Strand Symmetry Around the β-Globin Origin of Replication in Primates

M. Pilar Francino and Howard Ochman*†*

*Department of Biology, University of Rochester; and †Department of Ecology and Evolutionary Biology, University of Arizona

Certain mutations are known to occur with differing frequencies on the leading and lagging strands of DNA. The extent to which these mutational biases affect the sequences of higher eukaryotes has been difficult to ascertain because the positions of most replication origins are not known, making it impossible to distinguish between the leading and lagging strands. To resolve whether strand biases influence the evolution of primate sequences, we compared the substitution patterns in noncoding regions adjacent to an origin of replication identified within the β-globin complex. Although there was limited asymmetry around the β-globin origin of replication, patterns of substitutions do not support the existence of a mutational bias between the leading and lagging strands of chromosomal DNA replication in primates.

Introduction

Chromosomal DNA replication is an asymmetric process: on the leading strand it proceeds continuously, while on the lagging strand it proceeds discontinuously by the synthesis and joining of short Okazaki fragments (Kornberg and Baker 1992). In mammalian cells, three DNA polymerases (pol α, δ, and ϵ) are thought to participate in the replication of nuclear DNA, but the precise role of each has yet to be resolved. Pol α is probably restricted to replication initiation and priming of Okazaki fragments, whereas leading- and lagging-strand elongation could be accomplished by either pol δ alone (thereby restricting pol ϵ to Okazaki fragment maturation) or pol δ and pol ϵ, with each replicating one of the strands (Burgers 1998). In either case, and given the different processivities and error rates of the various polymerases (Kunkel 1992; Fijalkowska et al. 1998), the accuracy of lagging-strand replication may be affected by its more complex enzymatic requirements. Moreover, the replication fork is structurally asymmetric: leading-strand replication proceeds by the unwinding of very short templates prior to replication, whereas lagging-strand synthesis requires the exposure of long single-stranded regions, which can facilitate template-primer misalignments (Trinh and Sinden 1991; Kunkel 1992) and increase damage to the template strand.

Several experimental systems have revealed that the incidences of certain replication errors differ between the complementary DNA strands and that error rates are usually, but not always, higher on the lagging strand (Trinh and Sinden 1991; Veaute and Fuchs 1993; Roberts et al. 1994; Izuta, Roberts, and Kunkel 1995; Rosche, Trinh, and Sinden 1995; Thomas et al. 1996; Fijalkowska et al. 1998; Miret, Pessoa-Brandao, and Lahue 1998; Rosche, Ripley, and Sinden 1998). For bacteria, evidence for strand-specific mutational biases comes from the large-scale base composition patterns uncovered in complete genomic sequences. While equal frequencies of G and C, as well as A and T, are expected under no–strand bias conditions, most bacterial chromosomes exhibit deviations from these intrastrand parities, which often switch direction at the origin and terminus of replication (Lobry 1996; Mrazek and Karlin 1998). However, such deviations are also affected by strand-biased gene distribution coupled with (1) selection on amino acid content and/or synonymous codon usage and (2) strand-specific mutational biases introduced by transcription through transcription-coupled repair and the preferential deamination of the coding strand (Francino and Ochman 1997, 1999; Mrazek and Karlin 1998).

Unlike bacterial replication, DNA replication of eukaryotic chromosomes initiates at multiple origins, with an average spacing of 50–100 kb (Brewer and Fangman 1993). Eukaryotic chromosomes do not show large-scale deviations from intrastrand parity in base composition (Karlin, Campbell, and Mrazek 1998), although small local deviations are observable in the human β-globin region (Smithies et al. 1981; Bulmer 1991). This small-scale patterning might result from different mutational biases on the leading and lagging strands, but only if replicon size in germ line cells were restricted to a few kilobases (Bulmer 1991). Previous analyses of substitutional patterns in this region among primate species have yielded conflicting results regarding the existence of differences between the strands (Wu and Maeda 1987, but see Bulmer 1991). These were due, in part, to the lack of information about the actual positions of replication origins, which precluded the assignment of mutations to the leading or lagging strand. However, present knowledge of the location of a replication origin within the β-globin region (Kitsberg et al. 1993; Aladjem et al. 1998) enables the analysis of sequences for which the direction of replication can be established, thus facilitating the interpretation of such results.

To resolve whether mutational asymmetries exist between the leading and lagging strands of DNA, we compared the patterns of substitutions among primate species at noncoding sequences adjacent to the only origin of replication identified within the 200-kb β-globin region. The close proximity of the analyzed sequences to this replication origin minimizes the probability that these sequences are replicated by forks coming from al-

Key words: sequence evolution, substitutional asymmetry, strand bias, replication origin, primate globin genes.

