A Genomic Region Evolving Toward Different GC Contents in Humans and Chimpanzees Indicates a Recent and Regionally Limited Shift in the Mutation Pattern

Ingo Ebersberger[1] and Matthias Meyer

Max-Planck-Institute for Evolutionary Anthropology, Leipzig, Germany

DNA sequences evolving differently in the human and chimpanzee genomes signal recent and regionally limited changes in the process of DNA sequence evolution. Here we present the comparison of 90 kb from the nonrecombining part of the human Y chromosome to the corresponding part of the chimpanzee genome using gorilla as out-group. Our results reveal a significant difference in the region-specific substitution process among the human and chimpanzee lineages. As a consequence, this region experiences a change in its GC content on the human lineage while it resides in compositional equilibrium on the chimpanzee lineage. Based on our analysis, we suggest a recent and species-specific shift in the region's mutation pattern as the cause of its differing evolution in humans and chimpanzees.

Introduction

The human genome comprises a mosaic as its base composition varies little within genomic regions—frequently referred to as "isochores" (Thiery, Macaya, and Bernardi 1976)—but differs considerably among them (Bernardi 2000; Lander et al. 2001). The mechanisms responsible for the formation of region-specific GC contents are still controversial and to date mainly two models are discussed. The model of variable mutation bias suggests a variation in the ratio of GC → AT and AT → GC mutation rates among genomic regions to account for their specific base compositions (Wolfe, Sharp, and Li 1989). Assuming neutrality (Kimura 1983), GC-poor genomic regions are then due to a mutation process that is biased toward GC → AT mutations while the reverse situation forms GC-rich regions. Alternatively, a regional variation in the fixation probabilities of GC → AT and AT → GC mutations has been proposed to explain the existence of isochores. Initially connected to the hypothesis that selective forces act to increase the GC content in functional regions of the human genome (Bernardi 2000), more recently, a regional variation in the fixation probabilities of mutations has been related to the process of recombination. In the model of biased gene conversion (BGC; reviewed in Marais 2003), it is proposed that heteroduplex formation during recombination causes mismatches to occur between allelic GC and AT sequence variants. These are believed to be—due to a biased mismatch repair process (Birdsell 2002)—repaired preferentially in favor of the GC allele. As a consequence, the fixation probability of AT sequence variants decreases with increasing recombination rate, resulting in an increase of the local GC content.

Meanwhile, supporting evidences have accumulated for both the model of regional mutation bias (Casane et al. 1997; Francino and Ochman 1999) and the model of BGC (Eyre-Walker 1993; Fullerton, BernardoCarvalho, and Clark 2001; Galtier et al. 2001; Smith and Eyre-Walker

2001). More recently, the argumentation in favor of BGC gained substantial momentum by the observations that there is evidence for (1) a fixation bias in favor of GC alleles in GC-rich regions of the human genome (Duret et al. 2002; Lercher et al. 2002) and (2) the strong correlation between the equilibrium GC content of a genomic region and its recombination rate (Meunier and Duret 2004). However, various studies have suggested that the recombination process itself is mutagenic (Lercher and Hurst 2002; Hellmann et al. 2003a), which raises the possibility that new AT → GC mutations are introduced during recombination. Consequently, BGC does not exclude the mutational model (Meunier and Duret 2004). Therefore, the criticism remains that none of the studies presented so far provide exclusive support for either model because the respective other—disfavored—model failed to be rejected (Eyre-Walker and Hurst 2001; Fullerton, BernardoCarvalho, and Clark 2001).

The present study attempts to readdress the question how the GC content of a genomic region is set and maintained. For this purpose, we compare the lineage-specific amount and pattern of DNA sequence change of 90 kb from the human Y chromosome located outside the pseudoautosomal region between humans and chimpanzees. Our results indicate a recent change in the region's substitution process resulting in a tendency toward differing equilibrium GC contents in humans and chimpanzees.

