J Mol Evol (1993) 37:426–434

Journal of Molecular Evolution

Evolution of Protamine P1 Genes in Primates

J.D. Retief,[1] R.J. Winkfein,[1] G.H. Dixon,[1] R. Adroer,[2] R. Queralt,[2] J. Ballabriga,[2] R. Oliva[2]

Department of Medical Biochemistry, Faculty of Medicine, The University of Calgary, Health Sciences Centre, 3330 Hospital Drive N.W., Calgary, Alberta T2N 4N1 Canada, and [2]Molecular Genetics Research Group, Faculty of Medicine, University of Barcelona, Diagonal 643, 08025 Barcelona, Spain

Received: 15 May 1992 / Revised: 7 October 1992

Abstract. Protamine P1 genes have been sequenced by PCR amplification and direct DNA sequencing from 9 primates representing 5 major families, *Cebidae* (new world monkeys), *Cercopithecidae* (old world monkeys), *Hylobatidae* (gibbons), *Pongidae* (gorilla, orangutan, and chimpanzee), and *Hominidae* (human). In this recently diverged group of primates these genes are clearly orthologous but very variable, both at the DNA level and in their expressed amino acid sequences. The rate of variation amongst the protamine P1s indicates that they are amongst the most rapidly diverging polypeptides studied. However, some regions are conserved both in primates and generally in other placental mammals. These are the 13 N-terminal residues (including a region of alternating serine and arginine residues (the motif SRSR, res. 10–13) susceptible to Ser phosphorylation), a tract of six Arg residues (res. 24–29) in the center of the molecule, and a six-residue region (RCCRRR, res. 39–44), consisting of a pair of cysteines flanked by arginines. Detailed consideration of nearest-neighbor matrices and trees based on maximum parsimony indicates that P1 genes from humans, gorillas, and chimpanzees are very similar. The amino acid and nucleotide differences between humans and gorillas are fewer than those between humans and chimpanzees. This finding is at variance with data from DNA-DNA hybridization and extensive globin and mitochondrial DNA sequences which place human and chimpanzee as closest relatives in the super family, *Hominoidea*. This may be related to the fact that protamine P1s are expressed in germ line rather than somatic cells. In contrast to the variability of the exon regions of the protamine P1 genes, the sequence of the single intron is highly conserved.

Key words: Primate — Evolution — Protamine — Polymerase chain reaction — Sperm proteins

Protamines are small, arginine-rich polypeptides which appear at late stages of spermatogenesis in many but not all animal, and some plant, species (Bloch 1969; Dixon et al. 1985; Kasinsky 1989; Oliva and Dixon 1991). In those species in which they occur, some teleost fish (Bloch 1969; Dixon et al. 1985), birds (Oliva and Dixon 1989; Oliva et al. 1989; Chiva et al. 1988), some reptiles (Chiva et al. 1989) and amphibians (Kasinsky 1989), and all mammals (Oliva and Dixon 1991), they are involved in a major reorganization of spermatid chromatin during which the preexisting basic nuclear proteins complexed with DNA are replaced, in most cases quantitatively, by the newly synthesized protamines (Oliva and Dixon 1991). The consequence of this reorganization is an extreme condensation of the sperm chromatin, leading to markedly tighter packing of the male DNA.

Two classes of protamine polypeptides are observed amongst vertebrates, the "true" protamines (Bloch 1969; Oliva and Dixon 1991), which have been characterized from teleost fish (Bloch 1969; Dixon et al. 1985), birds (Oliva and Dixon 1989; Oliva et al. 1989; Chiva et al. 1988), reptiles (Chiva

Correspondence to: G.H. Dixon

et al. 1989), and marsupial mammals (Winkfein et al. 1991; Corzett et al. 1990). These range in size from 30 to 60 residues and typically contain up to 67% Arg residues and a restricted range of other residues but always including Ser (Oliva and Dixon 1991). A very characteristic feature of this group is the absence of any Cys residues and, therefore, an inability of these protamines to form disulfide bridges in the sperm chromatin.

By contrast, the protamines of the placental mammals, although still as Arg-rich as the "true" protamines and 50–60 residues long, always contain 6–9 Cys residues, which, in the mature sperm nucleus, are involved in both intra- and intermolecular disulfide bridges (Oliva and Dixon 1991). This ability of placental mammalian protamines to form a three-dimensional network of intermolecular disulfide bridges confers additional stability to the sperm nucleus which can only be dissociated into DNA and free protamine by chemical (or enzymatic) cleavage involving disulfide interchange with added thiols.

Protamine genes have been cloned and sequenced from teleost (Oliva and Dixon 1991) and elasmobranch fish (Berlot-Picard et al. 1986), birds (Oliva and Dixon 1989; Oliva et al. 1989), marsupial mammals (Winkfein et al. 1991), and a variety of placental mammals (Oliva and Dixon 1990, 1991). When the sequences of these protamine genes (and their corresponding polypeptides) were compared (Oliva et al. 1989; Oliva and Dixon 1989, 1990, 1991), it became clear that the protamines were far more variable in sequence than any other chromatin proteins such as the five classes of histones, some of which (H3 and H4) are amongst the least variable proteins (and genes) known (Wilson et al. 1977). In fact, the protamine genes were so variable that probes from different orders of birds or mammals usually did not cross-hybridize (Oliva and Dixon 1991).

This property of variability indicated to us that a study of the protamine P1 genes from the three monophyletic simian primate groups, *Cebidae* (new world monkeys), *Cercopithecidae* (old world monkeys), and the *Hominoidea* (apes and humans) might shed some additional light on their evolutionary relationships. Much previous work with extensive sequences in the primate globin clusters (Wilson et al. 1977; Perrin-Pecontal et al. 1992) has provided very strong evidence for a human-chimpanzee-gorilla clade in the *Hominoidea* and narrowly excludes gorilla from a human-chimpanzee clade (Goodman et al. 1989; Perrin-Pecontal et al. 1992).