Address for correspondence and reprints: Howard Ochman, Department of Ecology and Evolutionary Biology, 233 Life Sciences South, University of Arizona, Tucson, Arizona 85721. E-mail: hochman@u.arizona.edu.

Mol. Biol. Evol. 17(3):416–422. 2000

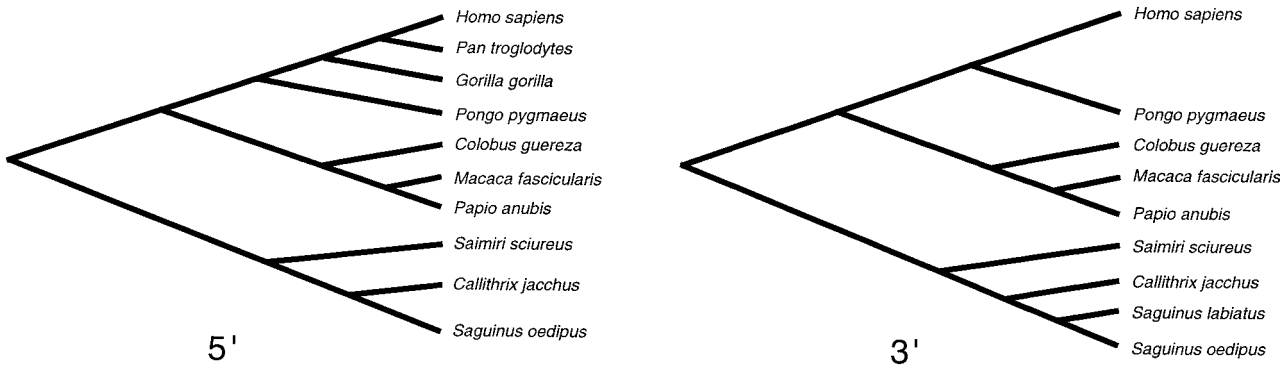

FIG. 1.—Phylogenies of the primate species examined for the regions 5′ (left) and 3′ (right) of the β-globin origin of replication.

ternative origins, even in face of dense origin recruitment in the germ line and in early embryogenesis, when the spacing between initiation sites is reduced to <10 kb (DePamphilis 1996). During these periods of rapid cell division, replication may initiate at many sites due to permissive conditions such as high concentrations of initiation proteins, but strong origins, like the one identified in the β-globin region, are still most likely to be used (DePamphilis 1996). Furthermore, several sequence motifs typical of eukaryotic replication origins are conserved at the same locations in the orthologous sequences from other primates, indicating that replication is very likely to initiate at the same site in all species considered.

Materials And Methods

Sequence Determination

The phylogenies of the species examined for the regions 5′ and 3′ of the β-globin origin of replication are shown in figure 1. The 5′ and 3′ sequences from the human (*Homo sapiens*), as well as the 5′ sequences from the chimpanzee (*Pan troglodytes*), the gorilla (*Gorilla gorilla*), and the cynomolgus monkey (*Macaca fascicularis*), were obtained from GenBank (accession numbers U01317, X02345, X61109, and X05665, respectively). All other sequences were amplified by the polymerase chain reaction (PCR) and sequenced for this study. Genomic DNA from the orangutan (*Pongo pygmaeus*), the cynomolgus monkey, the olive baboon (*Papio anubis*), the common marmoset (*Callithrix jacchus*), and the red-bellied tamarin (*Saguinus labiatus*) were provided by Todd Disotell (New York University), and genomic DNA from the black-and-white colobus (*Colobus guereza*) was provided by Caro-Beth Stewart (State University of New York at Albany). Blood samples from the squirrel monkey (*Saimiri sciureus*) and the cotton-top tamarin (*Saguinus oedipus*) were obtained from the New England Regional Primate Center (Harvard Medical School), and genomic DNA was isolated by means of the QIAamp Blood Kit (QIAGEN).

Figure 2 depicts the lengths and positions of the PCR-amplified and sequenced fragments. Alignments of GenBank sequences 5′ and 3′ of the β-globin origin were used in the design of the initial PCR primers. Sequences from the prosimian thick-tailed galago (*Galago crassicaudatus*, M73981) were included in these alignments but were avoided in subsequent analyses due to their high divergence from other orthologs. Purified PCR products (QIAquick PCR Purification or Gel Extraction Kits, QIAGEN) were used as templates for automated sequencing (ABI Prism BigDye Terminator Cycle Sequencing Kit, Applied Biosystems). Internal sequencing primers were designed as sequence data accumulated, until the PCR products were sequenced completely in both orientations. Primer design was aided by the OLIGO 4.01 software (National Biosciences, Plymouth, Minn.). Sequences were assembled and edited with SEQUENCHER 3.1RC12 (GeneCodes, Ann Arbor, Mich.).

