Nucleotide Sequence Evolution at the κ-Casein Locus: Evidence for Positive Selection Within the Family Bovidae

Todd J. Ward, Rodney L. Honeycutt* and James N. Derr*

*Department of Veterinary Pathobiology, *Faculty of Genetics, Texas A & M University, College Station, Texas 77843-4467*

Manuscript received July 2, 1997
Accepted for publication September 8, 1997

ABSTRACT

κ-Casein is a mammalian milk protein involved in a number of important physiological processes. In the gut, the ingested protein is split into an insoluble peptide (para κ-casein) and a soluble hydrophilic glycopeptide (caseinomacropeptide). Caseinomacropeptide is responsible for increased efficiency of digestion, prevention of neonate hypersensitivity to ingested proteins, and inhibition of gastric pathogens. Variation within this peptide has significant effects associated with important traits such as milk production. The nucleotide sequences for regions of κ-casein exon and intron four were determined for representatives of the artiodactyl family Bovidae. The pattern of nucleotide substitution in κ-casein sequences for distantly related bovid taxa demonstrates that positive selection has accelerated their divergence at the amino acid sequence level. This selection has differentially influenced the molecular evolution of the two κ-casein split peptides and is focused within a 34-codon region of caseinomacropeptide.

κ-CASEIN is a protein in mammalian milk that determines the size and specific function of milk micelles (GUTIERREZ *et al.* 1996). These micelles increase the solubility of minerals and facilitate the transfer of nutrients from mother to offspring (DEV *et al.* 1994). The mature κ-casein protein has a labile peptide bond that is cleaved in the gut by the action of rennin to produce an insoluble peptide (para κ-casein or PKC) as well as a soluble hydrophilic glycopeptide (caseinomacropeptide or CMP) (QIAN *et al.* 1995). The function of PKC is not well known. However, CMP is responsible for clotting milk in the gut, which increases retention time and results in more efficient digestion (MERCIER *et al.* 1976). Caseinomacropeptide also reduces the immune response of neonates, preventing hypersensitivity reactions to ingested food proteins (OTANI and MONNAI 1993). In addition, species-specific glycosylation patterns of CMP can result in differential inhibition of gastric pathogens such as *Helicobacter pylori* (STRÖMQVIST *et al.* 1995). Interestingly, the different alleles described in domestic cattle for the CMP peptide of κ-casein induce significantly different amounts of glycosylation (LODES *et al.* 1996) and have been shown to produce significant differences in milk yield and protein percent (MARZALI and NG-KWAI-HANG 1986).

At the nucleotide sequence level, most mammalian protein encoding genes demonstrate a ratio of synonymous to nonsynonymous substitutions per site of ~5:1 (LI 1997). Exon four of κ-casein appears to be an exception to this rule. A recent phylogenetic study of several mammalian taxa revealed similar rates of synonymous and nonsynonymous substitutions per site within exon four of κ-casein, with the divergence at each codon position being roughly equivalent (GATESY *et al.* 1996). Although this pattern of divergence at κ-casein was interpreted by GATESY *et al.* (1996) as supporting a strictly neutral model of evolution for this gene, these authors did not compare the level of variation within exon four to that within intron four, and performed a limited analysis of their data. Because κ-casein plays a critical role in several important physiological processes, it seems unlikely that this gene would be completely free of selective constraint. Therefore, we examined the pattern of nucleotide substitution within the κ-casein gene of representative bovid taxa to determine if the molecular evolution of this gene is influenced by processes of selection.

Corresponding author: James N. Derr, Department of Veterinary Pathobiology, Texas A & M University, College Station, TX 77843-4467. E-mail: jderr@cvm.tamu.edu

MATERIALS AND METHODS

The nucleotide sequence of a 782-base pair (bp) segment of the κ-casein gene was determined for *Bison bison* ($n = 1$), *B. bonasus* ($n = 2$), *Bos javanicus* ($n = 3$), *B. grunniens* ($n = 3$), *B. gaurus* ($n = 1$), *Syncerus caffer caffer* ($n = 1$), *S. caffer nanus* ($n = 1$), *Tragelaphus imberbis* ($n = 1$), *Boselaphus tragocamelus* ($n = 1$), and *Capra hircus* ($n = 1$). Numbers in parenthesis refer to the number of individuals sequenced. These nucleotide sequences encompass 377-bp from exon four and 407-bp from intron four of the κ-casein gene. In addition, the corresponding *B. taurus* sequence was obtained from ALEXANDER *et al.* (1988).

Total genomic DNA was isolated from white blood cells by proteinase K treatment followed by phenol chloroform extraction (SAMBROOK *et al.* 1989). Previously published oligo-

nucleotide primers (PINDER *et al.* 1991) were used to amplify DNA via PCR (SAIKI *et al.* 1988). Amplifications consisted of an initial 5-min denaturation at 95° followed by 35 cycles of 45 sec at 95°, 45 sec at 50°, and 2.5 min at 74°. Amplification products were run on 1.5% agarose gels and purified with QIAquick gel extraction columns (Qiagen). Nucleotide sequences were determined on an ABI PRISM 377 automated DNA sequencer (Applied Biosystems, Perkin Elmer) using the amplification primers as well as internal sequencing primers designed with MacVector 5.0 (International Biotechnologies Inc.). Primer sequences will be made available upon request. Sequences were aligned using Clustal V (HIGGINS and SHARP 1989) and then manually inspected.

Selection can be difficult to detect because its impact may be highly localized within a gene. Therefore, para κ-casein (PKC) and caseinomacropeptide (CMP) were analyzed separately because they are split into functionally different peptides in the gut and are potentially exposed to different selective forces. However, no detailed information regarding regions of structural or functional importance are available for these peptides. In the absence of such biological information, exon four sequences were analyzed with the computer program PLATO (GRASSLY and HOLMES 1997) in order to identify regions of possible selective importance. This program employed a statistical method that used site likelihoods to identify regions of nucleotide sequence that did not fit with the global maximum likelihood topology and substitution process. Regions of exon four identified as having a pattern of molecular evolution inconsistent with the overall pattern of the exon were examined separately in all subsequent analyses in order to determine if these regions were influenced by processes of selection. It should be noted that this analysis required a gene tree based on exon four sequences and did not require that this topology reflect the true phylogenetic relationships of the taxa examined.