Using a polymerase chain reaction (PCR) technique with a series of oligonucleotide primers, derived from protamine P1 sequences relatively conserved between mouse (Johnson et al. 1988) and human (Krawetz et al. 1989) P1 protamines, it has proved possible to amplify and sequence P1 protamine genes from nine primate species including human (*Homo sapiens*), chimpanzee (*Pan troglodytes* and *Pan paniscus*), gorilla (*Gorilla gorilla*), orangutan (*Pongo pygmaeus*), gibbon (*Hylobates lar*), African green monkey (*Cercopithecus patas*), marmoset (*Saguinus imperator*), and red howler (*Alouatta seniculus*). Our protamine P1 data confirm the close relationship of the human, gorilla, and chimpanzee genes but in contrast with the globin (Wilson et al. 1977; Koop et al. 1986; Miyamoto et al. 1987; Goodman et al. 1989; Perrin-Pecontal et al. 1992) and mitochondrial DNA (Hayasaka et al. 1988; Hasegawa et al. 1990) data, the protamine P1 genes of human and gorilla appear slightly more closely related than the human-chimpanzee and as related as the gorilla-chimpanzee pair.

Protamine genes are expressed postmeiotically at the early spermatid stage of spermatogenesis (Oliva and Dixon 1991) and code for a protein whose function relates to the condensation and genetic inactivation of the sperm nucleus (Oliva and Dixon 1991). A study of their evolution thus provides data for a gene set quite different from the globin or mitochondrial genes of somatic cells and upon which the functional constraints are likely to be quite different.

Methods

Most primate protamine P1 sequences were obtained by PCR amplification as follows: 50-μl reactions contained 0.5 μM each of the protamine specific 5′ (TTACAGGTTGGCTGGCTC) and 3′(TTGACAGGTCGGCATTGTTC) synthetic oligonucleotide primers, 10 mM Tris/HCl (pH 8.3 at 25° C), 50 mM KCl, 1.5 mM $MgCl_2$, 0.45% each of Tween 20 and NP40 and 200 μM each of dATP, dGTP, dCTP, and dTTP. Primate DNAs (0.5–1 μg) and 2 U of Taq DNA polymerase (Cetus) were added, mixed, overlain with 50 μl paraffin oil and heated at 94° C for 4 min. Thirty cycles of PCR were performed in a Perkin Elmer-Cetus DNA Thermocycler using the following parameters: denaturation for 1 min at 94° C, primer annealing for 1 min at 58° C, and primer elongation for 1 min at 72° C. The *C. patas* protamine P1 sequence was obtained using the primers and conditions described in Queralt and Oliva (1991). The *G. gorilla* P1 sequence was obtained independently from two different specimens using both the above primers and the primers described in Queralt and Oliva (1991).

Reaction mixtures were separated on 1.8% (w/v) agarose (BRL) gels in 40 mM Tris-acetate, 1 mM EDTA buffer. Bands were isolated by electrophoresis on to DEAE cellulose paper strips (type NA45, Schleicher & Schuell) and eluted as per manufacturer instructions. The DNA solution was extracted with a mixture of phenol/chloroform/isoamyl alcohol (25:24:1), twice repeated. Three volumes of 95% ethanol were added and the DNA was precipitated at −70° C for 2 h. DNA was pelleted in a microfuge (12,000 rpm, 5 min at room temperature), washed twice with 70% ethanol, and dried in vacuo. DNA was dissolved

in sterile water at a concentration of ~50–100 ng/μl and then heated at 100° C for 5 min and quick cooled on ice prior to sequencing. The *C. patas* and *G. gorilla* amplification products were purified, end-repaired, and phosphorylated using the Double Gene Clean procedure (Bio 101), cloned into the EcoR7 site of Bluescript KS+, and ssDNA prepared using helper phage R408.

DNA Sequencing. The dsDNA cycle sequencing system (BRL) was used to generate sequence ladders. One to 2 μl of PCR-amplified primate protamine DNA (~100 ng) was added into the system (containing ^{32}P end-labeled 5′ or 3′ primers) according to the manufacturers' "Normal cycle protocol" except that an initial incubation at 94° C for 4 min was included before PCR extension of the primers to ensure complete denaturation. DNA sequence ladders were separated on 40 cm × 35 cm × 0.4 mm 6% (w/v) polyacrylamide gels with multiple sample loadings at different times. Gels were fixed in 10% (v/v) methanol–7.5% (v/v) acetic acid for 20 min, transferred to Whatman 3MM paper, dried under vacuum, and exposed at −70° C to Kodak XAR film. Overnight exposures were usually sufficient to read sequences without intensifying screens. Several independent clones from *C. patas* and *G. gorilla* were sequenced with sequenase (USB) and ^{35}S-ATP using manufacturers' instructions.

Convention for Counting Differences. Changes in amino acids and nucleotides are counted as follows: substitutions are scored individually while gaps are scored as single changes regardless of length.

Results

Strong bands of the correct size were generated from human (*Homo sapiens*), common chimpanzee (*Pan troglodytes*), pygmy chimpanzee (*Pan paniscus*), gorilla (*Gorilla gorilla*), orangutan (*Pongo pygmaeus*), gibbon (*Hylobates lar*), African green monkey (*Cercopithecus patas*), marmoset (*Saguinus imperator*), and red howler (*Alouatta seniculus*) using the PCR primers described in Methods. It was possible to sequence seven PCR products directly without cloning since each major PCR band was sequentially pure. The marmoset (*Saguinus imperator*) sequence has been determined previously (Queralt and Oliva 1991). The *C. patas* protamine P1 was obtained by sequencing several independent clones which all showed identical sequences. The *G. gorilla* sequence, in addition to being sequenced directly, was also sequenced from an independent specimen using the approach of cloning and sequencing of several independent clones. The sequences with these two different approaches from the two independent specimens were identical.