Sequence Analysis and Statistical Treatment

Sequence alignments were performed with CLUSTAL X (Thompson et al. 1997), using default alignment parameters, and further refined by hand. Alignments

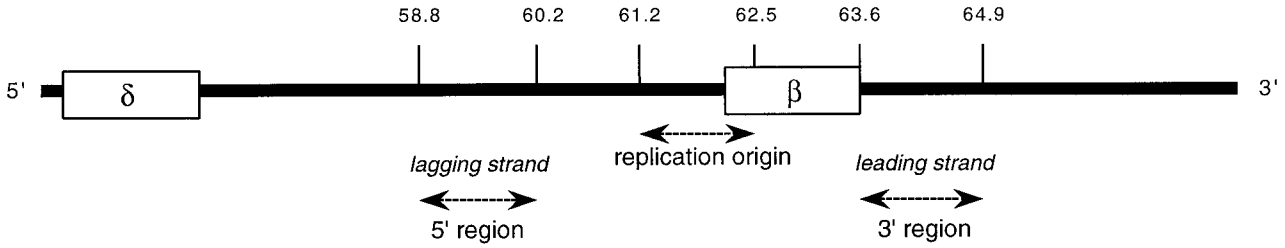

FIG. 2.—Lengths and positions of fragments analyzed 5′ and 3′ of the β-globin origin of replication. Numbers are in kilobases and correspond to the human sequence (GenBank accession number U01317). For New World monkeys, only half of the 5′ region (from kb 59.6 onward) was analyzed.

were then analyzed with the aid of MacClade 3.0 (Maddison and Maddison 1992) to reconstruct the most parsimonious ancestral states at the nodes of the species phylogenies and the directionality of the nucleotide substitutions. Minimal numbers of substitutions were used in subsequent analyses, i.e., substitutions that occurred with certainty but whose localizations within the tree topology may or may not have been determined. Minimal numbers of substitutions from nucleotide *i* to nucleotide *j* (n_{ij}, where *i* and *j* stand for T, C, A, or G and $i \neq j$) were converted to relative substitution frequencies, f_{ij} (Gojobori, Li, and Graur 1982; Imanishi and Gojobori 1992) by dividing by the average number of nucleotides *i* among all sequences and expressing the result as a percentage among all types of nucleotide substitutions.

In order to detect reversals in the substitutional patterns associated with the shift from lagging strand to leading strand, we compared the ratios of complementary substitutions (e.g., C→A vs. G→T) at either side (5′ and 3′) of the origin of replication. For every pair of complementary substitutions, we calculated the following ratio (ω):

$$\omega = \frac{f_{ij(5')}/f_{kl(5')}}{f_{ij(3')}/f_{kl(3')}},$$

where f_{ij} and f_{kl} are frequencies of complementary substitutions. If no strand bias existed, complementary substitutions would occur at equal frequencies within each region, and ω would be equal to 1. But under strand bias, complementary substitution frequencies would deviate from equality within each region, and the deviations would switch direction at the replication origin, yielding values of ω not equal to 1. To evaluate the statistical significance of ω values, we calculated ln ω, whose distribution is approximately normal and whose standard error can be obtained (Sokal and Rohlf 1995). Within each region, f_{ij} and f_{kl} were also compared by means of χ^2 tests. To evaluate the effects of neighboring nucleotides on the substitutional pattern, we obtained trinucleotide frequencies at each side of the origin with the aid of the GCG package (Devereux, Haberli, and Smithies 1984).

Results And Discussion

Sequence Features Around the β-Globin Origin of Replication

Two portions of the sequence around the origin were not included in the analysis of the substitutional pattern: a $(TG)_n$ repeat (Perrin-Pecontal et al. 1992) and a joint L1-Alu insertion.

The $(TG)_n$ repeat is located at the center of the 5′ region—nucleotides 678–731 in the 5′ alignment—in Old World monkeys (OWMs) and hominoids, with *n* varying from 10 in the chimpanzee to 27 in the gorilla. The orthologous segment could not be recovered from New World monkeys (NWMs). The sequence and length variability of this segment suggests that it evolves differently from the rest of the region due to such processes as slipped-strand mispairing.