The levels of total nucleotide sequence divergence within each gene region were determined for closely (Bos and Bison genera) and distantly related taxa separately. Distantly related taxa refer to pairs of taxa that diverged from each other at least 10 mya, while the closely related Bos and Bison species have divergence times of no more than 3 mya (MCDONALD 1981; JANECEK *et al.* 1996). *B. javanicus* was chosen to represent the Bos and Bison species in comparisons of distantly related taxa in order to simplify the analysis and to avoid biasing the results with repeated consideration of what would be nearly identical comparisons.

For distantly related taxa the number of nucleotide substitutions along each branch of three possible species trees (Figure 1) was determined for each gene region using PAUP 3.1.1 (SWOFFORD 1993). The program Fisher 6 (R. ADKINS, personal communication) was used to perform Fisher's exact tests to determine if the ratio of total nucleotide substitutions to total invariant sites was significantly different for pairwise comparisons of the different gene regions examined. The number of invariant sites was determined for each branch of the tree and then added across the entire tree in order to account for multiple substitutions at a single nucleotide position.

To further examine the pattern of nucleotide divergence within the κ-casein gene, MacClade 3.06 (MADDISON and MADDISON 1996) was used to determine the number of nucleotide substitutions at each codon position for each gene region individually using the three trees shown in Figure 1. Fisher 6 was then used to determine if the ratio of total nucleotide substitutions to total invariant sites was significantly different at first and second codon positions compared to third codon positions. The number of invariant sites was determined as the sum of the total nucleotides along each branch minus the number of total changes on the tree.

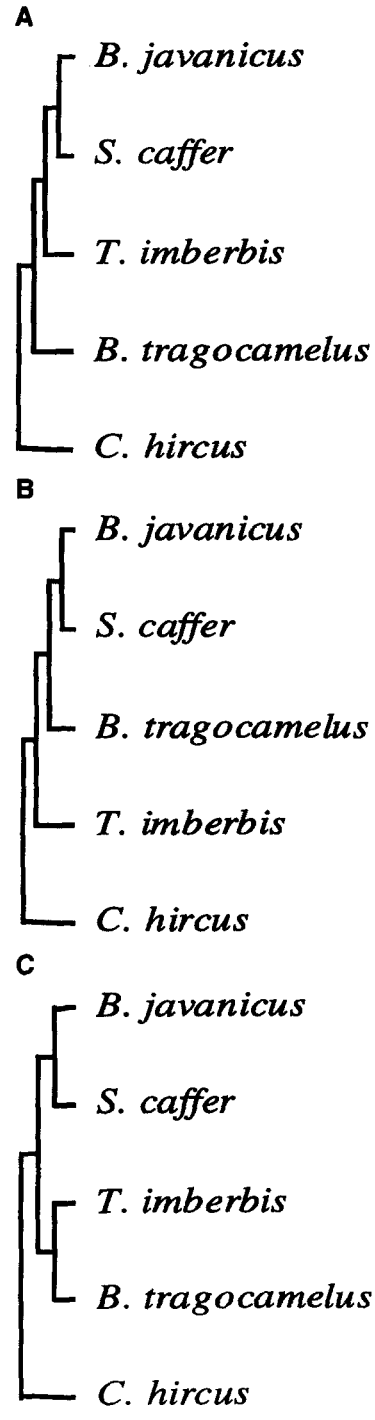

FIGURE 1.—Three possible species trees for distantly related bovid taxa showing (A) the tribes Bovini and Tragelaphini as sister taxa, (B) the tribes Bovini and Boselaphini as sister taxa, and (C) the tribes Boselaphini and Tragelaphini as sister taxa.

Three different species trees were used in the above analyses because there is no consensus regarding the phylogenetic relationships of the three bovine tribes (Bovini, Tragelaphini, and Boselaphini) to one another (ALLARD *et al.* 1992; JANECEK *et al.* 1996). However, there has been a general consensus

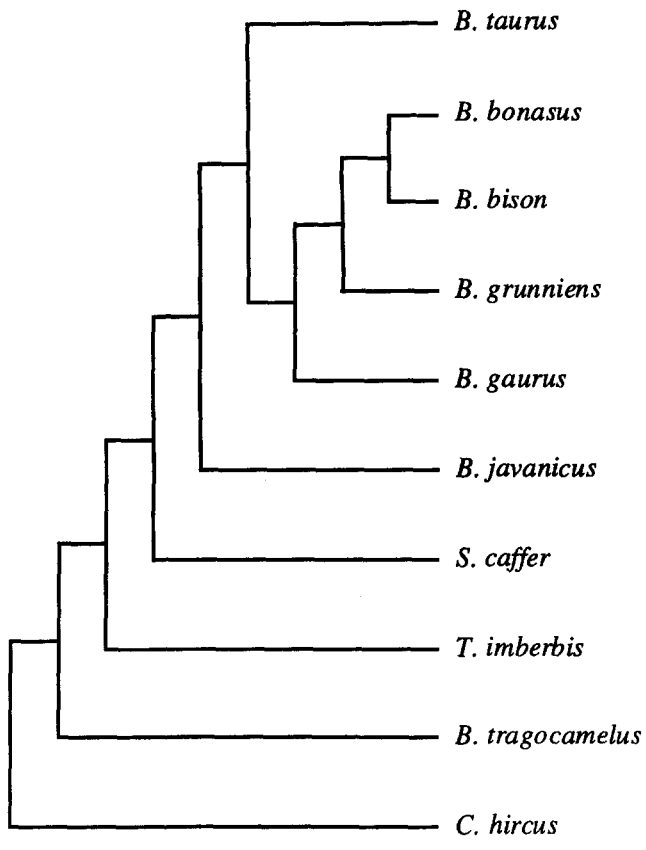

FIGURE 2.—Gene tree resulting from maximum likelihood analysis of κ-casein exon four sequences.

that the Bos and Bison species form a single monophyletic group and that *S. caffer* (as representative of the true buffaloes) is a sister taxon to this group (BOHLKEN 1958, 1961; GROVES 1981; WALL *et al.* 1992; JANECEK *et al.* 1996; MODI 1996). In addition, the placement of *C. hircus* (as representative of the Caprinae) as a sister taxon to the other species examined in this study is well supported (PILGRIM 1947; JANECEK *et al.* 1996). Therefore, the three trees used in the above analyses represent all of the probable species trees for the distantly related taxa examined.