Materials and Methods
BAC Library Screen

Polymerase chain reaction (PCR) products of the second and seventh exon of the human *ZFY* gene were radioactively labeled with α-^{32}P-deoxycytosine triphosphate by random prime labeling and hybridized to the male chimpanzee bacterial artificial chromosome (BAC) library RPCI-43 (BACPAC Resources Center, Oakland, Calif.) according to the manufacturer's protocol. Filters were exposed for 2–72 h at −80°C.

DNA Amplification

Fragments up to 3 kb were amplified from 20 to 30 ng of genomic DNA. Longer fragments were amplified from 100 to 200 ng of genomic DNA using the Expand 20 kb PCR System (Roche Diagnostics, Mannheim, Germany). All primers are available upon request. PCRs were purified

using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany), separated on agarose gels, and stained with ethidium bromide. In the case of reactions displaying multiple bands, the band of interest was excised from the gel. Gel extraction was performed using the QIAquick Gel Extraction Kit (Qiagen) (fragment length < 10 kb) and the QIAquick Gel Extraction Kit II (Qiagen) (fragment length > 10 kb).

Preparation of DNA Shotgun Libraries

Template DNAs were mechanically fragmented using a nebulizer (Invitrogen, Karlsruhe, Germany). Fragments were separated on agarose gels, and the size fraction between 1 and 2.5 kb was extracted from the gel using QIAquick Gel Extraction Kits (Qiagen) without prior staining with ethidium bromide. Overhanging DNA fragment ends were filled in for 30 min at 20°C with T4 polymerase (12 U) and Klenow enzyme (12 U). DNA fragments were purified using QIAquick PCR Purification Kits (Qiagen), ligated into dephosphorylated, *Sma*I-cleaved pUC18, and introduced into *Escherichia coli* using electroporation.

Preparation of Plasmid DNA

Bacterial colonies were picked, and plasmids were isolated from 1.2 ml Luria broth overnight cultures using the QIAprep 96 Turbo BioRobot Kit (Qiagen). Plasmid solutions were incubated for 20 min at 80°C to remove remaining ethanol. Twenty microliters of H_2O was added after heat incubation to compensate for evaporated liquid.

DNA Sequencing

A total of 100–500 ng of plasmid DNA or 50–100 ng of purified PCR product was used as templates in sequencing reactions. Sequencing reactions were set up using ABI Prism Big Dye™ Cycle Sequencing Ready Reaction Kits (Applied Biosystems, Foster City, Calif.) with 10 pmol of sequencing primer. Following cycle sequencing, sequencing reactions were precipitated with isopropanol, dissolved in 25 µl H_2O, and analyzed on an ABI 3700 DNA sequencer (Applied Biosystems).

Contig Building and Quality Assessment

Overlapping DNA sequences were assembled with the Phred/Phrap package (http://www.phrap.org). In the gorilla, in one instance, a region resisted PCR amplification. The resulting two separate contigs were ordered and oriented by aligning them to the human and chimpanzee consensus sequences. By this, the gap size between the two contigs was estimated to be approximately 500 base pairs. Homopolymeric stretches with 10 or more repeats in DNA sequences determined from PCR-amplified templates showed differing lengths in different subclones, presumably due to slippage during PCR amplification. Because the exact length of such homopolymers could not be determined, we excluded all homopolymeric stretches with 10 or more repeats from the analysis. Base positions with a Phred score below 40 (Ewing and Green 1998) were masked in order to restrict the analysis to nucleotide positions determined with a high confidence. To cross-check the accuracy of the chimpanzee sequence, we randomly chose 100 positions identified as chimpanzee-specific substitutions and inspected the corresponding trace files. In all cases the called base for the chimpanzee was unambiguously supported by the corresponding chromatograms.

Sequence data from this article have been deposited with the European Molecular Biology Laboratory/GenBank Data Libraries under accession numbers AY913763–AY913765.