Table 1
Minimal Numbers of Each Substitution Type in Regions 5′ and 3′ of the β-Globin Origin of Replication in Primates

	A	T	C	G
5′ region				
A	—	6	8	22
T	5	—	16	6
C	6	22	—	7
G	21	0	6	—
3′ region				
A	—	6	11	34
T	12	—	34	9
C	11	43	—	10
G	33	13	4	—

The NWMs harbor an insertion in this region of 634 bp—nucleotides 774–1407 in the 3′ alignment—consisting of a complete Alu element immediately preceded by the 3′ untranslated region (UTR) of L1. The joint L1-Alu insertion is flanked by perfect direct repeats (AGGACT). This arrangement suggests read-through transcription of an active L1 element into the adjacent Alu, coupled with retrotransposition by the L1-encoded enzymes into the 3′ region and extensive truncation of the 5′ end during cDNA synthesis, sparing only the 3′ UTR of L1 (H. Malik, personal communication). Therefore, this example can be added to the list of 3′ flanking DNA transductions mediated by L1 that have been detected in vivo (Miki et al. 1992; Holmes et al. 1994; McNaughton et al. 1997; Rozmahel et al. 1997) and represents the first case in which an Alu element has been transposed by the L1 machinery. The L1-Alu insertion occurred sometime between the split of NWMs from OWMs and hominoids 35 MYA and the split of *Saimiri* from the Callitrichinae (*Callithrix* and *Saguinus*) 17.5 MYA (Gingerich 1984; Fleagle 1988; Schneider et al. 1993). Because of the possibility of recombinational events between repetitive elements, we omitted the L1-Alu insertion from the estimation of the substitutional pattern in the 3′ region.

Patterns of Substitution Around the β-Globin Origin of Replication

The minimal numbers of substitutions detected were 125 in the region the 5′ and 220 in the region 3′ of the β-globin origin. Table 1 presents the numbers for every type of nucleotide substitution in either region, and figure 3 compares the frequencies (f_{ij}) for each pair of complementary substitutions. The hypothesis of unequal distribution of mutations between the leading and lagging strands predicts that complementary substitution frequencies will differ within each region and that deviations from equality will manifest opposite directions 5′ and 3′ of the replication origin. However, figure 3 shows that in most cases, complementary substitution frequencies are of similar magnitudes and/or that deviations from equality do not switch direction at the origin. Only two pairs of complementary transversions, T→A versus A→T and C→A versus G→T, behave in a manner similar to the strand-bias predictions.

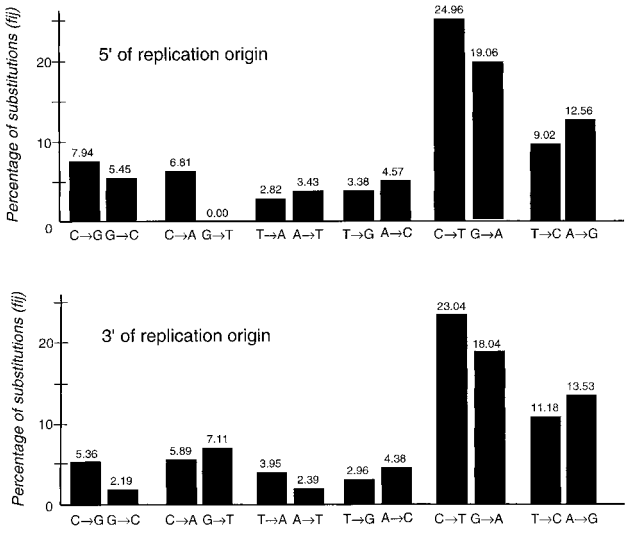

FIG. 3.—Substitutional patterns around the β-globin origin of replication. The frequencies (f_{ij}) for the members of each pair of complementary substitutions are graphed next to each other for ease of comparison.

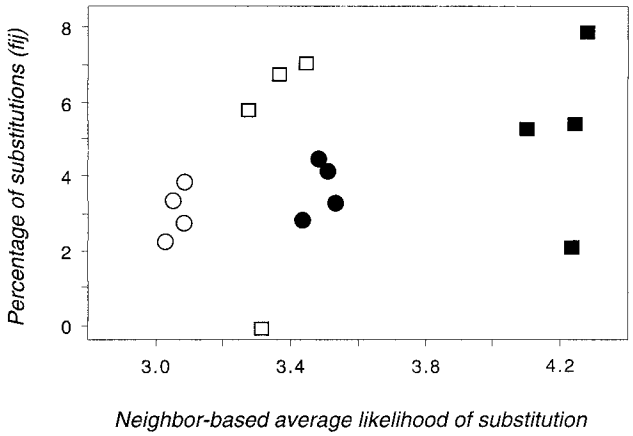

FIG. 4.—Neighboring-nucleotide effects on transversions on both sides of the β-globin origin of replication. Symbols are as follows: ○, T:A↔A:T; □, G:C↔A:T; ●, T:A↔G:C; ■, C:G↔G:C. For each transversion, the plotted likelihood is the average of the relative rate of substitution at every neighbor-pair environment (as reported by Hess, Blake, and Blake 1994) weighted by the frequency of that environment in the sequences, with relative rates normalized by the lowest rate of all substitutions (TAG(CTA)→TCG(CGA)).

