J Mol Evol (1996) 42:676–684

JOURNAL OF MOLECULAR EVOLUTION

Molecular Genetics and Evolution of Stomach and Nonstomach Lysozymes in the Hoatzin

Janet R. Kornegay*

Department of Molecular and Cell Biology, 401 Barker Hall, University of California, Berkeley, CA 94720-3202, USA

Received: 11 December 1995 / Accepted: 29 February 1996

Abstract. Multiple genes of the hoatzin encoding stomach lysozyme *c* and closely related members of this calcium-binding lysozyme *c* group were cloned from a genomic DNA library and sequenced. There are a minimum of five genes represented among these sequences that encode two distinct groups of protein sequences. One group of three genes corresponds to the stomach lysozyme amino acid sequences, and the remaining genes encode predicted proteins that are more basic in character and share several sequence identities with the pigeon egg-white lysozyme rather than with the hoatzin stomach lysozymes. Despite these structural similarities between some of the hoatzin gene products and the pigeon lysozyme, phylogenetic analyses indicate that all of the hoatzin sequences are closely related to one another. This is borne out by the relatively small genetic distances even in the intronic regions, which are not subject to the selective pressures operating on the coding regions of the stomach lysozymes. These results suggest that multiple gene duplication events have occurred during the evolution of hoatzin lysozymes.

Key words: Hoatzin — Stomach lysozyme — Foregut fermentation — Gene sequences — Gene duplication — Genetic variation — Lysozyme evolution

Introduction

Gene duplication events facilitate the development of new protein functions and have thus played a major role in the evolution of genes and genomes (Haldane 1932; Ohno 1970). One model for this phenomenon (Wistow 1993) proposes that a protein would acquire new utility through sequence drift, creating a selective pressure for gene duplication to maintain separation of functions. Alternatively, an unselected duplication can occur and the original function of a given gene is maintained by one copy, while the additional copy is free to mutate and evolve for the novel function. In either case, evolution of a new biochemical function may occur solely through nucleic acid substitution events or through more dramatic changes such as reshuffling of portions of genes via recombination (Gilbert 1978). This is an elegant recycling of biochemical structures and functions by molding and adapting existing gene products and by forming mosaic gene products through exon shuffling, all facilitating new ways and means of organismal evolution over eons of time.

In the case of the ruminant stomach lysozymes *c*, gene duplication events appear to have occurred concomitant with the evolution of the physiological adaptation leading to foregut fermentation (Irwin and Wilson 1989; Irwin et al. 1989, 1992). Not counting a putative calcium-binding lysozyme *c* gene, the number of lysozyme *c* genes present in the ruminant genome increases from the typical single copy found in mammals lacking foregut fermentation to at least 10 genes (including some pseudogenes) in the cow (Irwin et al. 1989; Irwin 1995). As one moves through the phylogenetic tree from the non-

* *Present address:* Roche Molecular Systems, 1145 Atlantic Ave., Alameda, CA 94501, USA

ruminants to those with a relatively simple foregut, and then on to those possessing the more complex chambered stomachs, the copy number of lysozyme increases accordingly (Irwin et al. 1989, 1992). The extra copies presumably facilitated a recruitment event for function in the stomach. An additional consequence of this gene amplification in the cow, for example, is an increase in the level of expression of lysozyme in the true stomach, as demonstrated by the presence of three distinct isozymes in the stomach tissue in relatively large amounts (Pahud and Widmer 1982; Dobson et al. 1984). There are at least four lysozyme genes expressed in the stomach (Irwin and Wilson 1989, 1990). Many of the cow genes exhibit evidence of gene conversion (Irwin and Wilson 1990), suggesting the maintenance of biochemical function as a result of strong selective pressures. This mechanism of concerted evolution prevents duplicate copies of the stomach lysozymes from straying away from their necessary function through further mutation. In contrast, colobine monkeys such as the langur possess only a single non-calcium-binding lysozyme *c* gene (Swanson et al. 1991), even though they exhibit parallel molecular adaptations for foregut fermentation such as the recruitment and structural evolution of lysozyme as a digestive enzyme (Stewart et al. 1987).

The neotropical hoatzin *(Opisthocomus hoazin)* is the only known avian foregut fermenter. Its enlarged crop contains symbiotic microflora and acts as a fermentative chamber, allowing it to thrive on a diet of leaves (Grajal et al. 1989). A stomach lysozyme has been isolated and characterized from the hoatzin proventiculus that demonstrates the hallmark structural adaptations required for adaptation to the novel stomach environment (Kornegay et al. 1994). In the hoatzin, a calcium-binding lysozyme *c* gene was recruited to express the stomach enzyme, rather than a conventional type as is the case for the mammalian foregut fermenters.[1] Despite this different genetic origin, it follows that an amplification of lysozyme genes could have occurred in the hoatzin genome as a part of the foregut adaptation. This paper reports nine genomic lysozyme sequences for the hoatzin, representing a minimum of five distinct gene loci. Six of these genomic sequences appear to represent expressed hoatzin stomach lysozymes, while the other three sequences encode putative lysozymes that are closely related, but would be biochemically distinct from those expressed in the stomach. Although complete or partial amino acid sequences are known for calcium-binding lysozymes *c* from nine vertebrate species (Prager and Jollès 1996), this is the first report of the structure and organization of genes encoding this type (as opposed to the conventional type) of lysozyme *c*.

[1] Among vertebrates, the lysozyme *c* gene family has at least two major branches which arose from some very ancient gene duplication event(s) preceding the divergence of birds and mammals (Dautigny et al. 1991; Prager and Jollès 1996). All of the conventional lysozymes *c* found in vertebrates arose from one of those very deep lineages. The functionally calcium-binding lysozymes and their relatives comprise the descendants of the other deep branch(es) among the lysozymes *c*.