Due to the fact that there is no generally agreed upon topology for the closely related Bos and Bison species (BOHLKEN 1958, 1961; GROVES 1981; WALL *et al.* 1992; JANECEK *et al.* 1996; MODI 1996), it was not possible to compare the levels of nucleotide divergence using the topology-based approach described for distantly related taxa. Therefore, the computer program MEGA (KUMAR *et al.* 1993) was used to calculate the average proportion of nucleotide differences for pairwise comparisons of sequences from the various gene regions of closely related taxa. While this method did provide an indication of the relative levels of divergence within the different gene regions, it was not possible to perform a detailed statistical analysis on these data because of problems associated with the nonindependence of pairwise comparisons of this type.

The Jukes-Cantor-corrected proportions of synonymous (Ds) and nonsynonymous differences (Dn) per site were calculated for closely and distantly related taxa separately using MEGA. Estimates of Ds and Dn were calculated by the method of NEI and GOJOBORI (1986), and their associated variances were calculated by the method of NEI and JIN (1989). The significance of differences in Ds and Dn were assessed with

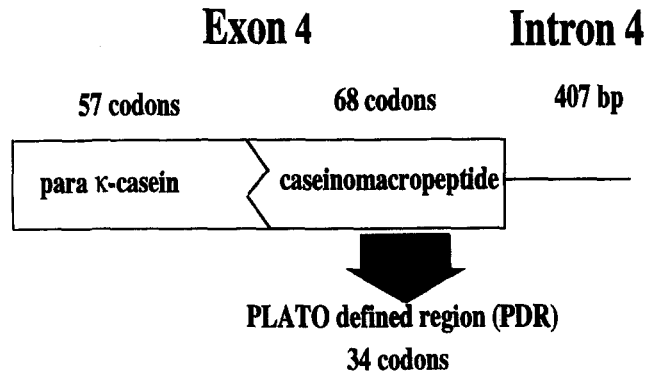

FIGURE 3.—Diagram showing the relative position of the various κ-casein gene regions examined.

one-tailed *t*-tests and infinite degrees of freedom (KUMAR *et al.* 1993).

RESULTS

Individuals from the two subspecies of *S. caffer* differed by a single transition in intron four. To simplify analysis, only *S. c. caffer* sequence was used. The three *B. grunniens* also differed by a single nucleotide substitution in intron four. Therefore, a single sequence representing the species was selected at random. In all other species for which multiple individuals were sequenced, the sequences were identical.

The nucleotide sequences from exon four as well as a single gene tree (Figure 2) resulting from a maximum likelihood analysis of these sequences using the DNAML program of the PHYLIP software package (FELSENSTEIN 1995) were used in the analysis by the computer program PLATO (GRASSLY and HOLMES 1997). Using this program, we identified a 34-codon region of anomalously evolving sequence located within the CMP portion of the κ-casein gene ($P < 0.01$). The location of this PLATO-defined region (PDR) as well as the relative position of the PKC and CMP peptides are shown in Figure 3.

Pairwise comparisons of sequences from closely related species within the Bos and Bison clade revealed levels of nucleotide sequence divergence within intron four that were equal to or greater than that within any region of exon four (Figure 4). In comparisons of distantly related taxa there was no significant difference between the level of total nucleotide divergence within the PKC region of the exon and that found within intron four. However, significantly higher levels of total nucleotide sequence divergence were found within the CMP, PDR, and exon four as a whole relative to levels within intron four in these comparisons of distantly related bovid taxa (Table 1). In addition, all branch lengths derived using the DELTRAN character state optimization option in PAUP 3.1.1 revealed a high amount of nucleotide substitution within these three

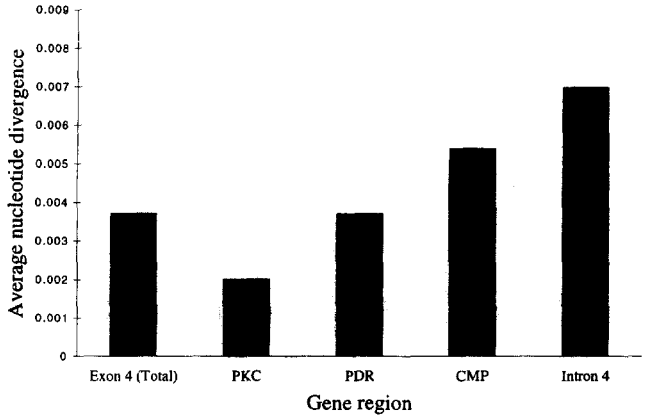

FIGURE 4.—Average values of pairwise divergences for comparisons of the closely related Bos and Bison species.

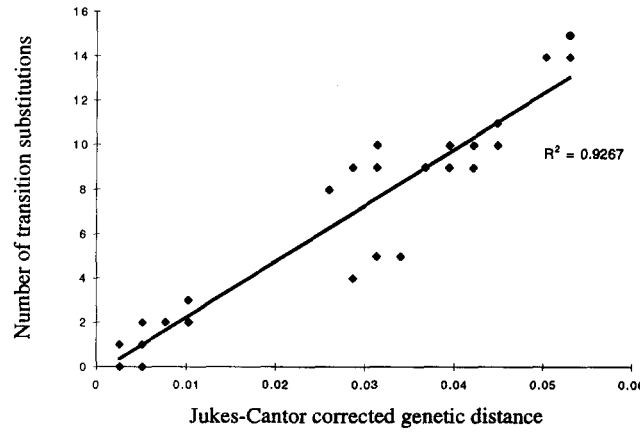

FIGURE 5.—Transitional substitution within the intron four sequences of all taxa.

coding regions relative to that within intron four for all three possible trees shown in Figure 1. These same results were obtained when using the ACCTRAN character state optimization option, with the exception of one branch within tree 1B where the level of nucleotide divergence within the PDR region was equal to that of intron four and one branch within tree 1C where the level of divergence within intron four was greater than that of exon four as a whole. As shown in Figure 5, saturation effects cannot account for the intron's low level of divergence relative to the PDR and CMP regions of exon four. Therefore, the fact that the levels of nucleotide substitution within exon four as a whole, specifically within the PDR and CMP regions of the exon, exceeded that of intron four suggests that positive Darwinian selection has accelerated the divergence of these coding sequences in distantly related bovid taxa.