The frequencies of T→A versus A→T differ in both regions, with A→T changes being more frequent upstream and T→A changes being more frequent downstream of the replication origin, but neither the differences within nor those between regions reach statistical significance. In the case of the C→A versus G→T pair, complementary frequencies differ significantly 5′ of the origin ($P = 0.032$), where no G→T changes occur, but they do not differ significantly 3′ of the origin, where G→T changes are more frequent than C→A changes. The pattern reversal at the origin is statistically significant ($\ln \omega = -2.679 \pm 1.156$; $P < 0.05$), although not complete. The directionality of transversions is opposite for the T→A versus A→T and the C→A versus G→T pairs: the purine→pyrimidine transversion is favored upstream of the origin in the first case, but downstream in the second. This counters simple explanations of strand bias, such as a higher mutation rate in one of the strands coupled with the differential incidence of purine : purine and pyrimidine : pyrimidine mismatches, which predict the same directionality in all deviations from equality between complementary transversions. Hence, there is no general pattern of asymmetry between the leading and lagging strands replicated from the β-globin origin.

Neighboring-Nucleotide Effects on the Substitutional Pattern

Because the pattern of substitutions around the β-globin origin of replication does not behave as expected from a strand-biased mutation input model, we investigated whether neighboring-nucleotide effects could explain the few individual cases in which deviations from symmetry do exist. The 5′ and 3′ neighboring nucleotides are known to influence the probability of specific substitutions (Hess, Blake, and Blake 1994), and, since nucleotides are likely to be exposed to different neighbors in each strand, substitutional asymmetries may arise. Based on the trinucleotide frequencies in the regions analyzed, we computed the average likelihood of each nucleotide substitution according to the relative rates reported by Hess, Blake, and Blake (1994) for every substitution in the environment of each of the 16 neighbor pairs. As shown in figure 4, the influence of neighboring nucleotides does not account for much of the variation in substitution frequencies or for the limited component of asymmetry around the origin. Indeed, the average likelihoods of substitution, given the neighboring environment, are similar for complementary transversion pairs within and between the 5′ and 3′ regions and do not covary with the observed f_{ij} (Spearman's coefficient $r_s = 0.374$, $P = 0.148$).

Base Composition Around the β-Globin Origin of Replication

Figure 5 illustrates the proportions of each base upstream and downstream of the β-globin origin and contrasts the observed values with those predicted from the reconstructed substitutional matrices (Sueoka 1995). The observed GC contents are in agreement with their expected values in both the 5′ and the 3′ regions (36% observed vs. 39% expected in 5′; 40% observed vs. 37% expected in 3′), and the proportions of all bases are close to equilibrium with the substitutional pattern, with 5′ G's and 3′ C's deviating most from expected frequencies. Most notably, there is no reversal of the base composition pattern at the replication origin, although the 5′ and 3′ relative proportions of A versus T and G versus C are not identical ($\ln \omega_{T/A} = -0.178 \pm 0.034$ and $\ln \omega_{C/G} = -0.243 \pm 0.044$, $P < 0.05$).

Therefore, examination of the base composition around the β-globin origin of replication confirms the findings based on analysis of the substitutional patterns within these regions. Even though slight differences are detected in composition and mutational input, they do