Materials and Methods

Genomic DNAs. High-molecular-weight genomic DNA was prepared from tissues of a single hoatzin (obtained from the frozen collection at the Museum of Natural Science, Louisiana State University), as described previously (Kornegay et al. 1994).

Lysozyme Probes. Two different hoatzin lysozyme probes were employed. The first probe was the relatively small fragment encoding amino acid residues 39–67, amplified from genomic DNA using degenerate oligonucleotide primers based on protein sequence data (cf. Kornegay et al. 1994). A second generation of hoatzin lysozyme probes was generated via PCR from a hoatzin stomach cDNA preparation using flanking primers that amplify the entire coding region of the cDNAs expressed in the hoatzin stomach (Kornegay et al. 1994). Probes were labeled with either ^{32}P-dCTP or digoxigenin-dCTP by incorporation with PCR using the following parameters: 1 min at 94°C, 1 min at 50–55°C, and 1 min 30 s at 72°C for 35 cycles.

Hoatzin Genomic Library Construction and Screening. The high-molecular-weight hoatzin genomic DNA was partially digested with the restriction enzyme *Sau*3A and the resulting products were fractionated on a sucrose gradient (Sambrook et al. 1989). The DNA fragments ranging from 15 to 20 kb in size were pooled and concentrated. With a bacteriophage λ cloning kit from Stratagene, these fragments were ligated to the arm fragments of a λDASH vector, packaged into bacteriophage particles, titered, and stored at 4°C. Once established, a portion of the library was plated out (~600,000 pfu) on a methylase-restriction-deficient *E. coli* host strain and screened by standard methods (Sambrook et al. 1989). The initial library screening utilized ^{32}P-radiolabeled probes, while nonradioactive digoxigenin-labeled probes were utilized in the three successive clone purification screenings. The hybridizations and detections using digoxigenin-labeled probes were done with the Genius system (Boehringer Mannheim) using Lumi-Phos 530.

Subcloning and Sequencing of Genomic Lysozyme Clones. Fragments of the λ genomic clones ranging from 3 to 6 kb in length were subcloned into a BlueScript plasmid vector (Sambrook et al. 1989). These fragments were selected based on Southern blot analysis of the λ clones restricted with various enzymes (data not shown). The HSLλ clone series was sequenced by working outward from the known sequence of the original probe region and inward from the ends of subcloned fragments. As sequence data became available, and after the stomach lysozyme cDNA sequence was determined (Kornegay et al. 1994), PCR and direct sequencing of overlapping regions in the λ clones were performed to complete the clone sequences. The primers used for amplification and sequencing of these HSLλ clones are shown in Fig. 1 and Table 1.

Sequence Analyses. Nucleic acid sequences were entered and aligned using the GEL program, version 5.4, in the IntelliGenetics software package. Overlapping fragment sequences for each clone were joined together in individual files and then the complete clone sequences were aligned with each other for each class of sequences.

Evolutionary Analyses. PAUP (Swofford 1993) was used to determine the phylogenetic relationship among the hoatzin nucleic acid

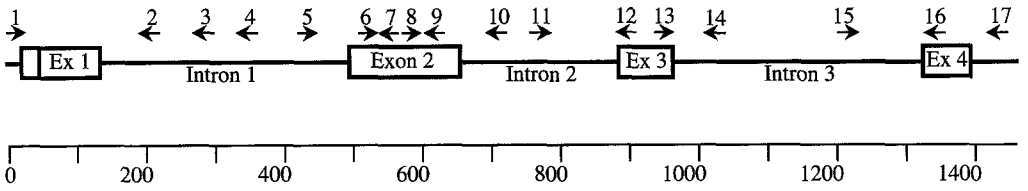

Fig. 1. Strategy for amplification and sequencing of the HSLλ series of hoatzin lysozyme genomic clones. In the schematic gene *(top), open bars* depict the coding regions of the four exons of the lysozyme genes; *lines* denote the flanking, 5′ and 3′ untranslated, and intronic regions. *Arrows* indicate the oligonucleotide primers employed (described in detail in Table 1). The units shown along the *scale bar (bottom)* correspond to the nucleotide numbering for the gene sequences in Fig. 2.

Table 1. Oligonucleotide primers used in the cloning, amplification, and sequencing of hoatzin lysozyme genes[a]

Number	Name	Sequence (5′ to 3′)
1	5′Leader	ATGAGAAAACTCAATGCTCT
2	Int1C	AGGCAGAGAGATGGTGAAGT
3	Int1B	AGAAAGCTCTGGATGGGTAT
4	Int1A	GATACGGTAAGTCTGAGGAA
5	ATOTX2	ATTTACCCTGAGCATCCCTTGACGT
6	Exon2E	AA**Y**AC**N**GA**R**GC**N**TACAA
7	AHSLX2–5′	CCCTGCTTGGACCATTGTTG
8	BHSLX2–3′	TTCAGATCAACAGCAAGTAC
9	Exon2C	TT**N**CC**R**TC**R**TY**R**CACCA
10	BTOTX2	ATCGAGAGGATGAAGGAAGGCTCGT
11	Int2A	CCTTCCTGATGAATGAACACCT
12	Exon3A	TCCTCAAGATCATTTGTCAT
13	Exon3B	GATGCTCACGGCCTCACTCC
14	Int3A	CCAGTCAACAGCAGGATTTTC
15	Int3B	GAGCTCATCAGAATTTTCTC
16	Exon4A	CCCTCGCAATGGTTTTTCCA
17	3′Flanker	GAGATCAGTGCTGCTTGCGGA

[a] Boldface type highlights degenerate bases. Primers 1, 6, 7, 9, and 17 have been described previously (Kornegay et al. 1994). Primers 6 and 9 were used to generate the initial lysozyme probe, as described in Materials and Methods. Primers 1, 2, 4–6, 8, 10–14, 16, and 17 were used for amplification via PCR. All primers except 6 and 9 were used for sequencing

sequences reported. Bootstrap analysis was performed using PAUP with random resampling of the data and parsimony reconstruction reiterated 1,000 times. The computer program MacClade (Maddison and Maddison 1992) was used to investigate the various possible combinations of alleles among the lysozyme sequences. To examine overall genetic diversity, pairwise differences between hoatzin sequences were calculated using the PAUP program.