In Fig. 1, the amino acid coding sequences deduced from the open reading frames in Fig. 3 are aligned. At this level, it is clear that, even with such a small 54-residue polypeptide as protamine P1, there are several amino acid substitutions even within the very closely related human-chimpanzee-gorilla clade. For example, there are 9 amino acid

```
     1.......10.......20.......30.......40.......50....
Hum  ARYRCCRSQSRSR--YYRQRQRSRRRRRRSCQTRRRAMRCCRPR--YRPRCRRH
Ppa            --C    -     K Q    Q        R   S L R
Ptr            --C          K Q    Q        R   S T R
Gor            --C     T              Q     R   N L R K
Ora          Q --CC R   CH      C            R     L
Gib            --C  RG                             L R
Cer            --CC   R C       R RA         R     L     Y
Sag            --C    R G       T RR -   S   R    KLT    Y
How        R L SRC    P C         RRP-   S   R     L R   Y
Mus        K   --CR R R C       C C-    RR   R RS TI  KKY
```

Fig. 1. Alignment of the deduced amino acid sequences of the primate protamine P1s. Hum = Human, Ppa = pygmy chimpanzee (*Pan paniscus*), Ptr = common chimpanzee (*Pan troglodytes*), Gor = gorilla (*Gorilla gorilla*), Ora = orangutan (*Pongo pygmaeus*), Gib = gibbon (*Hylobates lar*), Cer = African green monkey (*Cercopithecus patas*), Sag = marmoset (*Saguinus imperator*), How = red howler (*Alouatta seniculus*), Mus-mouse. The underlined residue indicates the position of the intron.

changes between human/*P. paniscus* (16.7%), 8 between human/*P. troglodytes* (14.8%), 8 between human/gorilla (14.8%), and 6 between both *P. paniscus* and *P. troglodytes* and gorilla (11.1%). Curiously, however, gibbon and human P1 protamines are the most similar of the primate group, differing by only 5 substitutions (9.3%), although much independent evidence comparing other DNA sequences (Wilson et al. 1977; Koop et al. 1986; Miyamoto et al. 1987; Goodman et al. 1989; Perrin-Pecontal et al. 1992) (as well as the protamine P1 DNA sequence, see below) indicates that gibbon is clearly more distant evolutionarily from human than either gorilla or chimpanzee.

In order to derive a quantitative estimate for the rate of evolution of the protamine polypeptides, we have compared the human P1 sequence with a more distantly related mammal, mouse P1 (Johnson et al. 1988), between which there are 18 amino acid changes per 54 residues or 33/100 residues. (We count the RS insert in mouse as a single event.) The primate-rodent divergence is generally accepted to be 75 million years ago (Wilson et al. 1977), leading to a rate of 0.22 changes/100 sites/10^6 years [33/(75 × 2)]. If we correct for back mutations (Holmquist 1972 formula 16), the rate increases to 0.25 changes/100 sites/10^6 years. When the more recent divergence of the human/chimp/gorilla clade is considered to be 5 million years ago (Wilson et al. 1977), the rate appears much faster at 1.5 changes/100 sites/10^6 years. This is an extremely rapid rate of evolutionary change and is compared graphically in Fig. 2 with other proteins (Behe 1990) with widely different rates of change for which quantitative estimates are available (Wilson et al. 1977; Behe 1990). The protamine P1 rate calculated from the rodent/primate divergence (0.22 changes/100 sites/10^6 years) lies between that for the fibrinopeptides, generally considered to be the most rapidly diverging proteins, and myoglobin (Wilson et al. 1977; Behe 1990). However, the rate calculated within the human/chimp/gorilla clade is much higher even than

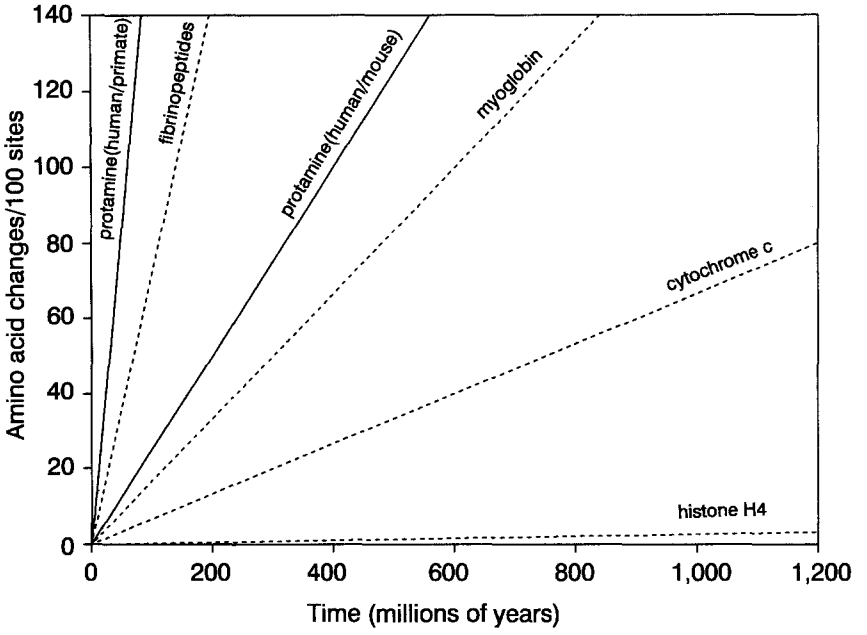

Fig. 2. Rates of amino acid change/100 sites for protamine P1 compared with other proteins after Behe (1990) using the data of Wilson et al. (1977). Protamine (human/mouse): the rate of amino acid change between human and mouse protamine P1 genes. Protamine (human/primate): the rate of change between humans and other primates. The higher rate of change is due to there being less chance of back mutation when changes are measured over a shorter time span. (See text.)

that for the fibrinopeptides at 1.6 changes/100 sites/ 10^6 years. Since there is more possibility of back mutations in the 75 million years since the divergence of the rodents and primates, it is likely that the higher rate within the human/gorilla/chimp clade may be more accurate.

The most conserved region of amino acid sequence is at the N-terminus where six protamines from human, chimpanzee, gorilla, orangutan, African green monkey, and marmoset are identical for the first 13 residues. Orangutan shows one R ↔ Q change at residue 11 and red howler has two changes Q ↔ R at residue 9 and R ↔ L at position 11. The mouse has one change Q ↔ K at position 9. The first four residues ARYR are also conserved in birds (Oliva and Dixon 1989, 1991). A very characteristic alternating sequence from residues 10–14 (SRSR) is part of this conserved region and is also seen in a number of other vertebrate protamines (Oliva and Dixon 1991). This serine-rich region has functional importance during the displacement of the somatic histones by protamine and the subsequent condensation of sperm nucleoprotamine by virtue of the apparently obligatory phosphorylation-dephosphorylation of the Ser residues during these processes (Oliva and Dixon 1991).