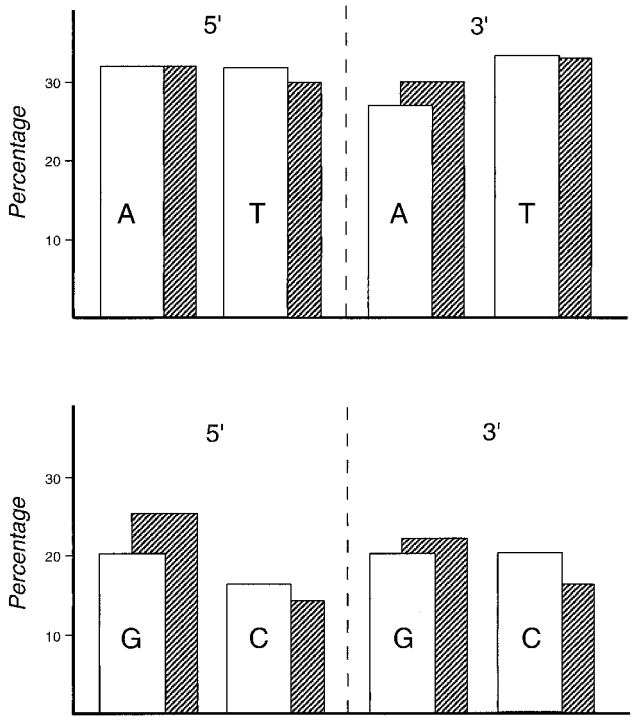

FIG. 5.—Base composition in regions upstream and downstream of the β-globin origin of replication. White bars represent observed values and hatched bars represent expected values at equilibrium with the reconstructed substitutional patterns (Sueoka 1995).

not support the existence of a mutational bias between the leading and lagging strands of chromosomal DNA replication in primates.

Origin Specificity in Multicellular Eukaryotes

Although site-specific origins of replication in multicellular eukaryotes have been difficult to localize, about 20 sites of replication initiation have now been identified. In particular, the β-globin origin of replication considered in the present study has been both biochemically (Kitsberg at al. 1993) and genetically (Aladjem et al. 1998) defined and shown to operate in both erythroid and nonerythroid cells. Moreover, although replication initiates at apparently random sites in the rapidly dividing early embryos of *Drosophila* and *Xenopus* (Shinomiya and Ina 1991; Hyrien, Maric, and Mechali 1995), this is not known to be the case in mammals, whose early embryogenesis differs from that of flies and frogs in several respects. First, mammal embryos divide at a much slower pace. For humans, the cleavage rate of embryos during their first week is only of one per day (England 1994). Furthermore, whereas no zygotic transcription takes place in *Drosophila* and *Xenopus,* transcription has been shown to begin immediately after zygotic formation in mouse embryos. In addition, consistent with the idea that active transcription and higher-order chromatin structure are involved in establishing origin specificity (DePamphilis 1996; Françon, Maiorano, and Mechali 1999), replication in mouse embryos requires specific origin and enhancer sequences (Wirak et al. 1985; Martinez-Salas, Cupo, and DePamphilis 1988).

The β-globin origin can initiate replication in different cell types independently of the transcriptional status of the β-globin gene cluster (Kitsberg at al. 1993; Aladjem et al. 1995), as well as upon experimental introduction at ectopic locations in the genomes of simian cell lines (Aladjem et al. 1998). Therefore, recruitment of the β-globin origin does not seem to require a specific chromatin organization, suggesting that this origin may function under varied chromosomal environments. Hence, we assume in our analysis that the β-globin origin is being preferentially used during the evolutionarily relevant cell divisions taking place in the germ line and in early embryos. Nevertheless, the possibility remains that the lack of strand biases around the β-globin origin is due to the utilization of alternative origins during such cell divisions. In particular, the β-globin replication initiation region may contain, in addition to the replication origin considered in this analysis, one or more secondary sites at which replication initiates with low efficiency (Aladjem et al. 1998). However, the frequency of firing of these putative origins would probably be too low to affect sequence evolution, unless relative origin efficiencies change in the germ line and/or in early embryos. Moreover, none of the potential origins lies upstream of the 5′ region analyzed in this study, and only sites with barely detectable activities reside downstream of the 3′ region, so the leading/lagging strand assignments here considered are not likely to vary. Therefore, the similarity of the patterns of substitution around the β-globin origin is likely to reflect the lack of strand biases in the input of mutations in the chromosomes of primates.

Acknowledgments

The authors thank Todd Disotell and Caro-Beth Stewart for DNA samples and Karen Gentile and Patrick O'Hara for assistance in procuring sequence data. Research supported through grants to H.O. from the National Institutes of Health. Sequence accession numbers are as follows: *P. pygmaeus,* AF205416 (5′) and AF205408 (3′); *C. guereza,* AF205417 (5′) and AF205409 (3′); *M. fascicularis,* AF205410 (3′); *P. anubis,* AF205418 (5′) and AF205411 (3′); *S. sciureus,* AF205419 (5′) and AF205412 (3′); *C. jacchus,* AF205420 (5′) and AF205413 (3′); *S. labiatus,* AF205414 (3′); *S. oedipus,* AF205421 (5′) and AF205415 (3′).

LITERATURE CITED

ALADJEM, M. I., M. GROUDINE, L. L. BRODY, E. S. DIEKEN, R. E. KEITH FOURNIER, G. M. WAHL, and E. M. EPNER. 1995. Participation of the human β-globin locus control region in initiation of DNA replication. Science **270**:815–819.