DNA Sequence Alignment and Analysis of DNA Sequence Differences

Multiple alignments of the human, chimpanzee, and gorilla DNA sequences were performed with the program "MultAlin" (Corpet 1988) using the default settings. The directions of nucleotide substitutions on the human and chimpanzee lineages were inferred by maximum parsimony with gorilla or baboon as out-group. When "non-CpG" sites were analyzed, all positions located in a CpG dinucleotide in at least one species were excluded.

Branch length estimations and clock test in a maximum likelihood framework were performed with the program Tree-Puzzle (Schmidt et al. 2002).

Estimation of the Equilibrium GC Content

To estimate the equilibrium GC content of a genomic region, we used three different approaches. (1) In the simple approach, the differences in the number of GC → AT substitutions ($N_{GC \to AT}$) and the number of AT → GC substitutions ($N_{AT \to GC}$) along a lineage were tested for significance using a two-tailed binomial test. The expected equilibrium GC content under the observed substitution pattern was then calculated as $GC_{equi} = 1/(1 + k)$, where $k = (N_{GC \to AT}/n_{GC})(N_{AT \to GC}/n_{AT})^{-1}$ and n_{GC} and n_{AT} are the numbers of analyzed GC and AT base pairs, respectively. (2) Alternatively, we applied the method described by Arndt, Petrov, and Hwa (2003), where the six-strand symmetric CpG-independent substitution rates and the transition rate at CpG dinucleotides are used to determine the equilibrium GC content of a DNA sequence. We estimated the substitution rates and the corresponding equilibrium GC content for the human and chimpanzee lineages, respectively, from the comparison of the extant DNA sequences to the ancestral sequence of humans and chimpanzees inferred by maximum parsimony. Parameter estimation and calculation of the equilibrium GC content was done with the tools provided at http://evogen.molgen.mpg.de/server/substitution_analysis. (3) We inferred the equilibrium GC content for the human ZFY region from the differences in the number of observed changes among the human and chimpanzee lineages. Assuming strand symmetry, we used the following modification of Felsenstein's (1981) neutral substitution model to describe the relation between substitution rate and equilibrium base frequency:

$$Q = \begin{pmatrix} & \alpha_1 \frac{\pi_{GC}}{2} & \beta_2 \frac{\pi_{GC}}{2} & \beta_4 \frac{\pi_{AT}}{2} \\ \alpha_2 \frac{\pi_{AT}}{2} & & \beta_3 \frac{\pi_{GC}}{2} & \beta_1 \frac{\pi_{AT}}{2} \\ \beta_1 \frac{\pi_{AT}}{2} & \beta_3 \frac{\pi_{GC}}{2} & & \alpha_2 \frac{\pi_{AT}}{2} \\ \beta_4 \frac{\pi_{AT}}{2} & \beta_2 \frac{\pi_{GC}}{2} & \alpha_1 \frac{\pi_{GC}}{2} & \end{pmatrix}.$$

From these, the expected number of substitutions N_{ij}, $i \neq j \in \{A, G, C, T\}$, in a DNA sequence can be calculated as $N_{ij} = f_i m_{ij} \pi_j$, where f_i is the frequency of the nucleotide i in the extant DNA sequence, $m_{ij} \in \{\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3, \beta_4\}$ is the mutation parameter for the replacement of i by j, and π_j is the equilibrium base frequency of j. Resolving for m_{ij}, we estimated the six parameters α_1–β_4 in the lineage displaying compositional equilibrium. We assessed π_j from the corresponding base frequency in the extant DNA sequence and N_{ij} from the observed number of substitutional changes in the respective direction. The equilibrium base frequency in the lineage undergoing the compositional change was then determined by resolving the equation for π_j and inserting the corresponding substitution parameter estimated in the first step.