Results

Isolation of Lysozyme Genes

In earlier Southern analyses of hoatzin genomic DNA compared with DNA of other birds (Kornegay 1994), the hoatzin stomach lysozyme probe appeared to light up more than one gene based on the abundance of bands distributed at higher molecular weights. The same probe was used in a primary screening of the hoatzin genomic library to isolate clones containing lysozyme genes. This screening yielded 15 different clones, labeled the HSLλ series, which were assumed and then found to represent at least two different types of lysozyme genes. One of these types would logically encode the stomach lysozyme, and any other genes represented in the HSLλ series would presumably be genetically related to the calcium-binding branch of the lysozyme *c* family. There are at least nine distinct sequences for lysozyme genes among the 15 HSLλ clones shown in Fig. 2. Allowing for two alleles per gene locus, one could conclude that the hoatzin genome contains at least five loci belonging to this calcium-binding/stomach lysozyme group.

The sequences presented here begin 10 bases upstream of the translation start site, encompass the entire coding region of the genes (including intronic sequences), and end 25 bases into the 3′ untranslated region. Four complete clone sequences were omitted from Fig. 2 because they are identical to other clones: HSLλ14 and HSLλ17 are the same as HSLλ15; HSLλ11 and HSLλ16 are identical to HSLλ8. The first 214 bases of the alignment were unattainable for one of the unique clone sequences presented, HSLλ15. For HSLλ1, HSLλ5, and HSLλ10, the sequences begin 47–50 bases into the alignment presented here. Only small fragments of sequence data were obtained for two of the HSLλ clones (HSLλ19 and HSLλ20—not included in Fig. 2), none of which contained any unique base substitutions when aligned with the other sequences.

The hoatzin genes are identical to those encoding conventional vertebrate lysozymes *c* with respect to their organization into four exons plus three introns and the locations of the introns. The entire length for the consensus of these fragments of DNA is 1,422 bases when all HSLλ sequences are aligned. Only a short fragment of the 3′ untranslated region is included in these data, although this region is normally a few hundred bases long in many genes, including lysozyme *c* genes (Irwin et al. 1996). If a length of up to 500 bp for the 3′ untranslated region of the hoatzin lysozyme gene is assumed, one can still infer that this is a relatively small gene. The conventional lysozyme gene of chicken is ~4 kb in length (Jung et al. 1980), and the mammalian stomach lysozyme genes characterized in cows range in length from ~6 to 8 kb (Irwin et al. 1993); these values span the entire size range for conventional vertebrate lysozyme *c* genes