ALADJEM, M. I., L. W. RODEWALD, J. L. KOLMAN, and G. M. WAHL. 1998. Genetic dissection of a mammalian replicator in the human β-globin locus. Science **281**:1005–1009.

BREWER, B. J., and W. L. FANGMAN. 1993. Initiation at closely spaced replication origins in a yeast chromosome. Science **262**:1728–1731.

BULMER, M. 1991. Strand symmetry of mutation rates in the β-globin region. J. Mol. Evol. **33**:305–310.

BURGERS, P. M. 1998. Eukaryotic DNA polymerases in DNA replication and DNA repair. Chromosoma **107**:218–227.

DEPAMPHILIS, M. L. 1996. Origins of DNA replication. Pp. 45–86 *in* M. L. DEPAMPHILIS, ed. DNA replication in eukaryotic cells. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

DEVEREUX, J., P. HABERLI, and O. SMITHIES. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.

ENGLAND, M. A. 1994. The human. Pp. 207–220 *in* J. BARD, ed. Embryos. Color atlas of development. Wolfe, London.

FIJALKOWSKA, I. J., P. JONCZYK, M. M. TKACZYK, M. BIALOSKORSKA, and R. M. SCHAAPER. 1998. Unequal fidelity of leading strand and lagging strand DNA replication on the *Escherichia coli* chromosome. Proc. Natl. Acad. Sci. USA **95**:10020–10025.

FLEAGLE, J. G. 1988. Primate adaptation and evolution. Academic Press, San Diego.

FRANCINO, M. P., and H. OCHMAN. 1997. Strand asymmetries in DNA evolution. Trends Genet. **13**:240–245.

———. 1999. A comparative genomics approach to DNA asymmetry. Ann. N.Y. Acad. Sci. **870**:428–431.

FRANÇON, P., D. MAIORANO, and M. MECHALI. 1999. Initiation of DNA replication in eukaryotes: questioning the origin. FEBS Lett. **452**:87–91.

GINGERICH, P. D. 1984. Primate evolution: evidence from the fossil record, comparative morphology, and molecular biology. Yearb. Phys. Anthropol. **27**:57–72.

GOJOBORI, T., W.-H. LI, and D. GRAUR. 1982. Patterns of nucleotide substitution in pseudogenes and functional genes. J. Mol. Evol. **18**:360–369.

HESS, S. T., J. D. BLAKE, and R. D. BLAKE. 1994. Wide variations in neighbor-dependent substitution rates. J. Mol. Biol. **236**:1022–1033.

HOLMES, S. E., B. A. DOMBROSKI, C. M. KREBS, C. D. BOEHM, and H. H. KAZAZIAN JR. 1994. A new retrotransposable human L1 element from locus LRE2 on chromosome 1q produces a chimaeric insertion. Nat. Genet. **7**:143–148.

HYRIEN, O., C. MARIC, and M. MECHALI. 1995. Transition in specification of embryonic metazoan DNA replication origins. Science **270**:994–997.

IMANISHI, T., and T. GOJOBORI. 1992. Patterns of nucleotide substitutions inferred from the phylogenies of the class I major histocompatibility complex genes. J. Mol. Evol. **35**: 196–204.

IZUTA, S., J. D. ROBERTS, and T. A. KUNKEL. 1995. Replication error rates for G.dGTP, T.dGTP, and A.dGTP mispairs and evidence for differential proofreading by leading and lagging strand DNA replication complexes in human cells. J. Biol. Chem. **270**:2595–2600.

KARLIN, S., A. M. CAMPBELL, and J. MRAZEK. 1998. Comparative DNA analysis across diverse genomes. Annu. Rev. Genet. **32**:185–225.

KITSBERG, D., S. SELIG, I. KESHET, and H. CEDAR. 1993. Replication structure of the human β-globin gene domain. Nature **366**:588–590.

KORNBERG, A., and T. BAKER. 1992. DNA replication. Freeman, New York.

KUNKEL, T. A. 1992. Biological asymmetries and the fidelity of eukaryotic DNA replication. Bioessays **14**:303–308.

LOBRY, J. R. 1996. Asymmetric substitution patterns in the two DNA strands of bacteria. Mol. Biol. Evol. **13**:660–665.

MCNAUGHTON, J. C., G. HUGHES, W. A. JONES, P. A. STOCKWELL, H. J. KLAMUT, and G. B. PETERSEN. 1997. The evolution of an intron: analysis of a long, deletion-prone intron in the human dystrophin gene. Genomics **40**:294–304.

