 5′ region alone (fig. 4), each of the cow pseudogenes has altered CCAAT and ATA boxes, and \( \psi^3 \) has a single base insertion between its eleventh and twelfth "codons," which would cause a frameshift.

These genes are more similar to each other and to the goat pseudogenes, \( \psi^{\beta} \) and \( \psi^{\beta x} \) (Cleary et al. 1981), than to any of the other bovine genes. Figure 4 shows an alignment of nucleotide sequences from the 5′ regions of the bovine \( \psi^1, \psi^2, \psi^3 \), and \( \gamma \); human \( \delta \) and \( \beta \); and goat \( \psi^{\beta x} \) genes. All the ruminant pseudogenes contain (1) a 9-
bp deletion relative to cow γ and human δ and β (as well as to cow β and goat βF, βA, and βC genes not shown in fig. 4) at position +19 and (2) an extra two “codons,” or 6 bp between codons 1 and 2 of the cow and goat fetal and adult genes. Apparently, these two codons were deleted in the ruminant fetal/adult gene ancestor, since other related β-like adult genes (human, mouse, and rabbit) have these two codons (Konkel et al. 1979; Lacy et al. 1979; Lawn et al. 1980). The goat ψsx and ψbx pseudogenes, which are located in the preadult and adult clusters, share an identical frameshift mutation with ψ3, which is situated in the cow adult cluster. ψ1, ψ2, and goat ψβY, which reside in the fetal gene clusters, do not have this single base insertion (Townes et al. 1984b). The relative sequence similarity, corresponding locations within the globin loci, and shared structural features indicate that all of the ruminant pseudogenes had common precursors before the cow/goat divergence.

Two related pseudogenes are adjacent in the cow fetal cluster, whereas neither the cow adult cluster nor any of the goat clusters contain such an arrangement. It is probable that a duplication event after the cow/goat divergence created a second gene. An analysis of the complete nucleotide sequences of ψ1 and ψ2 and their surrounding regions should allow us to determine the extent and nature of the duplication.

Discussion
Molecular Evolution of a Globin Gene Family

Present-day ruminant globin genes arose by a block duplication of an ancestral four-gene set. Genes in corresponding positions in each set are very similar, consistent with their recent creation and divergence (Schon et al. 1981). Alignment of the cow and goat globin gene families (fig. 3) shows that the entire bovine globin locus is homomorphic to the goat adult and fetal clusters and that the preadult cluster is absent from the cow.

Figure 5 illustrates a proposed history of the ruminant globin loci. Several lines of evidence support this. We have previously shown that the bovine fetal (γ) and adult (β) β-globin genes are orthologous to the fetal (βF) and adult/preadult (βA/βC) genes of the goat (Schimenti and Duncan 1984). It appears that the duplication event that created the caprine preadult cluster occurred after the cow/goat divergence. This is consistent with the estimated times of divergence between cows and goats (18–20 Myr ago) (Romero-Herrera et al. 1973; Gentry 1978; Reisner and Bucholtz 1983) and

![Diagram of proposed ruminant globin evolution](image)

**Fig. 5.—Schematic diagram of proposed ruminant globin evolution.** The goat lineage is taken from Townes et al. (1984b). The estimated time of cow/goat divergence is derived from multiple sources (see text). The rightmost gene of the ancient ruminant, designated β, was presumably a fetal gene. MYA = million years ago.
between the goat \( \beta^C \) and \( \beta^A \) genes (<15 Myr ago) (J. Lingrel, personal communication; Li and Gojobori 1983). However, the possibility that cows deleted the extra cluster cannot be eliminated. Nevertheless, the absence of a preadult or stress globin in cattle results from the absence of the structural gene, not from its silencing.

The cow and goat adult clusters are quite similar. A comparison of nucleotide sequence data indicates that the bovine \( \varepsilon^3, \varepsilon^4, \psi^3, \) and \( \beta \) genes are highly homologous, structurally similar, and probably directly related to the goat \( \varepsilon^3, \varepsilon^4, \psi^{Bz}, \) and \( \beta^A \) genes. The distances between the genes have been well conserved, except that the goat \( \psi^{Bz} \) and \( \beta^A \) genes are separated by ~3.8 kb more than are the corresponding cow genes. This difference probably arises from an insertion in the goat cluster after the cow/goat divergence, since the corresponding intergenic distance in the related goat preadult cluster is identical to that in the cow. Hardies et al. (1984) reached the same conclusion by comparing orthologous gene spacing in other mammals. The distance between the fetal and adult clusters (that separating \( \beta \) and \( \varepsilon^1 \)) is ~3.5 kb greater in the cow. At present we cannot say whether this arose because of insertion in the cow or deletion in the goat. The analogous genes in the fetal clusters also appear to be directly related; however, a marked difference is that compared to the goat the cow cluster has an extra pseudogene (fig. 3).

An Evolutionary Role for Pseudogenes?

The data indicate that \( \psi^1 \) and \( \psi^2 \) were derived from a duplication of a single ancestral gene in the fetal cluster. As described in the Results section, both are \( \beta \)-like in nature and share features characteristic of the other ruminant pseudogenes. They do not contain the frameshift mutation contained in \( \psi^3 \) and goat \( \psi^{Bz} \). This proposed duplication must have occurred after the primordial four-gene set duplicated to form separate fetal and adult clusters, since neither the cow nor the goat adult clusters contain this extra gene (fig. 5). It is also most likely that this duplication happened after the cow/goat divergence, because the goat fetal gene cluster does not have this arrangement; but since our preliminary sequence data indicate that these genes are 15%–20% divergent, the event has probably been conserved for millions of years.