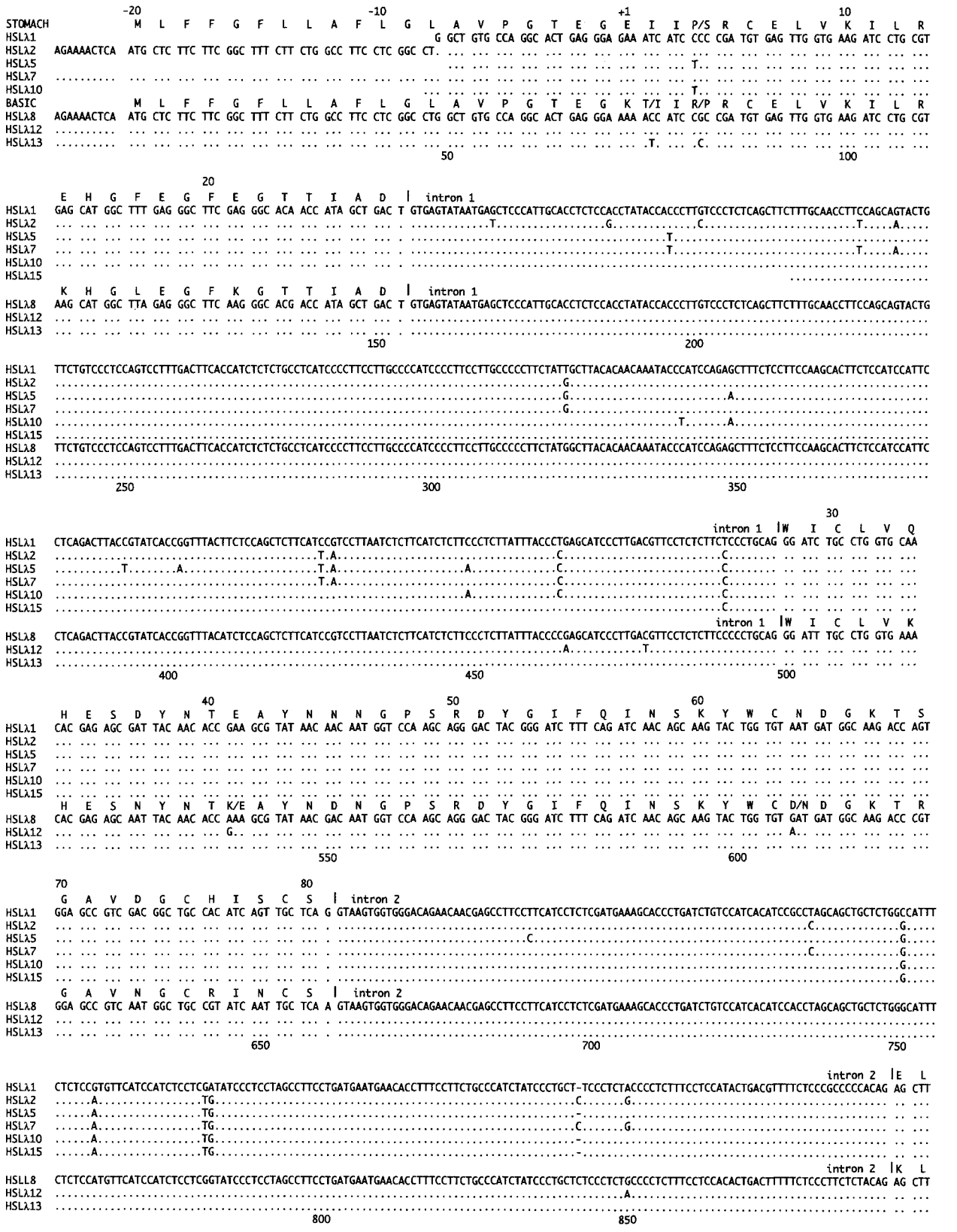

Fig. 2. Sequences of genomic clones encoding putative hoatzin lysozymes. The first six unique sequences shown correspond to a minimum of three genes for stomach lysozyme encoding the two isozymes found in the hoatzin proventriculus. *Dots* indicate identity with the first sequence (HSLλ1). The predicted amino acid sequences are shown above the coding regions for the six genomic clone sequences. The last three unique sequences correspond to a minimum of two genes encoding putative lysozymes that are biochemically more basic. *Dots* indicate identity to the HSLλ8 sequence and the predicted amino acid sequences are shown above the coding regions for the lysozyme genes. These genes display no evidence that they are inactive pseudogenes; however, it is unknown whether the predicted basic proteins are expressed in any hoatzin tissue(s). For all the sequences, *numbering* for the amino acid sequence is given *above*, while the *numbering* for nucleic acids is shown *below*. *Dashes* denote length variations due to the insertion or deletion of nucleotides.

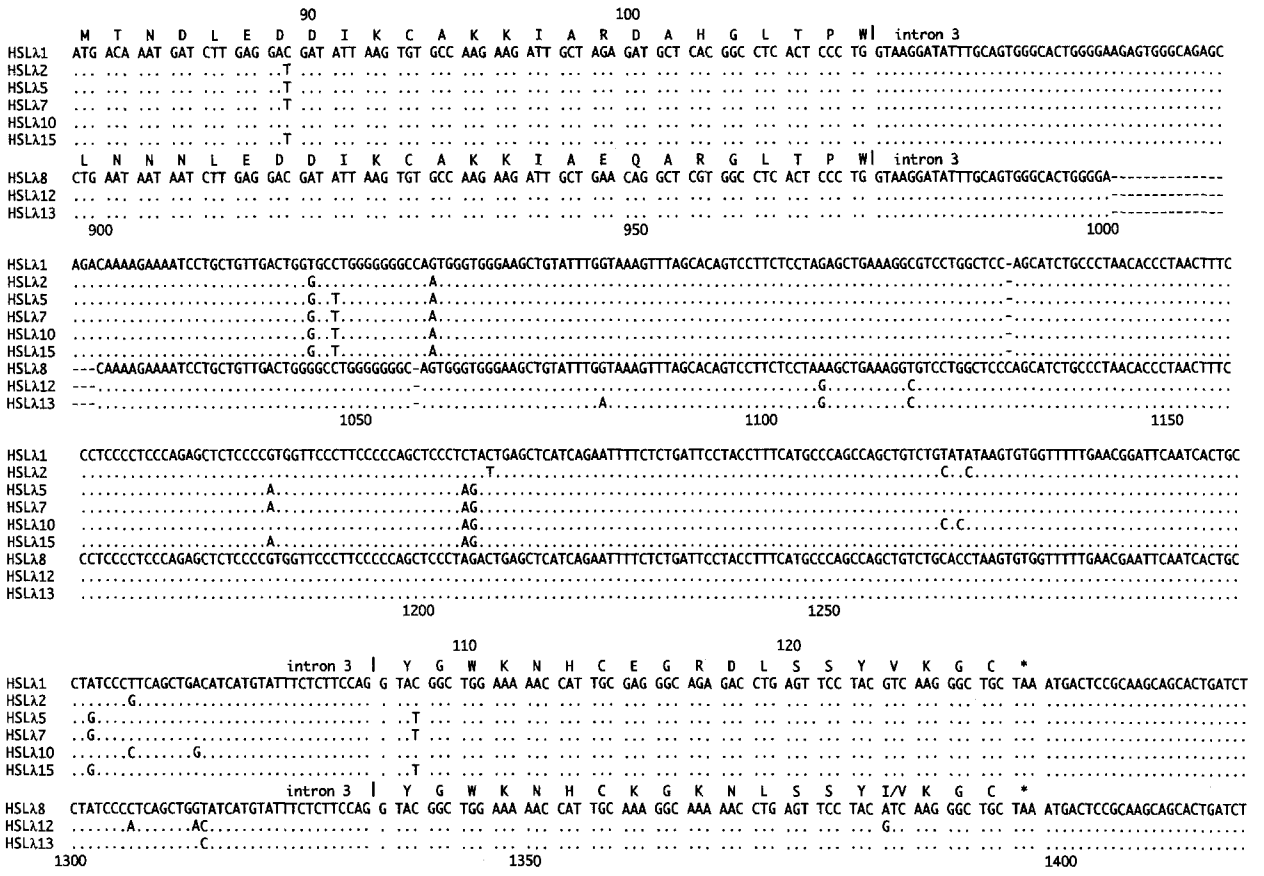

Fig. 2. Continued.