MADDISON, W. P., and D. R. MADDISON. 1992. MacClade, Version 3.0. Analysis of phylogeny and character evolution. Sinauer, Sunderland, Mass.

MARTINEZ-SALAS, E., D. Y. CUPO, and M. L. DEPAMPHILIS. 1988. The need for enhancers is acquired upon formation of a diploid nucleus during early mouse development. Genes Devel. **2**:1115–1126.

MIKI, Y., I. NISHISHO, A. HORII, Y. MIYOSHI, J. UTSONOMIYA, K. W. KINZLER, B. VOGELSTEIN, and Y. NAKAMURA. 1992. Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. Cancer Res. **52**:643–645.

MIRET, J. J., L. PESSOA-BRANDAO, and R. S. LAHUE. 1998. Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **95**:12438–12443.

MRAZEK, J., and S. KARLIN. 1998. Strand compositional asymmetry in bacterial and large viral genomes. Proc. Natl. Acad. Sci. USA **95**:3720–3725.

PERRIN-PECONTAL, P., M. GOUY, V.-M. NIGON, and G. TRABUCHET. 1992. Evolution of the primate β-globin gene region: 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.

ROBERTS, J. D., S. IZUTA, D. C. THOMAS, and T. A. KUNKEL. 1994. Mispair-, site-, and strand-specific error rates during simian virus 40 origin-dependent replication *in vitro* with excess deoxythymidine triphosphate. J. Biol. Chem. **269**: 1711–1717.

ROSCHE, W. A., L. S. RIPLEY, and R. R. SINDEN. 1998. Primer-template misalignments during leading strand DNA synthesis account for the most frequent spontaneous mutations in a quasipalindromic region in *Escherichia coli*. J. Mol. Biol. **284**:633–646.

ROSCHE, W. A., T. Q. TRINH, and R. R. SINDEN. 1995. Differential DNA secondary structure-mediated deletion mutation in the leading and lagging strands. J. Bacteriol. **177**:4385–4391.

ROZMAHEL, R., H. H. HENG, A. M. DUNCAN, X. M. SHI, J. M. ROMMENS, and L. C. TSUI. 1997. Amplification of CFTR exon 9 sequences to multiple locations in the human genome. Genomics **45**:554–561.

SCHNEIDER, H., M. P. C. SCHNEIDER, I. SAMPAIO, M. L. HARADA, M. STANHOPE, J. CZELUSNIAK, and M. GOODMAN. 1993. Molecular phylogeny of the New World monkeys (Platyrrhini, primates). Mol. Phylogenet. Evol. **2**:225–242.

SHINOMIYA, T., and S. INA. 1991. Analysis of chromosomal replicons in early embryos of *Drosophila melanogaster* by two-dimensional gel electrophoresis. Nucleic Acids Res. **19**: 3935–3941.

SMITHIES, O., W. R. ENGELS, J. R. DEVEREUX, J. L. SLIGHTOM, and S.-H. SHEN. 1981. Base substitutions, length differences, and DNA strand asymmetries in the human $^{G}\gamma$ and $^{G}\gamma$ fetal globin gene region. Cell **26**:345–353.

SOKAL, R. R., and F. J. ROHLF. 1995. Biometry. Freeman, New York.

SUEOKA, N. 1995. Intrastrand parity rules of DNA base composition and usage biases of synonymous codons. J. Mol. Evol. **40**:318–325.

THOMAS, D. C., D. L. SVOBODA, J. M. VOS, and T. A. KUNKEL. 1996. Strand specificity of mutagenic bypass replication of DNA containing psoralen monoadducts in a human cell extract. Mol. Cell. Biol. **16**: 2537–2544.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, and D. G. HIGGINS. 1997. The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment

aided by quality analysis tools. Nucleic Acids Res. **25**: 4876–4882.

TRINH, T. Q., and R. R. SINDEN. 1991. Preferential DNA secondary structure mutagenesis in the lagging strand of replication in *E. coli*. Nature **352**:544–547.

VEAUTE, X., and R. P. P. FUCHS. 1993. Greater susceptibility to mutations in lagging strand of DNA replication in *Escherichia coli* than in leading strand. Science **261**:598–600.

WIRAK, D. O., L. E. CHALIFOUR, P. M. WASSARMAN, W. J. MULLER, J. A. HASSELL, and M. L. DEPAMPHILIS. 1985. Sequence-dependent DNA replication in reimplantation mouse embryos. Mol. Cell. Biol. **160**:73–84.

WU, C.-I., and N. MAEDA. 1987. Inequality in mutation rates of the two strands of DNA. Nature **327**:169–170.

DAVID RAND, reviewing editor

Accepted November 18, 1999