Results
DNA Sequence Determination

We have focused on the genomic region around the gene *ZFY* located at Yp11.31 in the nonrecombining part of the human Y chromosome to compare rate and pattern of DNA sequence change between humans and chimpanzees. The second and the seventh coding exons of the gene were amplified by PCR and were used separately to probe a male chimpanzee BAC library (RPCI-43). A single clone (RPCI-43-107I4) hybridized to both probes. We confirmed the presence of the two exons by PCR and determined the nucleotide sequence of the BAC insert yielding 157 kb of contiguous DNA sequence. Sequencing of a partly overlapping genomic long-range PCR product extended the determined DNA sequence to a total of 167 kb. The comparison to the human Y chromosome sequence reveals that 90 kb of the chimpanzee DNA sequence encompassing the *ZFY* gene align to the region around the corresponding gene in humans. However, the human counterpart to the subsequent 77 kb of chimpanzee DNA sequence is located approximately 3.7 Mb downstream of the human *ZFY* gene (fig. 1). We confirmed by PCR from chimpanzee and gorilla genomic DNA that this discontinuous alignment does not reflect an artificial rearrangement in the BAC clone insert (data not shown). This indicates that the chimpanzee DNA sequence covers an evolutionary break point previously mapped to the region Yp11.31 (Page et al. 1984), where an ~3.7-Mb-large genomic segment was inserted along the human lineage.

The nucleotide sequence corresponding to the 90 kb upstream of the human insertion (fig. 1, shaded area) was determined in gorilla from multiple overlapping long-range PCR products. Further analysis was restricted to this region.

Analysis of DNA Sequence Differences in the ZFY Region

The alignment of the DNA sequences from human, chimpanzee, and gorilla and subsequent exclusion of positions that contain insertions, deletions, or masked nucleotides result in a total of 79,819 compared nucleotide positions with a GC content of 38.9%. Within these, 1,123 sites differ between humans and chimpanzees. In contrast, both species display substantially more differences when compared to the
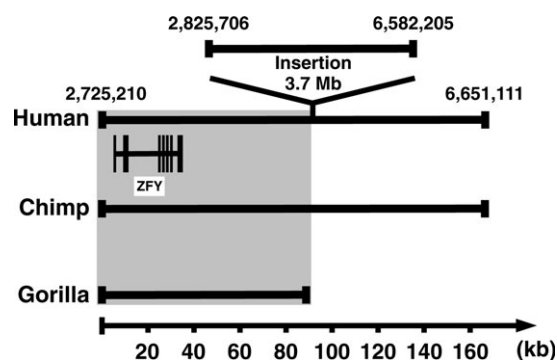

FIG. 1.—Overview of the DNA sequence alignment in the ZFY region. The human region differs from that of chimpanzees and gorillas by the presence of a 3.7-Mb-large human-specific insertion. The shaded area was used for the analysis of the substitution pattern. The position of the *ZFY* gene in the analyzed genomic region is shown. Numbers on the human graph represent the corresponding positions in the University of California–Santa Cruz human genome sequence (hg15; http://www.genome.ucsc.edu).

gorilla (1,569 and 1,645, respectively). This is indicative of a closer phylogenetic relationship between the human and chimpanzee DNA sequences to the exclusion of the gorilla. Therefore, we used gorilla as out-group to infer the ancestral state of DNA sequence differences between humans and chimpanzees by parsimony (Webster, Smith, and Ellegren 2003). Fifteen positions were excluded from the analysis because all three species differed at these positions. Among the remaining sites, we detected significantly less substitutional changes along the human lineage (516) than along the chimpanzee lineage (592) (table 1; relative rate test: $P = 0.03$; Tajima 1993). This difference remains significant when positions in a CpG context are excluded from the analysis ($P = 0.02$). Repeating the analysis in a maximum likelihood framework confirmed our findings from the parsimony approach. The null hypothesis of equal branch lengths on the human and chimpanzee lineages was rejected (likelihood ratio test: <0.05).