There is also strong conservation of a central tract (res. 24–29) of 6 Arg residues in human, gorilla, gibbon, all the monkeys and the mouse. The only changes are the 2 in the chimpanzees, residue 27 R ↔ K and residue R ↔ Q, and 1 in orangutan residue 24 R ↔ H. The only other strongly conserved region surrounds the pair of Cys residues in the sequence RCCRRR from residue 39–44 (this region is also the site of the intron which divides the Arg codon at position 39), but even here the Arg residue at position 43 is substituted by Pro in human and gibbon. The nature of the C-terminal residue neatly divides human and apes, which have His at this position, from the monkeys (and mouse) which have Tyr.

Finer discrimination between the orthologous P1 protamine gene sequences can be achieved by consideration of the alignment of the entire DNA sequence from position −54 to +361 in Fig. 3. The number refers to the human P1 gene (line 1) but differs for the other eight sequences as a result of gaps (deletions) introduced to maximize alignment. Identities of other primate sequences with human P1 are left blank; differences are indicated by the presence of an altered base and gaps by dashes. The sequences for marmoset (*Saguinus imperator*) and African green monkey (*Cercopithecus patas*) are slightly shorter than the others as a result of the use of different PCR primers (Queralt and Oliva 1991) and those 5′ and 3′ regions that were not determined are indicated by dots.

Within the total sequence is the coding region where evolutionary change is subject to constraints determined by the function of protamine P1 in condensing sperm DNA. At the mRNA level, the 5′ and 3′ untranslated regions also have constraints on function (Oliva and Dixon 1991; Kwon and Hecht 1991). However, we also have sequence information from the single intron which divides the coding region within codon 39 for Arg (A ↓ GG). It is generally considered that intron DNA is subject only to the functional constraints associated with maintenance of the splice sites at its termini (GT — — — — AG) together with a pyrimidine-rich region adjacent to the 3′ AG site (Mount 1982), although some introns with conserved controlling re-

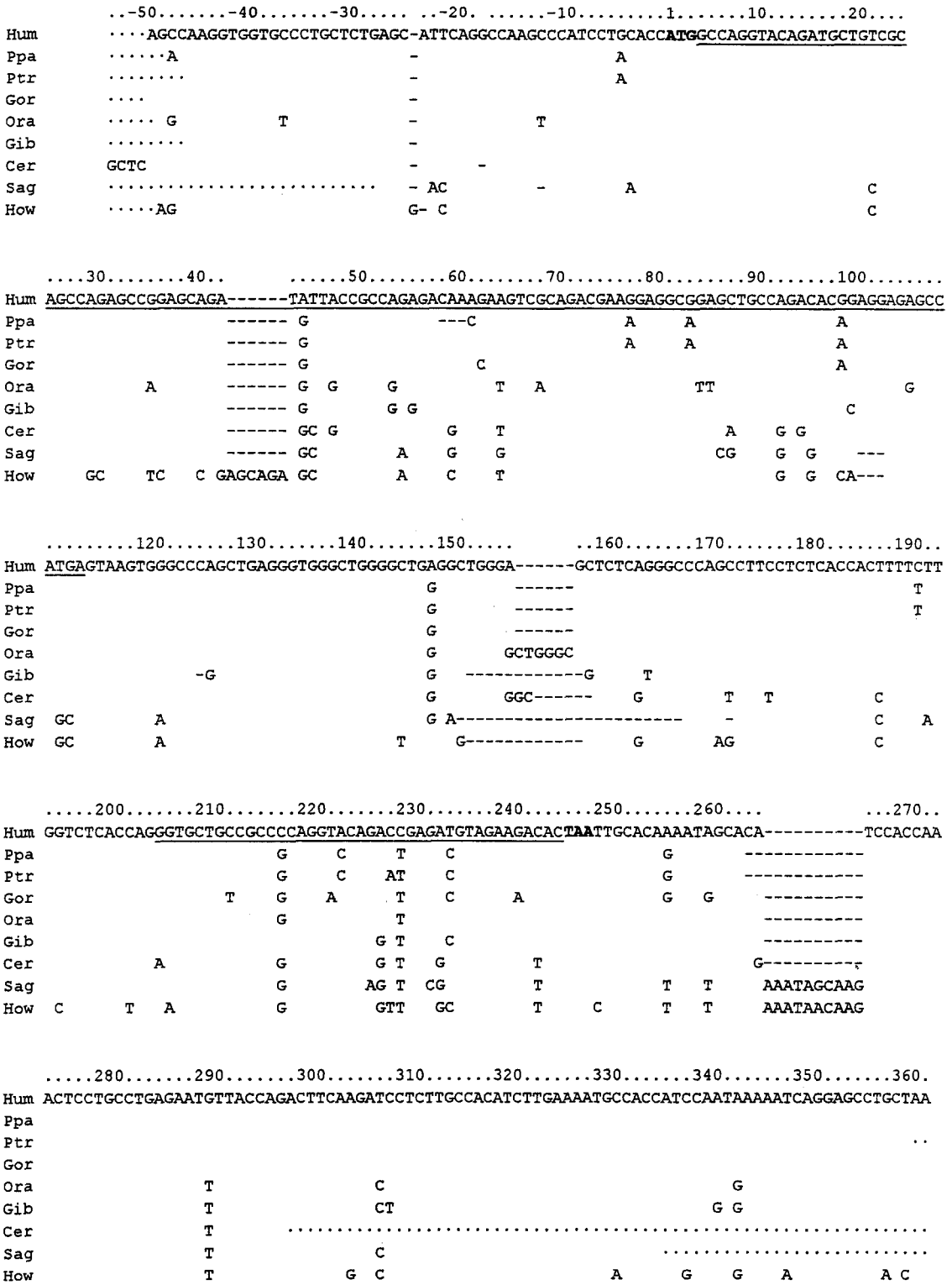

Fig. 3. Alignment of the DNA sequences of the primate protamine P1 genes. The start and the stop codons are *bold* and the coding areas *underlined*. The numbering is according to the human sequence and does not include gaps. The different sequence lengths, indicated by "•", where the sequence was not determined, reflect the various primers used in sequencing reactions. The alignments were adjusted to maintain the reading frames.

