ARTICLE

Symmetry Analysis of an X-palindrome in Human and Chimpanzee

Yan-jiao Qi, Wen-yuan Qiu*

Department of Chemistry, State Key Laboratory of Applied Organic Chemistry, Lanzhou University, Lanzhou 730000, China

(Dated: Received on January 22, 2009; Accepted on April 13, 2009)

We analyze for the first time the rules of breaking in an X-palindrome between human and chimpanzee. Results indicate that although the overall changes that occurred in the human X-palindrome are fewer than in the chimpanzee, mutations occurring between the left arm and right arm were nearly equivalent both in human and chimpanzee when compared with orangutan, which implies evolutionary synchronization. However, there are many more A/T→G/C changes than G/C→A/T in a single arm, which would lead to an increasing trend in GC content and suggest that the composition is not at equilibrium. In addition, it is remarkable to find that there are much more asymmetrical nucleotide changes between the two arms of the human palindrome than that of the chimpanzee palindrome, and these mutations are prone to occur between bases with similar chemical structures. The symmetry seems higher in the chimpanzee palindrome than in the human X-palindrome.

Key words: Palindrome, Symmetry, Sequence alignment, GC content, Mutation

I. INTRODUCTION

In nature, the exploration of symmetrical principle is a noteworthy topic, especially as DNA sequencing is currently progressing. One class of symmetrical DNA sequences existing in human chromosomes is called the palindrome sequences. A palindrome consists of two arms of similar DNA sequence—with one inverted and complemented relative to the other—around a central, usually nonhomologous spacer. It was amazing to find that there are abundant palindromes in human Y chromosome, which contains a lot of male-specific genes [1,2]. Deep understanding of the origin and structure of the palindrome is still a challenge.

Recently, there have been many interesting results [3,4] which demonstrated that the human X chromosome also contains a preponderance of large, highly homologous palindromes that contain a lot of male-specific genes but not female genes [3,5]. If there is no bias, the left arm is strictly reversed and complementary to the right arm of a palindrome. However, nucleotide mutations often occur in the corresponding sites and thus the perfect symmetry has disappeared. Studies have made some structural and statistical progress about the symmetry of DNA sequences [6−10]. However, this is still an incompletely solved curious problem, and further research is needed to elucidate it.

At present, the mutational pattern of the arms of palindrome along the evolutional process has never been reported systematically. Studies showed that the mutation direction is nonrandom [11]. Knowing the pattern will help us to better understand the genome evolution by analyzing the relative substitution frequency between GC→AT and AT→GC [12−14]. Given that we share more than 98% of our DNA and almost all of our genes, chimpanzees and our closest living relatives are the best comparison to study the human genomes. In this work, we give the first analysis of the mutational changes of an X-palindrome between human and chimpanzee. The frequency of transition was much higher than that of transversion in genomic sequences, and the frequencies of G/C→A/T were not equal to A/T→G/C, which indicated composition was not at equilibrium. Although overall mutational changes on both arms of this human X-palindrome were less than that of chimpanzee compared with orangutan, the symmetry between the two arms of chimpanzee sequences was higher.

II. MATERIALS AND METHODS

A. Subjects and data

The human complete sequence IRX-151.5, and chimpanzee and orangutan BAC clones (AC145689, AC144384 and AC146919, AC148185 respectively) was downloaded from http://www.ncbi.nlm.nih.gov/, the NCBI website, and arms of palindromes were obtained by dot matrix program alignment [15].

*Author to whom correspondence should be addressed. E-mail: wyqiu@lzu.edu.cn

TABLE I Frequencies (%) of nucleotide changes on the X-palindromes between human and chimpanzee.

Category	GC%	A→G	T→C	G→A	C→T	A→C	T→G	G→T	C→A	A→T	T→A	G→C	C→G
Left-H	46.53	22.83	21.92	10.50	13.24	5.94	5.02	3.65	2.74	1.37	3.20	5.02	4.57
Right-H	46.58	22.22	21.37	12.82	11.11	5.13	5.13	3.00	3.42	3.00	1.71	5.98	5.13
Left-C	46.91	18.45	17.34	12.55	12.18	6.27	3.69	3.32	4.06	3.32	4.43	7.38	7.01
Right-C	46.91	17.34	17.71	14.02	12.18	4.06	6.27	2.21	2.95	5.54	2.58	8.86	6.27

Left-H and right-H mean the left and right arms of human palindrome, respectively. Left-C and right-C mean the left and right arms of chimpanzee palindrome, respectively.