Substitution Pattern and GC Content in the ZFY Region

The difference in the number of substitutional changes between the human and chimpanzee ZFY regions is accompanied by a marked difference in the pattern of changes. In chimpanzees, a GC base pair has been equally often replaced by an AT base pair and vice versa (232 and 229, respectively). In contrast, on the human lineage, an excess of GC $\rightarrow$ AT changes over the reverse direction is seen (241 and 176, respectively; $P < 0.002$). Again, this observation remains when we restrict the analysis to non-CpG positions (chimpanzee, 165:163; human, 165:118). We subsequently inferred the frequencies a GC base pair is replaced by an AT base pair and vice versa from the observed numbers of respective changes at all sites. Using these, we determined how the difference in their substitution patterns affects the long-term GC content of the ZFY regions in humans and chimpanzees. A decrease of the GC content by ~7% tending to an equilibrium value of 31.7% is predicted in humans. In the chimpanzees, however, the GC content will remain virtually unchanged ($GC_{equi} = 38.6\%$).

Table 1
Observed Number of Substitutions in the Human and Chimpanzee ZFY Regions

	Human	Chimpanzee
G → A/C → T	192	185
G → T/C → A	49	47
G → C	66	76
A → G/T → C	144	176
A → C/T → G	32	53
A → T	33	55

Table 2
Substitution Frequencies per Site in the Human and Chimpanzee ZFY Regions

		Human	Chimpanzee
Non-CpG	G → A/C → T	0.0040	0.0039
	G → T/C → A	0.0015	0.0015
	G → C	0.0022	0.0025
	A → G/T → C	0.0030	0.0037
	A → C/T → G	0.0007	0.0011
	A → T	0.0006	0.0011
CpG	G → A/C → T	0.14	0.13
	π_{GC}[a]	33.6%	38.6%

[a] π_{GC} denotes the equilibrium GC content of the human and chimpanzee ZFY regions estimated from the six neighbor-independent substitution rates and the transition rate at CpG dinucleotides (Arndt, Petrov, and Hwa 2003).

Recently, it has been indicated that the above approach to infer the equilibrium GC content of a DNA sequence might be inappropriate, as it fails to take neighbor-dependent substitution rates into account (Arndt, Petrov, and Hwa 2003). We therefore followed the Arndt, Petrov, and Hwa (2003) approach to achieve a more accurate estimate of the equilibrium GC content. We estimated the six complementary substitution rates as well as the transition rate at CpG dinucleotides in the ZFY region for the human and chimpanzee lineages (table 2). We then calculated the lineage-specific GC content in equilibrium given the observed substitution rates. Notably, these estimates for the equilibrium base compositions in the human and chimpanzee ZFY regions reproduce the results from the naive approach (human: GC$_{equi}$ = 33.6%; chimpanzee: GC$_{equi}$ = 38.6%). However, the decrease in the equilibrium GC content in humans is with a value of 5% slightly less pronounced.

Discussion
Region-Specific Differences in the Substitution Pattern

The observation of differing substitution rates and patterns for the analyzed Y chromosomal region along the human and chimpanzee lineages contrasts the view that, in general, DNA sequences evolve alike in both species (Ebersberger et al. 2002). This suggests that at least one factor affecting the substitution rate and the GC content in this region has changed on either lineage.

Our data set includes the gene *ZFY*. Therefore, differing extents of purifying selection–maintaining gene function could account for the varying evolutionary rates of this region in humans and chimpanzees. However, when we exclude the 3% of positions in our data set that are protein coding, the differing substitution rates and patterns between the two lineages remain (data not shown). Thus, we believe that purifying selection acting differently on the *ZFY* gene in humans and chimpanzees does not account for the discrepant evolution of the respective genomic regions. Recently, evidence has accumulated that the frequency of transitions at CpG dinucleotides varies among regions in the human genome (Ebersberger et al. 2002; Hellmann et al. 2003b). A similar variation between the human and chimpanzee ZFY regions would cause both the region's overall substitution rate as well as its equilibrium GC content (Arndt, Petrov, and Hwa 2003) to differ between the two species. However, transitions at CpG dinucleotides occur with equal frequencies on the human and chimpanzee lineages (table 2), and consequently, both