gions are known (Balvay et al. 1992; Vikkula et al. 1992). In this gene there are many fewer changes in the 91-bp introns of the human, chimpanzee, and gorilla clade than in their exons. The two species of chimpanzee have identical 91-bp introns while human/chimp shows only 2 base changes (2.2%), human/gorilla 1 change (1.1%), and chimp/gorilla 1 change (1.1%). For the intron this leads to an evolutionary rate of 0.11–0.22 changes in base pairs/100 sites/10^6 years as compared with a rate of 2–4 times

faster at 0.4 changes/100 sites/10^6 years for the exon regions. With the more distantly related apes, gibbon, and orangutan, base changes by substitution remain very few; human/orangutan 2 (2.1%), human/gibbon 4 (4.8%), but in each there are also more extensive addition/deletion events such that the gibbon intron decreases to 84 bp and the orangutan intron increases to 97 bp. With the new world monkeys, marmoset, and red howler, there are both an increased number of substitutions and deletions of larger size reducing the intron length to 73 bp for marmoset and to 85 bp for red howler.

A further interesting feature of the protamine P1 intron is the presence of a 5-bp repeat, GGGCT, in the 5′ region of the intron which occurs 3 times as GGGCT(135–139, 141–145, and 147–151), but additionally, there are 3 derivatives GGGCC (twice at 119–123 and 163–167) and GAGCT (154–158) differing by only 1 base and 2 others, CAGCT (124–128) and AGGGT (130–134), differing at 2 positions. Thus in a sequence region of 48 bp (119–167), there are seven 5-bp repeats comprising 73% of the sequence. This repetitive structure is well conserved within the group of nine primate P1 protamine genes considered here and may provide a clue to the origin of the single intron in the P1 gene of mammals. There is no intron in the protamine genes of fish (Dixon et al. 1985; Oliva and Dixon 1991) or birds (Oliva and Dixon 1989, 1990, 1991) and it seems likely that the single intron of mammals (both marsupial and placental) appeared late in vertebrate evolution, possibly as an expansion of this 5-bp motif.

Discussion

Protamine P1 Gene and Protein Structure

The nine P1 protamine gene sequences described above (Fig. 3) and which predict corresponding protamine polypeptide sequences (Fig. 1) are clearly closely enough related (93.8% identical) to be regarded as orthologous. However, the marked sequence variations both at the DNA and protein levels indicate that sperm P1 protamines are amongst the proteins that diverge most rapidly during evolution (Fig. 2). This would indicate that the functional constraints on the structure of protamine P1s are relatively relaxed as compared, for example, with the arginine-rich histones H3 and H4 which interact with each other to form the "kernel" tetramer (H3.H4)2 of the nucleosome (Richmond et al. 1984).

At the spermatid stage of spermatogenesis, in fish (Dixon et al. 1985; Oliva and Dixon 1991) and birds (Oliva and Dixon 1989, 1990, 1991), newly synthesized protamines, phosphorylated on one to three serine residues, displace the entire complement of nucleosomal histones (Oliva and Dixon 1991). As a result of this displacement, a new, much more compact and transcriptionally inactive form of chromatin, nucleoprotamine, is produced. During this final compaction, the phosphate groups on the displacing phosphoprotamine are removed (Oliva and Dixon 1991). The functional attributes required of protamine for this process appear to be its strongly basic charge due to the high Arg content and the presence of two phosphorylatable serine residues in the alternating SRSR sequence.

In placental mammals, the displacement of histones by protamine is more complex, involving an additional step by which a set of arginine-rich proteins, distinct from protamines, the transitional proteins (Heidaran et al. 1989; Luerssen et al. 1988), appear first to displace the somatic histones and then are, in turn, displaced by protamine P1 [and protamine P2 (Elsevier 1982; Bellvé et al. 1988; McKay et al. 1986; Yelick et al. 1987) in some primates and rodents] to form the final compact sperm nucleoprotamine.

Comparisons of P1 protamine from several orders of mammals (Oliva and Dixon 1991) as well as the primate P1s described here (Fig. 1) indicate that the major conserved amino acid sequences are the N-terminal 13 residues, the alternating SRSR region, the tract of six Arg residues near the middle of the molecule, and a region RCCRRR flanking a pair of Cys residues (res. 40 and 41) near the C-terminus. Outside these conserved regions, it appears that the protamine amino acid sequence has diverged rapidly in placental mammals (Oliva and Dixon 1991). This rapid divergence, even of the coding regions of the protamine P1 in an order as recently evolved as the primates, makes them interesting evolutionary indicators. These genes also provide an example of a set of genes expressed only in male germ cells which complements the extensive data available from somatically expressed genes such as the globins (Wilson et al. 1977; Koop et al. 1986; Miyamoto et al. 1987; Goodman et al. 1989; Perrin-Pecontal et al. 1992), immunoglobulins (Ueda et al. 1985, 1989), mitochondrial genes (Hayasaka et al. 1988; Hasegawa et al. 1990), or ribosomal RNA genes (Gonzalez et al. 1990).