Estimates of sequence divergence at each of the three codon positions were calculated for each region of the exon using each of the three probable species trees of distantly related taxa. The number of substitutions at both first and second codon positions was greater than at third positions within all regions of the exon for all three trees (Table 2). However, the difference between divergence at first and third positions was found to be statistically significant ($P < 0.05$) only in the comparisons for exon four as a whole, while the difference between levels of divergence at second and third positions was found to be statistically significant ($P < 0.05$) for the PDR, CMP, and exon four as a whole using trees 1B and 1C. For species tree 1A the difference between levels of divergence at second and third positions was found to be statistically significant ($P < 0.05$) within the PDR and exon four as a whole and was just above the 5% level ($P = 0.06$) within the CMP region of the exon. Because all changes at second codon positions and most changes at first codon positions result in amino acid replacements, these results indicate that the accelerated rate of nucleotide substitution was caused by selection for divergence at the amino acid sequence level.

While the pattern of nucleotide substitution described above was indicative of positive selection, the standard test for deviation from neutrality has been to compare the rates of synonymous (Ds) and nonsynonymous (Dn) substitution per site (HUGHES 1992; LEE *et*

TABLE 1

Total nucleotide sequence divergence for distantly related taxa using species tree 1B

Gene regions compared		Ratio of nucleotide substitutions/total number of invariant sites		Significance of difference
Region 1	Region 2	Region 1	Region 2	
Exon 4 (total)	Intron 4	0.0235	0.0130	$P < 0.005$
CMP	Intron 4	0.0337	0.0130	$P < 0.001$
PDR	Intron 4	0.0464	0.0130	$P < 0.001$
Intron 4	PKC	0.0130	0.0125	NS
CMP	PKC	0.0337	0.0125	$P < 0.001$
PDR	PKC	0.0448	0.0125	$P < 0.001$

The results for species tree 1C were identical to those shown, and results for species tree 1A were not significantly different. NS, not significant.

TABLE 2

Nucleotide sequence divergence at each of the three codon positions for distantly related taxa

Gene region examined	Ratio of nucleotide substitutions/ total number of invariant sites			Significance of difference	
	Codon pos. 1	Codon pos. 2	Codon pos. 3	Pos. 1–3	Pos. 2–3
Tree 1A					
Exon 4 (total)	0.0092	0.0092	0.0044	$P < 0.05$	$P < 0.05$
PKC	0.0058	0.0041	0.0017	NS	NS
PDR	0.0161	0.0211	0.0064	NS	$P < 0.05$
CMP	0.0123	0.0138	0.0069	NS	NS
Tree 1B					
Exon 4 (total)	0.0088	0.0100	0.0044	$P < 0.05$	$P < 0.05$
PKC	0.0058	0.0050	0.0017	NS	NS
PDR	0.0161	0.0227	0.0064	NS	$P < 0.05$
CMP	0.0115	0.0146	0.0069	NS	$P < 0.05$

The results for species tree 1C were identical to those shown for species tree 1B. NS, not significant; pos., position.

al. 1995; SWANSON and VAQUIER 1995; METZ and PALUMBI 1996). In general, the rate of substitution at synonymous sites is many times greater than at nonsynonymous sites (KIMURA 1977; JUKES and KING 1979; MIYATA *et al.* 1980; TICHER and GRAUR 1989). However, if the rate of substitution at nonsynonymous sites is greater than at synonymous sites, positive selection may be indicated.

Estimates of Ds and Dn were not significantly different for pairwise comparisons of sequences from Bos and Bison species, although the proportion of synonymous substitutions per site over the entire exon was greater than that of nonsynonymous substitutions (Table 3). In these comparisons, nonsynonymous substitutions were seen only in comparisons to *B. taurus*. Within the PDR of exon four, *B. javanicus*, *B. gaurus*, *B. bison*, and *B. bonasus* sequences were identical, and the PDR sequence of *B. grunniens* differed from these other sequences only in the presence of a four-codon direct repeat.

The corrected proportions of synonymous and nonsynonymous substitutions for pairwise comparisons of sequences from *B. javanicus*, *S. caffer*, *T. imberbis*, *B. tragocamelus*, and *C. hircus* were determined. Again, *B. javanicus* was chosen to represent the Bos and Bison genera in order to facilitate analysis because identical results were found for all other species from these genera, except *B. taurus*. In these comparisons, average Dn exceeded average Ds within all regions of exon four, yet none of these differences were statistically significant (Table 3). The largest difference in Ds and Dn was seen within the PDR of exon four, where only 11 of 34 amino acids were conserved, and the value of Dn was more than twice the value of Ds. Again, this difference was not statistically significant.

We examined all 10 pairwise comparisons of PDR sequences from distantly related bovid taxa individually. The value of Dn exceeded the value of Ds in all 10 of these comparisons and this difference was statistically significant in three comparisons (Table 4), indicating that the accelerated nonsynonymous substitution rate was not limited to a single point in the evolution of these taxa. We examined this further by determining the value of Ds and Dn within the PDR for each of the branches in all three of the probable species trees (Figure 6). This was done by including the ancestral sequences inferred using PAUP 3.1.1 (both ACCTRAN and DELTRAN separately) into the alignment of se-

TABLE 3

Average proportion of synonymous and nonsynonymous substitutions per site for comparisons of closely related (Bos and Bison species) and distantly related taxa (divergence time ≥10 my)

Gene region	Closely related taxa		Distantly related taxa	
	Ds (%)	Dn (%)	Ds (%)	Dn (%)
Exon 4 (total)	1.16 ± 0.67	0.13 ± 0.13	4.48 ± 1.49	6.88 ± 1.06
PKC	0.82 ± 0.82	0	1.97 ± 1.41	3.49 ± 1.11
PDR	0	0.50 ± 0.50	7.12 ± 3.71	15.20 ± 3.29
CMP	1.49 ± 1.06	0.24 ± 0.24	6.87 ± 2.60	10.26 ± 1.83

Ds, synonymous; Dn, nonsynonymous.