(Irwin et al. 1996). Although the molecular genetics of the calcium-binding lysozymes have not been established, there is no reason, *a priori,* to assume a significantly smaller gene for a putative hoatzin lysozyme gene. However, it is possible that some or even all calcium-binding lysozyme genes are characteristically short compared to those encoding the conventional lysozymes.[2]

Two Classes of Predicted Proteins

Among these nine sequences are two classes, based on their predicted amino acid sequences: One group of six sequences representing a minimum of three genes that correspond to the stomach lysozyme amino acid sequences, and a second group of three sequences representing at least two loci encoding lysozymes that are more basic in character. As indicated in Fig. 2, there are only two slightly different proteins encoded by the six stomach sequences, and they differ by only a single amino acid at position 4. Both of these species appear to be expressed in the stomach (Kornegay et al 1994).[3] The three "basic" lysozyme gene sequences correspond to three distinct predicted proteins, which differ from one another by amino acid replacements at two to five positions. In contrast, pairwise amino acid differences between representatives of the different classes range from 23 to 26.

Discussion

Evolutionary Relationships to Other Vertebrate Lysozymes

When the amino acid sequences for these five putative hoatzin lysozymes are aligned with pigeon lysozyme

[2] Irwin et al. (1996) point out that the genes for α-lactalbumin (which has long been recognized as sharing a common ancestor with lysozyme *c*) are short—only 2–2.5 kb, with most introns ranging from 301 to 507 bp. A model for the evolutionary relationships among members of the lysozyme *c* superfamily that involves two ancient calcium-binding lysozyme *c* lineages is presented by Prager and Jollès (1996); this model postulates the origin via gene duplication of α-lactalbumin along the calcium-binding lysozyme lineage that also gave rise to the hoatzin lysozymes. It is tempting to speculate that the observation of short genes for hoatzin lysozymes and for α-lactalbumins provides additional support for this evolutionary model. Sequences of other genes encoding calcium-binding lysozymes *c,* notably those representing the second deep lineage, should permit further assessment of this possibility.

[3] The genomic sequences presented here are in complete agreement with the previously reported cDNA sequence, with the exception of a very short region in the leader sequence. Reexamination of the cDNA sequence indicates that this region is prone to compressions during direct DNA sequencing of that fragment. The genomic sequences presented here reflect the actual signal peptide of 20 amino acids, thereby correcting an error in the 19-residue signal peptide sequence reported earlier (Kornegay et al. 1994).

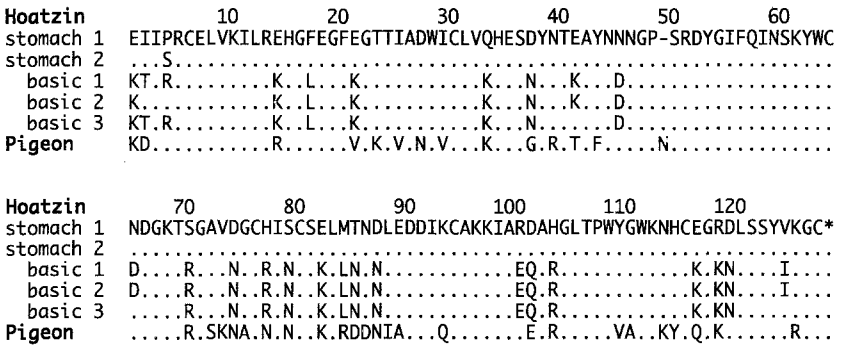

Fig. 3. Predicted amino acid sequences for all six hoatzin lysozyme genes aligned with pigeon egg-white lysozyme. Amino acids are represented by the *single-letter code. Dots* indicate identity to the residues in hoatzin stomach 1 lysozyme. Amino acid *numbering* is shown *above*. The presence of a deletion in the hoatzin lysozymes at position 49 of this alignment is indicated with a *dash* and accounts for the discrepancy between the numbers shown here compared to those in Fig. 2.

(Fig. 3), it is apparent that there is greater similarity between the basic hoatzin sequences and the pigeon egg-white lysozyme than between the hoatzin stomach and pigeon sequences. There are nine amino acid replacements between the hoatzin stomach and basic classes that the basic class shares with the pigeon sequence, and nearly all of these result in a change to a more basic residue such as arginine or lysine. In comparison, there are four amino acid replacements between the two hoatzin classes that the stomach class shares with the pigeon sequence, and only one of those results in a more basic residue. Overall, there are 34–36 differences (72% amino acid identity, on average) between the pigeon and hoatzin basic lysozyme sequences, compared to 38–39 differences (69–70% identity) between the pigeon and hoatzin stomach enzymes. The two hoatzin classes are, however, most similar to each other, with an average shared amino acid identity of 81%. This is supported by evolutionary analyses (Kornegay et al. 1994 and others not shown) in which the predicted protein sequences for the hoatzin lysozymes were aligned to pigeon and other vertebrate lysozymes and subjected to parsimony analysis. Consistent with the tree shown in Fig. 4, the hoatzin amino acid sequences cluster together, with the pigeon egg-white lysozyme coming out as the expected sister group.

Genetic Variation Among Hoatzin Lysozyme Genes

The three predicted amino acid sequences in the "basic" class of lysozymes are more variable than those of the stomach class, with at least two replacement changes between any two sequences, and five variable codon positions. This is in contrast to a single variable codon position among six different sequences in the stomach group. Conversely, when one considers the total amount of nucleic acid sequence variation in these two classes, the stomach lysozyme genes are actually more divergent. The pairwise distances between the HSLλ genomic clones are summarized in Table 2. The average observed difference between the stomach clones and the "basic" clones is 4.7%. The average observed divergence among the six stomach sequences is 1.3%, while the average divergence observed among the three basic sequences is roughly half that, or 0.67%. At least three possible explanations can account for the difference in divergence between the two classes. One possibility is that the initial stomach lysozyme gene underwent further duplication events at an earlier time than did the basic lysozyme gene (see Fig. 4). This model assumes that there was an initial duplication event giving rise to the two classes of lysozyme genes in the hoatzin, followed by additional duplication events later on within each class. The second possibility is that the two classes of lysozymes simply lie in different regions of the genome; it appears that genomic sequences may vary in mutation rate by as much as two-fold (Wolfe et al. 1989). Lastly, it may just be a reflection of the different "sample sizes" for the two classes. Since some of the intraclass pairwise divergence values overlap between the two classes, this possibility cannot be ruled out.