Phylogeny of Higher Primates

An objective and quantitative comparison of the sequences in Fig. 3 has been done by the nearest-neighbor (Sattath and Tversky 1977) and maximum parsimony (Fitch 1971) methods. Figure 4A shows a

(A)

	Hum	Ppa	Ptr	Gor	Ora	Gib	Cer	Sag	How	Mus
Hum		14	13	12	19	16	22	31	66	177
Ppa	3.5		2	12	24	21	26	34	65	173
Ptr	3.2	0.5		11	23	21	26	35	64	171
Gor	2.9	3.0	2.7		23	20	26	33	67	176
Ora	4.6	5.9	5.7	5.6		19	25	32	66	171
Gib	4.0	5.3	5.3	5.0	4.7		22	31	63	165
Cer	6.4	7.7	7.7	7.6	7.3	6.6		23	55	125
Sag	9.1	10.1	10.4	9.7	9.4	9.1	7.6		33	132
How	16.5	16.5	16.3	16.7	16.6	15.9	16.4	9.5		156
Mus	45.2	44.8	44.2	44.8	43.1	43.2	37.3	39.4	39.7	

(B)

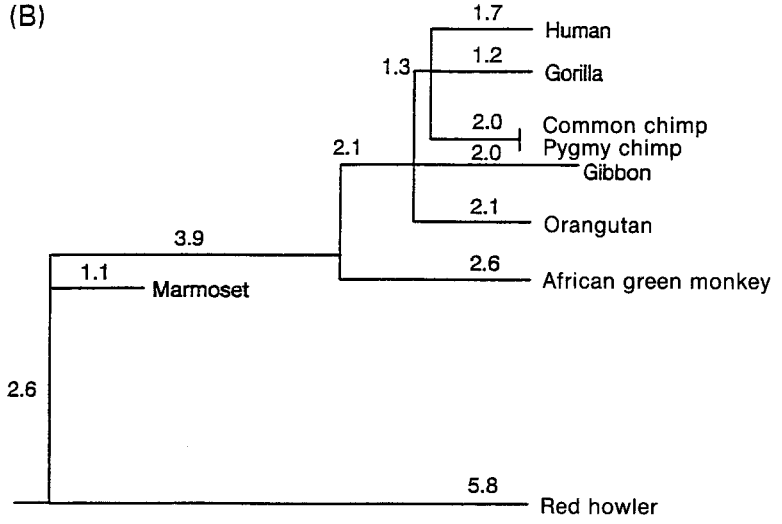

Fig. 4. **A** Distance matrix of protamine P1 DNA sequences. The total numbers of differences are indicated above the diagonal and the differences per 100 bp below. Gaps are not counted. **B** Rooted phylogenetic tree using the mutational distances from Fig. 4A. The tree was constructed by the two-step procedure of Li (1981), which corrects for unequal rates of evolution. The branch lengths were calculated according to Fitch and Margoliash (1967). The mouse was used as an outgroup only to root the tree. Only positive branch lengths were allowed.

nearest-neighbor matrix including the nine primate sequences and the published mouse sequence (Johnson et al. 1988). The matrix is divided into two regions by the diagonal. Above and to the right the total number of base-pair differences is shown while below and to the left of the diagonal, these differences are calculated as percentages (changes per 100 bp). From these data, it is possible to construct a rooted tree relating the nine primate protamine P1 gene sequences, Fig. 4B, giving the branching order and the length of the branches. The mouse outgroup was used only to root the tree. The most interesting result is within the very closely related human-gorilla-chimpanzee clade and indicates that human and gorilla P1 protamine genes with a total of the two branch lengths since divergence of 2.9 (1.7 + 1.2) are more closely related than human and chimpanzee with a corresponding branch length sum of 3.7 (2.0 + 1.7). Gorilla and chimpanzee P1s are intermediate at 3.2 (2.0 + 1.2).

We have also constructed a DNA parsimony tree, independent of any assumptions involving a molecular clock, by the method of Felsenstein (Felsenstein 1989) using the "DNABOOT" program of the PHYLIP 3.4 package. The results are illustrated in Fig. 5, which also places human closer to gorilla than either of the two chimpanzee species.

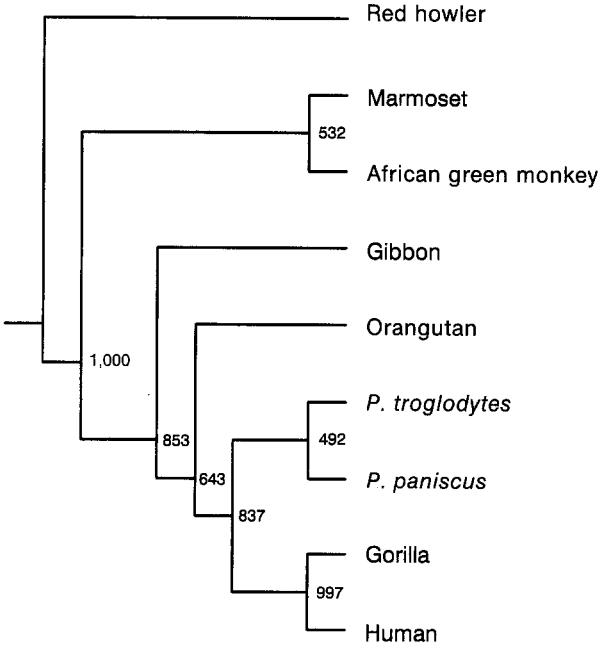

Fig. 5. A DNA parsimony tree constructed by the bootstrap method of the Felsenstein "DNABOOT" program in the PHYLIP 3.4 package (Felsenstein 1985, 1989). The data are randomly resampled and the numbers at the forks indicate the number of times the group occurred among the 1,000 bootstrap replicates. The branch lengths are arbitrary.

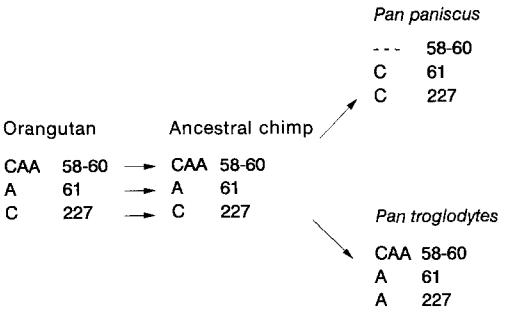

Fig. 6. A hypothetical scheme for the evolution of the protamine P1 gene from an ancestral gene to those of *P. paniscus* and *P. troglodytes*.