TABLE 4

Proportion of synonymous and nonsynonymous substitutions per site for pairwise comparisons of PDR sequences from distantly related taxa (divergence times ≥10 my)

Pairwise comparison	Ds (%)	Dn (%)	d = Dn − Ds (%)
B. javanicus × *S. caffer*	9.31 ± 6.70	11.22 ± 4.33	1.91 ± 7.98
B. javanicus × *T. imberbis*	9.24 ± 6.65	16.64 ± 5.43	7.40 ± 8.59
B. javanicus × *B. tragocamelus*	14.22 ± 8.43	16.69 ± 5.45	2.47 ± 10.04
B. javanicus × *C. hircus*	13.56 ± 8.27	24.77 ± 6.93	11.21 ± 10.79
S. caffer × *T. imberbis*	0.00 ± 0.00	11.22 ± 4.33	11.22 ± 4.33***
S. caffer × *B. tragocamelus*	4.48 ± 4.51	11.25 ± 4.34	6.77 ± 6.26
S. caffer × *C. hircus*	3.77 ± 4.16	15.02 ± 5.11	11.25 ± 6.59*
T. imberbis × *B. tragocamelus*	4.45 ± 4.48	11.28 ± 4.35	6.83 ± 6.24
T. imberbis × *C. hircus*	3.74 ± 4.13	18.80 ± 5.84	15.06 ± 7.15**
B. tragocamelus × *C. hircus*	8.43 ± 6.32	15.11 ± 5.14	6.68 ± 8.15
Average	7.12 ± 3.71	15.20 ± 3.29	8.08 ± 4.96

Significance of comparison: * $P < 0.05$; ** $P < 0.025$, and *** $P < 0.005$.

quences for distantly related terminal taxa. MEGA was then used to determine the value of Ds and Dn along every branch in the tree. As can be seen in Figure 6, Dn is greater than Ds along the majority of branches in each of the three trees. Graphical measures of divergence at synonymous sites within the PDR did not reveal any evidence of saturation for synonymous substitutions within this region (Figure 7). Therefore, these results provide strong evidence for the acceleration of divergence among bovid κ-casein proteins being a consequence of positive Darwinian selection.

The difficulty in statistically demonstrating that the observed differences in the values of Dn and Ds (Table 4) are due to the influence of positive selection is largely attributable to the small number of nucleotides examined and the collective variance associated with averages of pairwise comparisons that differ by a factor of two in time since divergence and consequently their overall level of divergence. In terms of the small number of nucleotides examined, if selection is affecting a very narrowly defined region of a gene, as we believe is the case within κ-casein exon four, then there is little that can be done to increase the resolution of the analysis.

Because nucleotide composition bias can constrain the rate of substitution at silent sites (METZ and PALUMBI 1996), the nucleotide composition of the various regions of the κ-casein gene were determined (Table 5). A strong A+T bias was seen in both intron four and the third codon positions of exon four. This result is consistent with the theory that mutational bias should favor the accumulation of A+T at sites free of selective constraint (LI 1997). Given that there was little variation in nucleotide composition among species or among regions of the exon, it is unlikely that differences in nucleotide composition can account for the highly accelerated rates of nucleotide and amino acid sequence divergence focused within the PDR region of exon four.

Synonymous codon usage bias for exon four sequences was moderate (0.42) as measured with a scaled chi square (SHIELDS *et al.* 1988). While equal codon usage would result in a relative synonymous codon usage (RSCU) value of one (SHIELDS *et al.* 1988), the RSCU value for exon four sequences was greater for A+T (1.31 ± 0.17) than for G+C (0.67 ± 0.13) ending codons. This result is consistent with previous findings that codon usage bias in vertebrates is correlated with mutational bias and is not the result of selective constraint on third position sites (HASTINGS and EMERSON 1983; HATFIELD and RICE 1986; WELLS *et al.* 1986; BULMER 1988). Therefore, it is unlikely that the higher rate of nonsynonymous change relative to synonymous change is due to selective constraints on silent sites within the exon.

DISCUSSION

If a strictly neutral model was appropriate for the evolution of κ-casein exon four as GATESY *et al.* (1996) have suggested, one would expect to see an equal number of substitutions at each of the three codon positions and in the intron as well. In comparisons of distantly related bovid taxa, there was an uneven distribution of changes along the gene. In fact, there was an increased rate of nonsynonymous substitution concentrated within a 34-codon region of caseinomacropeptide, and this region as well as the exon as a whole had a rate of total nucleotide substitution that exceeded that of intron four. Because there is no evidence that intron four is saturated, the overall increase in rate of substitution described above indicates that positive Darwinian selection has accelerated the divergence of κ-casein in distantly related bovid taxa. The apparent conflict between this conclusion and that of GATESY *et al.* (1996) may be due to the fact that they compared Ds and Dn for closely related taxa only. In addition, these authors did not examine the different regions of exon four individually