the differences in substitution rate and pattern remain significant when we exclude changes observed in a CpG context ($P < 0.02$ and $P < 0.01$, respectively). Thus, we find no evidence that the differing evolutionary rates and patterns of the ZFY region in humans and chimpanzees are accomplished only by a particular subset of the analyzed positions. Rather, it seems that the neighbor-independent substitution process affecting all positions in this genomic region has changed either on the human or on the chimpanzee lineage.

The observation that the GC content of the ZFY region in chimpanzees is not altered by the substitution process indicates that the substitution pattern in this region has remained constant sufficiently long to allow equilibrium to be reached. In contrast, the human ZFY region tends to an equilibrium GC content that is significantly below the currently observed value. This implies that the change in the substitution process has occurred on the human lineage.

Trigger of the Shift in the Substitution Process

We have shown that the human ZFY region is located right adjacent to a 3.7-Mb-large human-specific insertion. This event has changed the genomic landscape such that the ZFY region now resides next to a GC-poor genomic sequence block (60 kb, mean GC content of 34.1%). In chimpanzees, however, the subsequent 60 kb display a mean GC content of 40.6%. Thus, the tendency of the human ZFY region toward an ~5% lower equilibrium GC content coincides with a drop of the GC content by a similar extent in the adjacent genomic region. The co-occurrence of both events suggests the rearrangement of the genomic landscape the ZFY region is embedded in as a likely trigger of the observed change in the substitution pattern. Notably, this change is of a quality that it adapts the GC content of the ZFY region to that of its new flanking region. It resembles, therefore, a previously proposed "influence" in the substitution pattern that levels differences in the GC content within regions of the human genome (Gu et al. 2000; Kumar and Subramanian 2002).

Model Choice

From the two models that are currently discussed to explain region-specific influences in the GC content, (1)

variable mutation bias and (2) BGC, the latter has been recently favored to explain the existence of region-specific GC contents (Eyre-Walker and Hurst 2001; Duret et al. 2002; Meunier and Duret 2004). However, it appears that it does not apply to our data set. The ZFY region is located outside the pseudoautosomal region on the human Y chromosome. Therefore, recombination due to allelic crossing-over does not occur. Furthermore, no second copy of this region is known in the human genome (Skaletsky et al. 2003), indicating that recombination by nonallelic gene conversion does not occur either in this region. The occurrence of at least one of these processes is a necessary prerequisite for BGC to occur. Thus, it seems reasonable to exclude this model in our case.

Rejecting BGC, we can think of no reason to assume that the differing substitution patterns among the human and chimpanzee ZFY regions are due to differences in the fixation process of mutations. Thus, the adaptation of the ZFY region in humans to a new equilibrium GC content is most likely accomplished by a recent change in the underlying mutation pattern.

Variable Mutation Bias and Substitution Rate

We subsequently used a modified version of Felsenstein's (1981) model of DNA sequence evolution to relate the observed change in the equilibrium GC content in the human ZFY region to the region's difference in the substitution rate. Because the model assumes that nucleotide positions evolve independently, we restricted the analysis to sites outside a CpG context. Given the model, we estimated for all six pairs of complementary substitution types at non-CpG sites how the equilibrium GC content of the ZFY region should have changed in humans in order to explain the observed difference in the number of the respective substitutional changes between the two species. Three out of six substitution types argue for a reduced equilibrium GC content on the human lineage relative to the chimpanzee lineage, and only the difference in the number of A $\to$ T changes suggests a reverse scenario (table 3). We subsequently applied a least square fit to determine the equilibrium GC content in humans that explains best the relative change of the overall substitution rate in humans. The such obtained value of 34.2% agrees perfectly with the GC content of 33.6% we have previously inferred as the new equilibrium base composition in the human ZFY region (see above). This indicates that the differing equilibrium base frequencies between the human and chimpanzee ZFY regions are sufficient to explain their differing evolutionary rates. Consequently, we propose that a recent shift in the mutation pattern in the human ZFY region accounts for both the region's tendency to a reduced GC content and its reduced substitution rate when compared to the ZFY region in chimpanzees.