Within the P1 protamine genes of human, gorilla, and the two species of chimpanzee there are base changes at a total of 25 positions. In *P. paniscus* three of these changes result from the deletion of nucleotides 58–60 (Fig. 3), comprising the CAA codon for glutamine (Fig. 1 residue 21). We count this as a single event. There is also a second deletion, in both the chimpanzee species, of the two bases, CA present at positions 263–264 (Fig. 3) in human and gorilla P1s, which we also score as a single event. This brings the total number of changes within the group to 22. As would be expected, the two species of chimpanzee, *P. paniscus* and *P. troglodytes*, are most closely related but not identical. Using orangutan as an outgroup it is possible to construct a P1 sequence for a common ancestor of the two chimpanzee species. Thus an ancestral chimpanzee would be predicted to have CAA at position 58–60, A at position 61, and C at position 227. After the separation of *P. paniscus* and *P. troglodytes* from their common ancestor, CAA at 58–60 would be deleted and A mutated to C in *P. paniscus*. In *P. troglodytes* CAA 58–60 and A at 61 are retained, while C 227 is mutated to A (Fig. 6).

The phylogenetic branching pattern of the higher hominoids is still a matter of debate (Goodman et al. 1990; Holmquist et al. 1988; Ellis et al. 1990; Clemente et al. 1990) although extensive data from the β-globin locus (Koop et al. 1986; Miyamoto et al. 1987; Goodman et al. 1989, 1990, Perrin-Pecontal et al. 1992; Holmquist et al. 1988; Ellis et al. 1990), particularly of nontranscribed intergenic regions (thought to be free of functional constraints in their evolution), tends to a consensus favoring human and chimpanzee as slightly more closely related than human-gorilla although all three species are very closely related.

The data in the present paper confirm that human-gorilla-chimpanzee P1 protamines are indeed very similar but a consideration of both amino acid changes and DNA base changes and construction of phylogenetic trees by two independent methods, maximum parsimony and distance matrix, suggest that for these genes, the human-gorilla relationship is slightly favored. This view of the close relationship of gorilla and human genes is supported by two other gene sets that have been examined. Ueda et al. (1985) found a truncated immunoglobulin ϵ pseudogene in gorilla and man but not in chimpanzee and more recently, human leucocyte antigen (HLA) class II typing by Gyllensten and Ehrlich (1989) suggests that gorilla and human HLA-DXA-DQA1 loci are more closely related to each other than to chimpanzee.

It will be necessary to examine the evolutionary relationships between many more genes, expressed in both somatic and germ cells, before a final conclusion can be reached about the phylogenetic relationships amongst the higher primates.

Acknowledgments. We wish to thank Doug Nickel, Animal Genetics Laboratory, University of Calgary, and Drs. A.T. Garber, D.S. Hoar, and K. Tynan, Molecular Diagnostic Laboratory, Alberta Children's Hospital, Calgary, for kind gifts of various primate DNAs. We also thank Dr. J.L. Slightom, Division of Molecular Biology, The Upjohn Company, Kalamazoo, Michigan, for his gift of chimpanzee (*Pan troglodytes*) DNA. We are very grateful to Dr. Joseph Felsenstein, Department of Genetics SK-50, University of Washington, Seattle, for providing the PHYLIP program. We wish to thank Salvador Filella, Maria Jose Noto, and Justo Garasa from the Barcelona zoo for providing the *C. patas* and one of the *G. gorilla* samples. The work was supported by faculty position starting funds granted by the University of Barcelona to R. Oliva and by CICYT (Comision Interministerial de Ciencia y Tecnologia) and by the Medical Research Council of Canada.

References

Balvay I, Libri D, Gallego M, Fiszman MY (1992) Intronic sequence with both negative and positive effects on the regulation of alternative transcripts of the β tropomyosin transcripts. Nucleic Acids Res 20:3987–3992

Behe MJ (1990) Histone deletion mutants challenge the molecular clock hypothesis. Trends Biochem Sci 15:374–376

Bellvé AR, McKay DJ, Renaux BS, Dixon GH (1988) Purification and characterization of mouse protamines P1 and P2. Amino acid sequence of P2. Biochemistry 27:2890–2897

Berlot-Picard F, Vodjdani G, Doly J (1986) Nucleotide sequence of a cDNA clone encoding *Scylliorhinus caniculus* protamine Z2. Eur J Biochem 160:305–310

Bloch DP (1969) A catalog of sperm histones. Genetics Suppl 61:93–111

Chiva M, Kasinsky HE, Mann M, Subirana J (1988) On the diversity of sperm basic proteins in vertebrates: VI. Cytochemical and biochemical analysis in birds. J Exp Zool 254: 304–317

Chiva M, Kulak D, Kasinsky HE (1989) Sperm basic proteins in the turtle *Chrysemys picta:* characterization and evolutionary implications. J Exp Zool 249:329–333

Clemente IC, Ponsa M, Garcia M, Egozcue J (1990) Evolution of the Simiiformes and the phylogeny of human chromosomes. Hum Genet 84:493–506

Corzett M, Mazrimas JA, Cummins L, Lung B, Balhorn R (1990) Primary sequence of two protamines isolated from the bush-tailed possum, *Trichosurus vulpecula*. J Biol Chem 111:108a

Dixon GH, Aiken JM, Jankowski JM, McKenzie DI, Moir RD, States JC (1985) Organization and evolution of the protamine genes of Salmonid fishes. In: NATO Adv. Study Inst. Ser. 101. Plenum, New York/London, pp 287–314

Ellis N, Yen P, Neiswanger K, Shapiro LJ, Goodfellow PN (1990) Evolution of the pseudoautosomal boundary in Old World monkeys and great apes. Cell 63:977–986

Elsevier SM (1982) Messenger RNA encoding basic chromosomal proteins of mouse testis. Dev Biol 90:1–12

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–789

Felsenstein J (1989) PHYLIP—phylogeny inference package (version 3.2). Cladistics 5:164–166

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155:279–284

Fitch WM (1971) Toward defining the course of evolution: minimum change for a specific tree topology. Syst Zool 20:406–416

Gonzalez IL, Sylvester JE, Smith TF, Stambolian D, Schmickel RD (1990) Ribosomal RNA gene sequences and hominoid phylogeny. Mol Biol Evol 7:203–219

Goodman M, Koop BF, Czelusniak J, Fitch DH, Tagle DA, Slightom JL (1989) Molecular phylogeny of the family of apes and humans. Genome 31:316–335