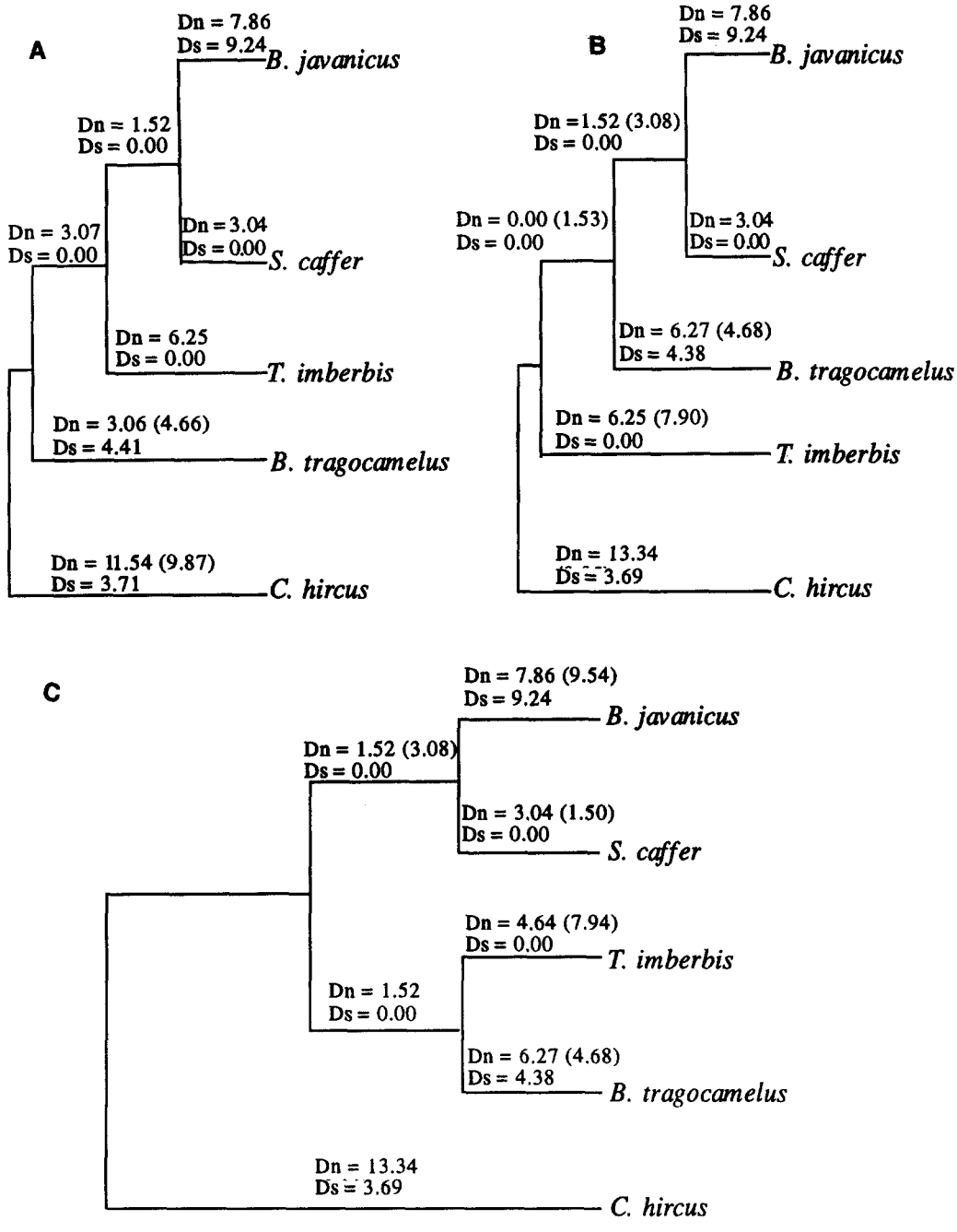

FIGURE 6.—Three possible species trees for distantly related bovid taxa with Jukes-Cantor-corrected proportions of synonymous (Ds) and nonsynonymous (Dn) substitutions per site shown along each branch. The results of using the DELTRAN optimization option to reconstruct the interior nodes are shown in parenthesis on branches where they differ from the results obtained with ACCTRAN optimization.

and did not compare the level of variation within intron four to that within exon four. Therefore, a more detailed statistical analysis of exon four sequences reported by GATESY *et al.* (1996) and the addition of intron four sequences for the taxa included in their study would be an interesting avenue for future investigation.

The pattern of nucleotide substitution seen in comparisons of closely related bovid taxa provides no evidence of positive selection, however total nucleotide sequence divergence was nearly twice as great within intron four than within the PDR or exon four as a whole. This may indicate that purifying selection has constrained the divergence of sites within exon four in comparisons of closely related bovid taxa. However, the extremely small number of nucleotide changes within κ-casein exon four between these closely related taxa makes it impossible to demonstrate that differences in Ds and Dn are statistically significant.

One possible explanation for the observed pattern of nucleotide substitution is that the selectively optimal state for κ-casein has changed in different bovid lineages as taxa diverged and differentiated. Positive selection then accelerated interspecific divergence at the amino acid level as alleles approximating these new and divergent optima were favored. This would explain the high level of divergence within the PDR relative to in-

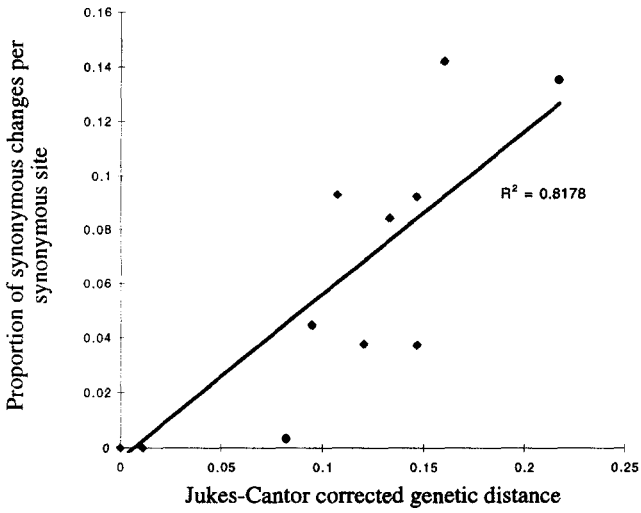

FIGURE 7.—Sequence divergence at synonymous sites within the PDR of exon four for pairwise comparisons of distantly related bovid taxa.

tron four in comparisons of distantly related bovid taxa. Selective sweeps (MAYNARD SMITH and HAIGH 1974) of neutral or nearly neutral variation that had developed during the process of species differentiation at sites close to the site under strong positive selection would be expected to result in low levels of polymorphism within these species.

The very closely related Bos and Bison species may be biologically similar enough that a common optimum for κ-casein has been retained. Purifying selection would then be expected to maintain an optimal allele in all of these species, resulting in low levels of divergence within the PDR relative to intron four as was observed. In addition, directional selection may have recently fixed an optimal allele, resulting in a selective sweep of neutral or nearly neutral ancestral variation at sites linked to the site under selection.

The scenario described above would result in high interspecific divergence (between species with some minimal amount of differentiation) and low intraspecific polymorphism, violating the predictions of a strictly neutral model (KIMURA 1983) that levels of divergence and polymorphism should be concordant. Although data pertaining to intraspecific variation within the κ-casein gene are limited, they do seem to fit the expectation of low polymorphism. KAWAMOTO *et al.* (1992) report that, of the κ-casein genes examined from 22 individual water buffalo (*Bubalus bubalis*), there was no polymorphism detected by isoelectric focusing. In addition, we found no evidence of polymorphism within exon four of the three species in which multiple individuals were sequenced or between the two subspecies of *S. caffer*. However, more extensive investigations of these species need to be conducted.