Conclusion

In conclusion, we find that individual genomic regions can differ in rate and pattern of nucleotide substitutions between humans and chimpanzees, which is indicative of an evolutionary recent and regionally limited shift in the sub-

Table 3
Observed Number of Substitutions at Non-CpG Sites in the Human and Chimpanzee ZFY Regions

	Human	Chimpanzee	π_{GC}[a]
G $\to$ A/C $\to$ T	119	119	38.9
G $\to$ T/C $\to$ A	46	46	38.9
G $\to$ C	51	56	35.4
A $\to$ G/T $\to$ C	96	119	31.4
A $\to$ C/T $\to$ G	22	44	19.5
A $\to$ T	33	55	63.3

[a] π_{GC} denotes the equilibrium GC content of the human ZFY region calculated from the difference in the observed number of substitutions.

stitution process. Previously, it was speculated that such local differences in the substitution pattern between humans and chimpanzees are caused by a recent change in the local recombination rate (Meunier and Duret 2004). Our results, however, point toward a process that can occur independently of recombination. Based on the analysis of the ZFY region, we provide evidence for a locally restricted variation of the mutation pattern between these regions in the human and chimpanzee genomes. The implications of our findings for the discussion on what mechanism accounts for the existence of region-specific base compositions are twofold. First, they support the hypothesis that the mutation pattern varies in the human genome. Second, they suggest that the local mutation pattern in primates is amenable to evolutionary change with an accompanied effect on a region's base composition also in recent evolutionary timescales, as has been previously shown in flies (Takano-Shimizu 2001). Adding the observation that the altered mutation pattern acts to homogenize the base composition in the rearranged human ZFY region results in a hitherto unique support in favor of the hypothesis of regional mutation bias as a determinant of the local base composition.

When we relate our findings to those from other studies of DNA sequence evolution in the human genome, it seems that influences acting both on the mutational level and on the level of allele fixation determine pattern and rate of evolutionary DNA sequence. In large parts, both may be mediated by the process of recombination. Further insights into this matter are likely to emerge from the observation that the extent to which individual factors influence the process of DNA sequence evolution itself is subject to changes in recent evolutionary timescales. Therefore, a comprehensive analysis of genomic regions evolving differently in humans and chimpanzee genomes is likely to contribute substantially to the understanding of DNA sequence evolution in the human genome.

Acknowledgments

Literature Cited

Arndt, P. F., D. A. Petrov, and T. Hwa. 2003. Distinct changes of genomic biases in nucleotide substitution at the time of mammalian radiation. Mol. Biol. Evol. **20**:1887–1896.

Bernardi, G. 2000. Isochores and the evolutionary genomics of vertebrates. Gene **241**:3–17.

Birdsell, J. A. 2002. Intergrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. Mol. Biol. Evol. **19**:1181–1197.

Casane, D., S. Boissinot, B. H. Chang, L. C. Shimmin, and W. Li. 1997. Mutation pattern variation among regions of the primate genome. J. Mol. Evol. **45**:216–226.

Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. **16**:10881–10890.

Duret, L., M. Semon, G. Piganeau, D. Mouchiroud, and N. Galtier. 2002. Vanishing GC-rich isochores in mammalian genomes. Genetics **162**:1837–1847.

Ebersberger, I., D. Metzler, C. Schwarz, and S. Paabo. 2002. Genomewide comparison of DNA sequences between humans and chimpanzees. Am. J. Hum. Genet. **70**:1490–1497.

Ewing, B., and P. Green. 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. Genome Res. **8**:186–194.