Goodman M, Tagle DA, Fitch DH, Baily W, Czelusniak J, Koop BF, Benson P, Slightom JL (1990) Primate evolution at the DNA level and a classification of hominoids. J Mol Evol 30:260–266

Gyllensten UB, Erlich HA (1989) Ancient roots for polymorphism at the HLA-DQ alpha locus in primates. Proc Natl Acad Sci USA 86:9986–9990

Hasegawa M, Kishino H, Hayasaka K, Horai S (1990) Mitochondrial DNA evolution in primates: transition rate has been extremely low in the lemur. J Mol Evol 31:113–121

Hayasaka K, Gojobori T, Horai S (1988) Molecular phylogeny and evolution of primate mitochondrial DNA. Mol Biol Evol 5:626–644

Heidaran MA, Kozak CA, Kistler WS (1989) Nucleotide sequence of the Stp-1 gene coding for rat spermatid nuclear transition protein (TP1): homology with protamine P1 and assignment of the mouse Stp-1 gene to chromosome 1. Gene 75:39–46

Holmquist R, Miyamoto MM, Goodman M (1988) Higher-primate phylogeny—why can't we decide? Mol Biol Evol 5: 201–216

Holmquist R (1972) Theoretical foundations for a quantitative approach to paleogenetics. J Mol Evol 1:115–133

Johnson PA, Peschon JJ, Yelick PC, Palmiter RD, Hecht NB (1988) Sequence homologies in the mouse protamine 1 and 2 genes. Biochim Biophys Acta 950:45–53

Kasinsky HE (1989) Histones and other basic nuclear proteins. CRC Press, Boca Raton, FL

Koop BF, Goodman M, Xu P, Chan K, Slightom JL (1986) Primate beta-globin DNA sequences and man's place among the great apes. Nature 319:234–238

Krawetz SA, Herfort MH, Hamerton JL, Pon RT, Dixon GH (1989) Chromosomal localization and structure of the human P1 protamine gene. Genomics 5:639–645

Kwon YK, Hecht NB (1991) Cytoplasmic protein binding to highly conserved sequences in the 3′ untranslated region of mouse protamine 2 mRNA, a translationally regulated transcript of male germ cells. Proc Natl Acad Sci USA 88:3584–3588

Li W-H (1981) Simple method for constructing phylogenetic trees from distance matrices. Proc Natl Acad Sci USA 78: 1085–1089

Luerssen H, Hoyer-Fender S, Engel W (1988) The nucleotide sequence of human transition protein 1 cDNA. Nucleic Acids Res 16:7723–7723

McKay DJ, Renaux BS, Dixon GH (1986) Human sperm protamines. Amino-acid sequences of two forms of protamine P2. Eur J Biochem 156:5–8

Miyamoto MM, Slightom JL, Goodman M (1987) Phylogenetic relations of humans and African apes from DNA sequences in the psi beta-globin region. Science 238:369–373

Mount SM (1982) A catalogue of splice junction sequences. Nucleic Acids Res 10:459–472

Oliva R, Dixon GH (1989) Chicken protamine genes are intronless. The complete genomic sequence and organization of the two loci. J Biol Chem 264:12472–12481

Oliva R, Dixon GH (1990) Vertebrate protamine gene evolution I. Sequence alignments and gene structure. J Mol Evol 30: 333–346

Oliva R, Goren R, Dixon GH (1989) Quail (*Coturnix japonica*) protamine, full-length cDNA sequence, and the function and evolution of vertebrate protamines. J Biol Chem 264:17627–17630

Oliva R, Dixon GH (1991) Vertebrate protamine genes and the histone-to-protamine replacement reaction. Prog Nucleic Acid Res Mol Biol 40:25–94

Perrin-Pecontal P, Gouy M, Nigon V-M, Trabucket G (1992) Evolution of the primate β-globin gene reaction: nucleotide sequence of the δ-β-globin intergenic region of gorilla and phylogenetic relationships between African apes and man. J Mol Evol 34:17–30

Queralt R, Oliva R (1991) Protamine 1 gene sequence from the primate Saguinus imperator isolated with PCR using consensus oligonucleotides. Nucleic Acids Res 19:5786–5786

Richmond TJ, Finch JT, Rushton BM, Rhodes D, Klug A (1984) Structure of the nucleosome core particle at 7Å resolution. Nature 311:532–537

Sattath S, Tversky A (1977) Additive similarity trees. Psychometrica 42:319–345

Ueda S, Takenaka O, Honjo T (1985) A truncated immunoglobin ϵ pseudogene is found in gorilla and man but not in chimpanzee. Proc Natl Acad Sci USA 82:3712–3715

Ueda S, Matsuda F, Honjo T (1988) Multiple recombinational events in primate immunoglobulin epsilon and alpha genes suggest closer relationship of humans to chimpanzees than to gorillas. J Mol Evol 27:77–83

Ueda S, Watanabe Y, Saitou N, Omoto K, Hayashida H, Miyata T, Hisajima H, Honjo T (1989) Nucleotide sequences of immunoglobulin-epsilon pseudogenes in man and apes and their phylogenetic relationships. J Mol Biol 205:85–90

Vikkula M, Metsaranta M, Syvanen AC, Ala-Kokko L, Vuorio E, Peltonen L (1992) Structural analysis of the regulatory elements of the type-II procollagen gene. Conservation of promoter and first intron sequences between human and mouse. Biochem J 285:287–294

Williams SA, Goodman M (1989) A statistical test that supports a human/chimpanzee clade based on noncoding DNA sequence data. Mol Biol Evol 6:325–330

Wilson AC, Carlson SS, White TJ (1977) Biochemical evolution. Ann Rev Biochem 46:573–639

Winkfein RJ, Nishikawa S, Connor W, Oliva R, Dixon GH (1991) An opossum protamine gene. J Cell Biol 115:89a

Yelick PC, Balhorn R, Johnson PA, Corzett M, Mazrimas JA, Kleene KC, Hecht NB (1987) Mouse protamine 2 is synthesized as a precursor whereas mouse protamine 1 is not. Mol Cell Biol 7:2173–2179