The distribution of changes found in the stomach lysozymes in comparison with the basic sequences may be of greater interest than the overall difference in the level of variation. Table 3 summarizes the number of synony-

Table 2. Pairwise differences in percent among HSLλ genomic sequences encoding hoatzin lysozyme genes[a]

Sequences compared	Stomach genes						Basic genes		
	1	2	5	7	10	15	8	12	13
HSLλ1	—	1.7	1.8	1.7	1.4	1.2			
HSLλ2		—	1.8	1.0	1.7	1.6			
HSLλ5			—	0.8	1.2	0.7			
HSLλ7				—	1.5	0.7			
HSLλ10					—	1.0			
HSLλ15						—			
HSLλ8	4.9	4.8	5.4	4.8	4.5	4.9	—	0.8	0.4
HSLλ12	4.4	4.5	4.9	4.6	4.3	4.4		—	0.8
HSLλ13	4.6	4.5	5.1	4.6	4.2	4.8			—

[a] Intraclass comparisons among the stomach and basic sequences are shown above the diagonal, and interclass comparisons are shown below. The tabulated values were computed over the entire sequences shown in Fig. 2 using the computer program PAUP 3.1.1. All base substitutions were weighted equally; insertions and deletions were taken to be equivalent to base substitutions; the 17-bp deletion in intron 3 of the basic genes was counted as a single change

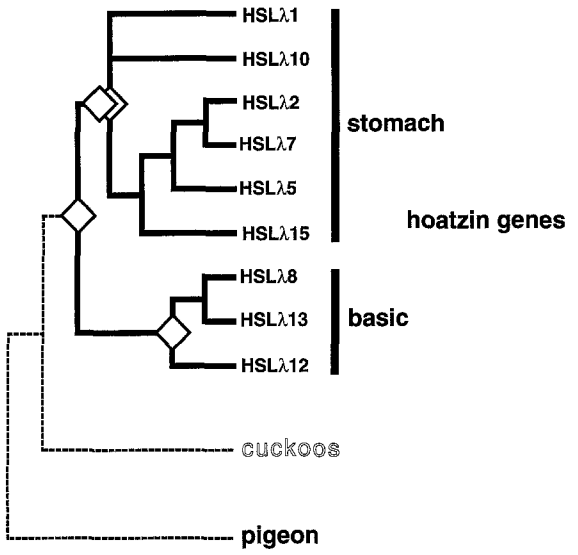

Fig. 4. Evolutionary relationships among the hoatzin lysozyme nucleotide sequences. The network drawn in *solid lines* shows the relationships of the hoatzin genes in a network generated by a parsimony analysis. This is a strict consensus of two equally parsimonious trees with a total length of 117. The presumed relationships of cuckoo and pigeon lysozymes are drawn with *dashed lines* to illustrate schematically the temporal placement of gene duplication events leading to the multiple lysozyme genes in the hoatzin. The branch separating the hoatzin stomach lysozyme genes from the more basic lysozyme genes was supported in bootstrapping analysis at a level of 100%. Other bootstrap values were lower than 70% with the exception of the grouping of HSLλ8 and HSLλ13 at 80% support. Putative gene duplication events are indicated with *open diamonds* and are presumed to have occurred in the hoatzin after this lineage diverged from other avian lineages (see text for discussion). There have been at least two duplications among the stomach lysozyme sequences; however, these events cannot be placed on nodes within that group because the allelic liaisons among those sequences could not be established. The various possible allelic combinations among the HSLλ1, HSLλ5, HSLλ10, and HSLλ15 sequences were tested by rearranging the branches using MacClade (Maddison and Maddison 1992). Any coupling among these four sequences into possible allelic loci results in a tree that is at least three steps longer than the one shown here. These differences in tree length are not statistically significant, but they raise the possibility that some of these sequences may represent different genes for which only a single allele was detected. In summary, there are at least three genes encoding the stomach lysozymes, but there may be as many as five or six genes represented among these sequences.

mous and noncoding variant positions vs the number of nonsynonymous variant positions within each of the two classes of sequences, as well as those observed between the sequences for each class. The sequence alignment in Fig. 2 reveals 37 noncoding and two silent variant positions compared to a single replacement site among the six stomach lysozyme sequences. A high ratio of synonymous to nonsynonymous changes suggests that strong purifying selection is operating at the protein-coding level (Li and Graur 1991). The total number of mutational events in the coding regions of the hoatzin lysozyme genes is too small to conduct a significant test of this ratio. However, if one includes the presumably neutral changes in the noncoding regions, the ratio observed for the stomach lysozyme intraclass comparison

Table 3. Synonymous and noncoding variant positions vs nonsynonymous variant positions among HSLλ sequences[a]

	Variant sites			
Comparison	Noncoding (NC)	Silent (S)	Replacement (R)	Ratio (NC + S)/(R)
Intrastomach	37	2	1	39.0
Intrabasic	9	0	5	1.8
Interclass	54	9	30	2.1

[a] The 17-bp deletion within intron 3 of the basic lysozyme genes was omitted from the interclass comparison

is significantly higher than that of the basic intraclass or the interclass comparisons. This is consistent with the strong selective pressures being exerted on these lysozymes in the stomach environment (Kornegay et al. 1994). The three more basic lysozyme sequences contain only nine silent and noncoding variant positions and have five replacement sites, and this lower ratio of 1.8 implies less functional constraint on the class as a whole. It is possible, for example, that only one of the basic genes is actively expressed in hoatzin tissues, resulting in low selective pressures on additional copies. Or perhaps the conventional lysozyme *c* (or a lysozyme *g;* cf. Prager and Jollès 1996) is performing the more conventional functions in egg white and other tissues, so that the additional "calcium-binding" type *c* genes are not active at all, making them effectively pseudogenes. Either of these two scenarios would account for the observed changes, however, and it should be noted that the pattern of substitution noted for these basic sequences is not unusual among other lysozymes. Rather, it is the stomach lysozymes which appear to depart from a relatively neutral pattern of evolution to one which is indicative of purifying selection. A similar bias has been observed in ruminant stomach lysozyme genes (Irwin and Wilson, 1989), for which concerted evolutionary processes such as gene conversion appear to be active (Irwin and Wilson 1990; Irwin et al. 1992, 1993).