In most of the previous examples of positive selection, the selective agent was identified as playing a role in immunity or some process of self recognition (HUGHES 1992; LEE *et al.* 1995; SWANSON and VAQUIER 1995). Identification of the selective agent for κ-casein is not easy because it is known to be involved in a variety of physiological processes and consists of two separate peptides, which may be exposed to very different selective pressures. However, the role that caseinomacropeptide plays in the downregulation of immune response in neonates seems to be a likely candidate for the selective agent. It also is conceivable that the selective pressure could come from caseinomacropeptide's antimicrobial activity. Different genetic variants of κ-casein produce changes in glycosylation patterns (LODES *et al.* 1996) that are known to be species specific and that influence the inhibition of infection by gastric pathogens (STRÖMQVIST *et al.* 1995). While it is not known if the functions of caseinomacropeptide described above are attributable to the amino acids within the PDR, it is interesting to note that caseinomacropeptide is responsible for all of these functions and that the PDR segment of this peptide is the region revealing the strongest evidence of positive selection. Thus, the biological and physiological information that we have is consistent with the observed pattern of evolution at the molecular level.

Although positive selection has been demonstrated in a number of previous examples, most of these involve comparisons of allelic variation within a single species or within a very closely related group of species (HUGHES and NEI 1988; HUGHES and NEI 1989; METZ and PALUMBI 1996). Thus, in the majority of the proteins for which positive selection has been demonstrated, the level of polymorphism is exceptionally high.

TABLE 5

Nucleotide composition of κ-casein gene regions by codon position

	Codon position 1		Codon position 2		Codon position 3		All positions	
	A + T	G + C	A + T	G + C	A + T	G + C	A + T	G + C
Exon 4 (total)	47.1	52.9	51.0	49.0	66.8	33.2	55.0	45.0
PKC	45.4	54.6	53.1	46.9	64.4	35.5	54.3	45.7
PDR	39.8	60.1	49.0	51.0	66.4	33.6	51.8	48.3
CMP	48.7	51.3	49.1	50.9	69.1	31.0	55.6	44.4
Intron 4	N/A	N/A	N/A	N/A	N/A	N/A	72.5	27.5

The results of our examination of the molecular evolution of κ-casein differ from most of these previous studies in that interspecific divergence is high, yet there is no evidence to suggest that intraspecific polymorphism is elevated. In addition, the results of this study are unique in that they demonstrate that positive selection at the κ-casein locus is a general factor in the evolution of taxa representing an entire mammalian family.

We are grateful to JOE BIELAWSKI, ALEX ROONEY, and JERRY TAYLOR for helpful comments and discussion and to NICK GRASSLY for instruction on the use of the computer program PLATO. We also thank RON ADKINS for advice on the use of the Fisher's exact test. In addition, we would like to acknowledge the two anonymous reviewers for their helpful comments. We also thank CLAIRE KOLENDA and DOUG MELENDY for technical assistance and JIM WOMACK for donations of DNA from various species. This work was supported by National Science Foundation grants DEB-9622126 (J.N.D.) and DEB-9615163 (R.L.H.).

LITERATURE CITED

ALEXANDER, L. J., A. F. STEWART, A. G. MACKINLAY, T. V. KAPELINSKAYA, T. M. TKACH *et al.*, 1988 Isolation and characterization of the bovine kappa-casein gene. Eur. J. Biochem. **178:** 395–401.

ALLARD, M. W., M. M. MIYAMOTO, L. JARECKI, F. KRAUS and M. R. TENNANT, 1992 DNA systematics and evolution of the artiodactyl family Bovidae. Proc. Natl. Acad. Sci. USA **89:** 3972–3976.

BOHLKEN, H., 1958 Vergleichende untersuchengen an wildrindern (TRIBUS BOVINI SIMPSON, 1945). Zool. Jahrb. **68:** 113–202.

BOHLKEN, H., 1961 Haustiere und Zoologische systematik. Z. Tierz. Zuechtungsbiol. **76:** 107–113.

BULMER, M., 1988 Are codon usage patterns in unicellular organisms determined by selection-mutation balance? J. Evol. Biol. **1:** 15–26.

DEV, B. C., S. M. SOOD, S. DEWIND and C. W. SLATTERY, 1994 Kappa-casein and beta-caseins in human milk micelles: structural studies. Arch. Biochem. Biophys. **314:** 329–336.

FELSENSTEIN, J., 1995 PHYLIP version 3.572c. Department of Genetics, University of Washington, Seattle.

GATESY, J., C. HAYASHI, M. A. CRONIN and P. ARCTANDER, 1996 Evidence from milk casein genes that cetaceans are close relatives of hippopotamid artiodactyls. Mol. Biol. Evol. **13:** 954–963.

GRASSLY, N. C., and E. C. HOLMES, 1997 A likelihood method for the detection of selection and recombination using nucleotide sequences. Mol. Biol. Evol. **14:** 239–247.

GROVES, C. P., 1981 Systematic relationships in the Bovini (Artiodactyla, Bovidae). Z. Zool. Syst. Evolutionforsch. **19:** 264–278.

GUTIERREZ, A. A., E. A. MAGA, H. MEADE, C. F. SHOEMAKER, J. F. MEDRANO *et al.*, 1996 Alterations of the physical characteristics of milk from transgenic mice producing bovine kappa-casein. J. Dairy Sci. **79:** 791–799.

HASTINGS, K. E. M., and C. P. EMERSON, 1983 Codon usage in muscle genes and liver genes. J. Mol. Evol. **19:** 214–218.

HATFIELD, D., and M. RICE, 1986 Aminoacyl-tRNA (anticodon): codon adaptation in human and rabbit reticulocytes. Biochem. Int. **13:** 835–842.

HIGGINS, D. G., and P. M. SHARP, 1989 Fast and sensitive multiple sequence alignments on a microcomputer. CABIOS **5:** 151–153.

HUGHES, A. L., 1992 Positive selection and interallelic recombination at the merozoite surface antigen-1 (MSA-1) locus of *Plasmodium falciparum.* Mol. Biol. Evol. **9:** 381–393.