Even if the \( \psi^1/\psi^2 \) ancestor was a pseudogene at the time of its duplication, a possibility that is consistent with data on goat globin lineages derived by Townes et al. (1984b), there still may have been an evolutionary reason for fixation of this duplication. On the basis of studies of the human histocompatibility genes, Larhammar et al. (1985) have proposed that pseudogenes may act as mutation “reservoirs” during evolution. Pseudogenes may, via microgene conversion, confer multiple changes on regions of an evolving functional gene. It is conceivable that the extra pseudogene intrachromosomally converted a portion of the early duplicated fetal or adult genes, resulting in amino acid changes more suitable to either role. It would then be maintained through close linkage to the new, selected-for functional gene.

Evolutionary Relationships of the Cow Pseudogenes

Through a comparison of mammalian globin DNA sequences, Hardies et al. (1984) determined an ancestry of members in the mammalian \( \beta \)-globin gene family. They concluded that the goat pseudogenes, \( \psi^{Bx} \) and \( \psi^{Bz} \), were “delta-like” in origin, being orthologous to the human \( \delta \), rabbit \( \psi^{Bz} \), and mouse \( \beta h 2 \), and that the goat fetal and adult genes were “beta-like,” being most closely related to the adult genes of other mammals. They also determined that the coding regions of what are now \( \psi^{Bx} \) and \( \psi^{Bz} \) had undergone gene conversion with the “proto-\( \beta \)” gene in the ancient ruminant.
These conclusions also apply to the orthologous bovine genes. In the region upstream of the CCAAT box, the ruminant pseudogenes show no similarity to the β-like genes, human δ and cow β (fig. 4). They are more than 70% identical to human δ in this region. However, downstream of CCAAT, the similarity of both δ and the ruminant pseudogenes to the β-like genes increases dramatically. These data support Hardies et al.’s (1984) conclusion that gene conversion prior to the mammalian radiation occurred between the primordial δ and β genes up to a boundary near the CCAAT box.

Evolutionary Strategy of the Ruminant Globin Genes

The most intriguing aspect of ruminant β-globin gene evolution is the manner in which the various developmentally regulated genes were created. The data we present here, in agreement with those from the goat system (Townes et al. 1984b), show that an immediate ancestor to present-day ruminants had a basic four-gene set: ε-ε-ψ-β. As noted by Shapiro et al. (1983), the first gene in the ruminant cluster most closely resembles other embryonic genes, human ε and rabbit β4. The second gene in the cluster shows approximately as much similarity to other mammalian embryonic genes as to the human fetal genes (Shapiro et al. 1983), although its evolutionary relationship to these genes has been proposed to be indirect (Goodman et al. 1984). The ruminant third and fourth genes in the set are apparently orthologous to the descendants of what are now pseudo- and adult genes (Hardies et al. 1984). Thus, the early ruminants, having only a four-gene set, may still have expressed three distinct globins in development: ε(embryonic)-ε(fetal)-pseudo(none or adult δ-like)-β(adult). Another possibility is that the ancient ruminant, like the rabbit, did not have an embryonic-to-fetal switch and that both ε genes were embryonic specific. In either case, what was the driving force that caused ruminants to undergo a block duplication event forming separate fetal and adult clusters? Whether the evolutionary advantage, if any, for block duplication was primarily recruitment of different embryonic genes or creation of a new fetal gene or both, a single duplication of the primordial β(adult) gene alone could have produced the same result: ε(embryonic)-ε(embryonic)-ψ(pseudo)-β(fetal)-β(adult).

Perhaps, more plausibly, the block duplication moved the newly created fetal and adult genes into chromosomal domains that are “opened” during appropriate developmental periods (Schon et al. 1981; Townes et al. 1984b). Alternatively, the evolutionary drive may have been toward creation of a large bank of embryonic genes, since in both cow and goat these are the only redundant genes. It is unknown whether all the presumptive embryonic genes are expressed and, if they are expressed, whether multiple embryonic globins would benefit the developing embryo.

Ideal Model for Studying Developmental Regulation and Evolution

Ruminant globin gene loci provide distinct advantages over other mammalian globin systems in studies of the molecular bases of hemoglobin switching and gene-family evolution. Fetal and adult genes arose via relatively recent duplication and divergence. Thus, it is possible that the structural changes responsible for the differential developmental specificities of these homologous genes can be revealed by pairwise DNA sequence analysis. We are currently in the process of sequencing DNA flanking the bovine fetal and adult genes. Comparison of these sequences, as well as those available from the goat, may shed light on the nature of DNA sequences that may be important in developmental regulation. Furthermore, complete nucleotide sequence analysis of all the members of the bovine globin family will allow us to make conclusions on the role of gene conversion in the dynamics of gene-family evolution.
Acknowledgments

The excellent technical assistance of J. Eichhold and the secretarial skills of M. Loescher were greatly appreciated. We also thank R. Ganschow and J. Lingrel for advice and discussion. This work was supported by the following grants to C.H.D.: American Cancer Society grant no. NP-409, March of Dimes Birth Defects Foundation Basil O'Connor Starter Research Grant no. 5-354, and National Institutes of Health grant no. HL15996-11 to the Comprehensive Sickle Cell Center of Cincinnati.

LITERATURE CITED


MASATOSHI NEI, reviewing editor

Received April 23, 1985; revision received June 18, 1985.