Temporal Estimate for Gene Duplication Events

The absence of a close phylogenetic relative to the hoatzin among other birds and a lack of definitive rates of mutation for the various types of sequences make it difficult to estimate the timing of gene duplication events leading to the various hoatzin lysozymes. However, it should be possible to ascertain relative times for these events based on evolutionary rates calculated for diverse genes of other organisms and the current view of hoatzin systematics. One assumption for this comparison is that changes at synonymous and intronic sites are relatively neutral and will therefore evolve in a clock-like fashion (Kimura 1983). The estimates for the rate of nucleotide divergence in such genomic DNA regions for vertebrates

vary from 0.3% to 1% per Myr (Wilson et al. 1987; Li et al. 1987; Li and Graur 1991; Satta et al. 1993). If only synonymous and noncoding sites are considered, the average divergence among hoatzin stomach lysozyme genes is 1.6%. One can infer that the duplication events leading to this group of genes therefore took place between 1.6 and 5.3 Myr ago. Among the more basic lysozyme gene sequences, the average pairwise divergence at silent plus noncoding sites is 0.5%, suggesting that the duplication occurred 0.5–1.7 Myr ago. To estimate the time of the duplication resulting in the stomach and basic classes of hoatzin lysozyme genes, one must consider all interclass pairwise distances for silent and noncoding sites. The average value is 3.5% divergence for these comparisons, which results in a window of 3.5–11.7 Myr ago for the initial duplication event among the hoatzin lysozyme genes.

It is reasonable to argue that all of these gene duplication events have occurred during the evolution of the hoatzin after this bird diverged from other modern birds, as shown in Fig. 4. Although the systematics of the hoatzin has long been an issue of debate, recent molecular genetic data indicate that the hoatzin is most closely related to the cuckoos (Hedges et al. 1995). This alliance is consistent with other observations, including various morphological data, behavioral similarities, and other molecular studies, including DNA-DNA hybridization (Sibley and Ahlquist 1990) and mitochondrial DNA cytochrome *b* sequence data (Avise et al. 1994). The fossil record for cuckoos is very sparse, and only a single fossil has been assigned to the hoatzin family, Opisthocomidae. This is the incomplete skull of *Hoazinoides magdelenae* from the middle Miocene (~11–16.5 Myr ago) of Colombia (Miller 1953). Nonetheless, it is generally accepted that the modern hoatzin lineage represents a relatively ancient divergence from other groups of birds (see Olson 1985; Avise et al. 1994). One view (C.G. Sibley, personal communication) is that the hoatzin diverged from the South American cuckoo lineages within a few Myr after the isolation of some common ancestor in South America. This isolation event could have occurred ~100 Myr ago, when South America separated from Africa, or it could have occurred later, after South America moved from close proximity with Australia and Antarctica some 50 Myr ago. As for all ordinal diversifications among birds, there are alternative possibilities to these biogeographical scenarios; nonetheless, the hoatzin can be conservatively estimated to have split from other cuckoos no later than ~25 Myr ago. The genetic divergence among the genes presented above suggests that the first lysozyme duplication on the hoatzin lineage occurred no earlier than 12 Myr ago, well after its divergence from other birds.

The proposal that the diversity observed among the hoatzin lysozymes has arisen in only 12 Myr may seem startling, given the relatively high number of amino acid differences between the classes of hoatzin lysozymes.[4] These observations are probably due at least in part to variations in the rate of amino acid sequence evolution along different lineages. Upon apportioning replacement changes onto a parsimony tree with the echidna calcium-binding lysozymes *c* as an outgroup, Prager (1996—see footnote *o* to Prager's Table 2) found that amino acid replacement occurred twice as fast along the lineages leading to the hoatzin stomach and basic lysozymes as along the lineage leading to the pigeon lysozyme. Furthermore (E.M. Prager, personal communication), the average rate along both hoatzin lysozyme lineages in the 12 Myr since the initial gene duplication was estimated to be more than twice as fast (1.14 changes per Myr) as the rate computed for the hoatzin lineage during the 38 Myr preceding this split (0.49 changes per Myr). In this comparison, where a date of 50 Myr ago is assumed for the organismal split, the rate along the pigeon lineage (0.33 changes per Myr) is comparable to the rate along the hoatzin lineage up to the point of the lysozyme gene expansion. The apparent episodic evolution as well as the calculated absolute and relative rates parallel the scenario for ruminant stomach lysozymes [see Irwin et al. (1992), Prager (1996), and references therein]. In the case of the hoatzin stomach lysozyme lineage, rapid sequence evolution with purifying selection was likely involved in remodeling lysozyme to function as a digestive enzyme, before the gene duplication(s) occurred. After the new function was established, there was almost no fixation of replacement mutations among these genes (see Table 3). In the case of the basic hoatzin lysozymes, rapid amino acid sequence change may have been involved in remodeling for one or more other purposes and/or be explainable by the loss of functional constraints.

Acknowledgments. I thank S.B. Hedges, D.M. Irwin, I. Matsumura, E.M. Prager, C.G. Sibley, A. Sidow, C.-B. Stewart, and W.K. Thomas for technical advice and discussion, and D.M. Irwin, T.J. White, and especially E.M. Prager for comments on the manuscript. This work received support from National Institutes of Health and National Science Foundation grants to the late A.C. Wilson and from a National Science Foundation predoctoral fellowship to the author. The nucleotide sequences reported in this paper have been submitted to the GenBank with accession numbers L77803–L77811.