Eyre-Walker, A. 1993. Recombination and mammalian genome evolution. Proc. R. Soc. Lond. B Biol. Sci. **252**:237–243.

Eyre-Walker, A., and L. D. Hurst. 2001. The evolution of isochores. Nat. Rev. Genet. **2**:549–555.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17**:368–376.

Francino, M. P., and H. Ochman. 1999. Isochores result from mutation not selection. Nature **400**:30–31.

Fullerton, S. M., A. BernardoCarvalho, and A. G. Clark. 2001. Local rates of recombination are positively correlated with GC content in the human genome. Mol. Biol. Evol. **18**:1139–1142.

Galtier, N., G. Piganeau, D. Mouchiroud, and L. Duret. 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. Genetics **159**:907–911.

Gu, Z., H. Wang, A. Nekrutenko, and W. H. Li. 2000. Densities, length proportions, and other distributional features of repetitive sequences in the human genome estimated from 430 megabases of genomic sequence. Gene **259**:81–88.

Hellmann, I., I. Ebersberger, S. E. Ptak, S. Pääbo, and M. Przeworski. 2003*a*. A neutral explanation for the correlation of diversity with recombination rates in humans. Am. J. Hum. Genet. **72**:1527–1535.

Hellmann, I., S. Zollner, W. Enard, I. Ebersberger, B. Nickel, and S. Paabo. 2003*b*. Selection on human genes as revealed by comparisons to chimpanzee cDNA. Genome Res. **13**:831–837.

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

Kumar, S., and S. Subramanian. 2002. Mutation rates in mammalian genomes. Proc. Natl. Acad. Sci. USA **99**:803–808.

Lander, E. S., L. M. Linton, B. Birren et al. (254 co-authors). 2001. Initial sequencing and analysis of the human genome. Nature **409**:860–921.

Lercher, M. J., and L. D. Hurst. 2002. Human SNP variability and mutation rate are higher in regions of high recombination. Trends Genet. **18**:337–340.

Lercher, M. J., N. G. Smith, A. Eyre-Walker, and L. D. Hurst. 2002. The evolution of isochores. Evidence from snp frequency distributions. Genetics **162**:1805–1810.

Marais, G. 2003. Biased gene conversion: implications for genome and sex evolution. Trends Genet. **19**:330–338.

Meunier, J., and L. Duret. 2004. Recombination drives the evolution of GC-content in the human genome. Mol. Biol. Evol. **21**:984–990.

Page, D. C., M. E. Harper, J. Love, and D. Botstein. 1984. Occurrence of a transposition from the X-chromosome long arm to the Y-chromosome short arm during human evolution. Nature **311**:119–123.

Schmidt, H. A., K. Strimmer, M. Vingron, and A. von Haeseler. 2002. TREE PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics **18**:502–504.

Skaletsky, H., T. Kuroda-Kawaguchi, P. J. Minx et al. (37 co-authors). 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature **423**:825–837.

Smith, N. G., and A. Eyre-Walker. 2001. Synonymous codon bias is not caused by mutation bias in G+C-rich genes in humans. Mol. Biol. Evol. **18**:982–986.

Tajima, F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics **135**:599–607.

Takano-Shimizu, T. 2001. Local changes in GC/AT substitution biases and in crossover frequencies on Drosophila chromosomes. Mol. Biol. Evol. **18**:606–619.

Thiery, J. P., G. Macaya, and G. Bernardi. 1976. An analysis of eukaryotic genomes by density gradient centrifugation. J. Mol. Biol. **108**:219–235.

Webster, M. T., N. G. Smith, and H. Ellegren. 2003. Compositional evolution of noncoding DNA in the human and chimpanzee genomes. Mol. Biol. Evol. **20**:278–286.

Wolfe, K. H., P. M. Sharp, and W. H. Li. 1989. Mutation rates differ among regions of the mammalian genome. Nature **337**:283–285.

David Goldstein, Associate Editor