HUGHES, A. L., and M. NEI, 1988 Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature **335:** 167–170.

HUGHES, A. L., and M. NEI, 1989 Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc. Natl. Acad. Sci. USA **86:** 958–962.

JANECEK, L. L., R. L. HONEYCUTT, R. M. ADKINS and S. K. DAVIS, 1996 Mitochondrial gene sequences and the molecular systematics of the artiodactyl subfamily Bovinae. Mol. Phylogenet. Evol. **6:** 107–119.

JUKES, T. H., and J. L. KING, 1979 Evolutionary nucleotide replacements in DNA. Nature **281:** 605–606.

KAWAMOTO, Y., T. AMANO, T. NAMIKAWA, T. NISHIDA and H. B. RAJUBHANDARY, 1992 Milk protein polymorphisms of water buffaloes in Nepal. Anim. Sci. Technol. **63:** 270–276.

KIMURA, M., 1977 Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature **267:** 275–276.

KIMURA, M., 1983 *The Neutral Theory of Evolution.* Cambridge University Press, Cambridge.

KUMAR, S., K. TAMURA and M. NEI, 1993 MEGA: molecular evolutionary genetics analysis, version 1.01. The Pennsylvania State University, University Park, PA.

LEE, Y., T. OTA and V. D. VACQUIER, 1995 Positive selection is a general phenomenon in the evolution of abalone sperm lysin. Mol. Biol. Evol. **12:** 231–238.

LI, W., 1997 *Molecular Evolution.* Sinauer Associates, Sunderland, MA.

LODES, A., I. KRAUSE, J. BUCHBERGER, J. AUMANN and H. KLOSTERMEYER, 1996 The influence of genetic variants of milk proteins on the compositional and technological properties of milk: 1. Casein micelle size and the content of non-glycosylated κ-casein. Milchwissenschaft **51:** 368–373.

MADDISON, W. P., and D. R. MADDISON, 1996 MacClade version 3.06. Sinauer Associates, Sunderland, MA.

MARZALI, A. S., and K. F. NG-KWAI-HANG, 1986 Relationships between milk protein polymorphisms and cheese yielding capacity. J. Dairy Sci. **69:** 1193–1201.

MAYNARD SMITH, J., and J. HAIGH, 1974 The hitch-hiking effect of a favorable gene. Genet. Res. **23:** 23–35.

MCDONALD, J. N., 1981 *North American Bison: Their Classification and Evolution.* Univ. of California Press, Berkeley, CA.

MERCIER, J. C., J. M. CHOBERT and F. ADDEO, 1976 Comparative analysis of the amino acid sequences of the caseinomacropeptides from seven species. FEBS Lett. **72:** 208–214.

METZ, E. C., and S. R. PALUMBI, 1996 Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein binding. Mol. Biol. Evol. **13:** 397–406.

MIYATA, T., T. YASUNAGA and T. NISHIDA, 1980 Nucleotide sequence divergence and functional constraint in mRNA evolution. Proc. Natl. Acad. Sci. USA **77:** 7328–7332.

MODI, W. S., D. S. GALLAGHER and J. E. WOMACK, 1996 Evolutionary histories of highly repeated DNA families among the Artiodactyla (Mammalia). J. Mol. Evol. **42:** 337–349.

NEI, M., and T. GOJOBORI, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:** 418–426.

NEI, M., and L. JIN, 1989 Variances of the average numbers of nucleotide substitutions within and between populations. Mol. Biol. Evol. **6:** 290–300.

OTANI, H., and M. MONNAI, 1993 Inhibition of proliferative responses of mouse spleen lymphocytes by bovine milk κ-casein digests. Food Agric. Immunol. **5:** 219–229.

PILGRIM, G. E., 1947 The evolution of the buffaloes, oxen, sheep, and goats. J. Linn. Soc. Lond. Zool. **41:** 272–286.

PINDER, S. J., B. N. PERRY, C. J. SKIDMORE and D. SAVVA, 1991 Analysis of polymorphism in the bovine casein genes by use of the polymerase chain reaction. Anim. Genet. **22:** 11–20.

QIAN, Z. Y., P. JOLLS, D. MIGLIORE-SAMOUR, F. SCHOENTGEN and A. FIAT, 1995 Sheep κ-casein peptides inhibit platelet aggregation. Biochem. Biophys. Acta **1244:** 411–417.

SAIKI, R. K., D. H. GELFEND, S. STOFFEL, S. J. SCHARF, R. HIGUCHI *et al.*, 1988 Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:** 487–491.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual,* 2nd ed. Cold Spring Harbor Laboratory Press, Plainview, NY.

SHIELDS, D. C., P. M. SHARP, D. G. HIGGINS and F. WRIGHT, 1988 "Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. Mol. Biol. Evol. **5:** 704–716.

STRÖMQVIST, M., P. FALK, S. BERGSTROM, L. HANSSON, B. LONNERDAL *et al.*, 1995 Human milk κ-casein and inhibition of *Helicobacter*

pyloi adhesion to human gastric mucosa. J. Pediatr. Gastroenterol. Nutr. **21:** 288–296.

SWANSON, W. J., and V. D. VACQUIER, 1995 Extraordinary divergence and positive Darwinian selection in a fusagenic protein coating the acrosomal process of abalone spermatozoa. Proc. Natl. Acad. Sci. USA **92:** 4957–4961.

SWOFFORD, D. L., 1993 PAUP: phylogenetic analysis using parsimony, version 3.1.1. Illinois Natural History Survey, Champaign, IL.

TICHER, A., and D. GRAUR, 1989 Nucleic acid composition, codon usage, and the rate of synonymous substitution in protein-coding genes. J. Mol. Evol. **28:** 286–298.

WALL, D. A., S. K. DAVIS and B. M. READ, 1992 Phylogenetic relationships in the subfamily Bovinae (Mammalia: Artiodactyla) based on ribosomal DNA. J. Mammal. **73:** 262–275.

WELLS, D., W. BAUIS and L. KEDES, 1986 Codon usage in histone gene families of higher eukaryotes reflects functional rather than phylogenetic relationships. J. Mol. Evol. **23:** 224–241.

Communicating editor: W-H. LI