References

Avise JC, Nelson WS, Sibley CG (1994) Why one-kilobase sequences from mitochondrial DNA fail to solve the hoatzin phylogenetic enigma. Mol Phylo Evol 3:175–184

Dautigny A, Prager EM, Pham-Dinh D, Jollès J, Pakdel F, Grinde B,

[4] There are about two-thirds as many amino acid differences between the classes of hoatzin lysozymes as between the hoatzin and pigeon enzymes, even though the pigeon–hoatzin organismal divergence is likely to have occurred at least 50–60 Myr ago.

Jollès P (1991) cDNA and amino acid sequences of rainbow trout (*Oncorhynchus mykiss*) lysozymes and their implications for the evolution of lysozyme and lactalbumin. J Mol Evol 32:187–198

Dobson DE, Prager EM, Wilson AC (1984) Stomach lysozymes of ruminants. I. Distribution and catalytic properties. J Biol Chem 259:11607–11616

Gilbert W (1978) Why genes in pieces? Nature 271:501

Grajal A, Strahl SD, Parra R, Dominguez MG, Neher A (1989) Foregut fermentation in the hoatzin, a neotropical leaf-eating bird. Science 245:1236–1238

Haldane JBS (1932) The causes of evolution. Harper & Row, New York

Hedges SB, Simmons MD, van Dijk MAM, Caspers G-J, de Jong WW, Sibley CG (1995) Phylogenetic relationships of the hoatzin, an enigmatic South American bird. Proc Natl Acad Sci USA 92: 11662–11665

Irwin DM (1995) Evolution of the bovine lysozyme gene family: changes in gene expression and reversion of function. J Mol Evol 41:299–312

Irwin DM, Sidow A, White RT, Wilson AC (1989) Multiple genes for ruminant lysozymes. In: Smith-Gill SJ, Sercarz EE (eds) The immune response to structurally defined proteins: the lysozyme model. Adenine Press, Schenectady, NY, pp 73–85

Irwin DM, Prager EM, Wilson AC (1992) Evolutionary genetics of ruminant lysozymes. Anim Genet 23:193–202

Irwin DM, White RT, Wilson AC (1993) Characterization of the cow stomach lysozyme genes: repetitive DNA and concerted evolution. J Mol Evol 37:355–366

Irwin DM, Yu M, Wen Y (1996) Isolation and characterization of vertebrate lysozyme genes. In: Jollès P (ed) Lysozymes: model enzymes in biochemistry and biology. Birkhäuser Verlag, Basel, pp 225–242

Irwin DM, Wilson AC (1989) Multiple cDNA sequences and the evolution of bovine stomach lysozyme. J Biol Chem 264:11387–11393

Irwin DM, Wilson AC (1990) Concerted evolution of ruminant stomach lysozymes. Characterization of lysozyme cDNA clones from sheep and deer. J Biol Chem 265:4944–4952

Jung A, Sippel AE, Grez M, Schütz G (1980) Exons encode functional and structural units of chicken lysozyme. Proc Natl Acad Sci USA 77:5759–5763

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Kornegay JR (1994) Evolution of avian lysozymes. PhD thesis, University of California, Berkeley

Kornegay JR, Schilling JW, Wilson AC (1994) Molecular adaptation of a leaf-eating bird: stomach lysozyme of the hoatzin. Mol Biol Evol 11:921–928

Li W-H, Graur D (1991) Fundamentals of molecular evolution. Sinauer, Sunderland, MA

Li W-H, Tanimura M, Sharp PM (1987) An evaluation of the molecular clock hypothesis using mammalian DNA sequences. J Mol Evol 25:330–342

Maddison WP, Maddison DR (1992) MacClade: analysis of phylogeny and character evolution, version 3.0. Sinauer, Sunderland, MA

Miller AH (1953) A fossil hoatzin from the Miocene of Colombia. Auk 70:484–489

Ohno S (1970) Evolution by gene duplication. Springer-Verlag, Berlin

Olson SL (1985) The fossil record of birds. Avian Biol 8:79–238

Pahud J-J, Widmer F (1982) Calf rennet lysozyme. Biochem J 201: 661–664

Prager EM (1996) Adaptive evolution of lysozyme: changes in amino acid sequence, regulation of expression, and gene number. In: Jollès P (ed) Lysozymes: model enzymes in biochemistry and biology. Birkhäuser Verlag, Basel, pp 323–346

Prager EM, Jollès, P (1996) Animal lysozymes *c* and *g:* an overview. In: Jollès P (ed) Lysozymes: model enzymes in biochemistry and biology. Birkhäuser Verlag, Basel, pp 9–31

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Satta Y, O'hUigin C, Takahata N, Klein J (1993) The synonymous substitution rate of the major histocompatibility complex loci in primates. Proc Natl Acad Sci USA 90:7480–7484

Sibley CG, Ahlquist JE (1990) Phylogeny and classification of birds. A study in molecular evolution. Yale University Press, New Haven, CT

Stewart C-B, Schilling JW, Wilson AC (1987) Adaptive evolution in the stomach lysozymes of foregut fermenters. Nature 330:401–404

Swanson KW, Irwin DM, Wilson AC (1991) Stomach lysozyme gene of the langur monkey: tests for convergence and positive selection. J Mol Evol 33:418–425

Swofford DL (1993) PAUP: phylogenetic analysis using parsimony, version 3.1.1. Illinois, Natural History Survey, Champaign, IL

Wilson AC, Ochman H, Prager EM (1987) Molecular time scale for evolution. Trends Genet 3:241–247

Wistow G (1993) Lens crystallins: gene recruitment and evolutionary dynamism. Trends Biochem Sci 18:301–306

Wolfe KH, Sharp PM, Li W-H (1989) Mutation rates differ among regions of the mammalian genome. Nature 337:283–285