B. DNA sequence alignment and statistical analysis

Sequence alignments were performed with the program ClustalX [16] using the default values. In order to make our alignments reliable, we realigned our sequences using different alignment parameters with the same program, and also using the default parameters of a different alignment program, the Chaos/Dialign software, which achieved both speed and alignment accuracy [17].

C. GC content equilibrium analysis in the human and chimpanzee genomes

According to Sueoka [18], the mean GC content at equilibrium ($P_{\rm GC}$) is defined as

$$P_{\rm GC} = \frac{v}{u+v} \tag{1}$$

where u and v are the rates of G/C→A/T and A/T→G/C, respectively.

In addition, Jiang and Zhao [19] transformed the formula to

$$\frac{u}{v} = \frac{1-P_{\rm GC}}{P_{\rm GC}} \tag{2}$$

$$\frac{u}{v} = \frac{N_u/f_{\rm GC}}{N_v/1-f_{\rm GC}} = \frac{N_u}{N_v}\frac{1-f_{\rm GC}}{f_{\rm GC}} \tag{3}$$

where N_u and N_v represent the number of G/C→A/T and A/T→G/C, respectively, and $f_{\rm GC}$ is the GC content. Eq.(3) becomes Eq.(2) when N_u is close to N_v, which means the sequence is under the GC content equilibrium.

III. RESULTS

A. Mutational asymmetric pattern on the X-palindromes along the evolutionary process

The alignment of the DNA sequences from human, chimpanzee, and orangutan and subsequent exclusion of positions that contain insertions, deletions, or masked nucleotides resulted in a total of 43865 and 43834 compared nucleotide positions on the left arm and the right arm respectively. Within these, there were 495 and 508 different sites between human and chimpanzee left arm and right arm, respectively. However, more differences appeared when human and chimpanzee were compared to the orangutan left one (1347 and 1402, respectively) and the right one (1348 and 1386, respectively). This indicated that there was a closer phylogenetic relationship between the human and chimpanzee palindromes. Therefore, we used orangutan as the out-group to infer the directions of nucleotide mutations on the human and chimpanzee lineages by maximum parsimony. Among these, 12 and 10 sites were excluded from the analysis on the left arm and the right arm respectively, because the three sited were all different absolutely and the directions could not be determined.

In addition, CpG sites are known to be more hypermutable and are subjected to more different mutational mechanism than non-CpG sites [20], and thus they can be considered separately [21,22]. We use the CpG island searcher program available at http://www.cpgislands.com/ [23], and the modified original criteria [24] which exclude *Alus* and many CpG islands not located within the promoters of genes. Regions of DNA of greater than 500 bp with a G+C equal to or greater than 55% and the observed/expected CpG of 0.65 were considered. When "non-CpG" sites were analyzed, all positions located in a CpG dinucleotide in at least one species were excluded.

Among all types of observed frequencies of nucleotide changes in the human and chimpanzee palindromes from Table I, we saw that the frequencies of A to G and T to C were higher than others. The differences between all frequencies of G to A and C to T were even more than fourfold that for the transversional type, especially A to T and T to A. It was interesting to notice that conversional frequencies observed between purine and pyrimidine on the left arm were not much different from that on the right arm of palindromes. However, there was a very large difference between all transversional frequencies on both arms of the X-palindromes for human and chimpanzee.

B. The relationship between mutational pattern and GC content

In general, the genomes of human and many other eukaryotes exhibit extremely different genomic landscapes

DOI:10.1088/1674-0068/22/04/401-405

[25], which are correlated with many other important features: gene density, intron size, distribution of transposable elements, replication timing, and so on [26−31]. Recently, many biologists have become interested in determining whether the GC content of respective sequences has reached equilibrium [32−37]. Studies suggested that the mutational pattern was the major cause that affected DNA sequences evolving differently in the human and chimpanzee genomes [38]. Elucidating the relationship between the mutational changes and the region-specific GC contents is still an interesting subject, which may help us to better understand the evolution of human genomes.

In order to examine a possible correlation between the GC content and the mutation patterns on the X-palindromes, we further analyzed the substitutions of A or T to G or C and their compositional equilibrium. For our comparisons, we moved a sliding window of 5 kbp in non-overlapping 5 kbp steps along the palindromes, and thus determined the GC content in each window. The resulting GC-content curves for the two palindromes are shown in Fig.1, they are almost, but not exactly, identical, implying that they have a common ancestor. From the curves, we can easily observe repeated segments, which clearly reflect the reverse and complementary property of the palindrome.

From the number of substitutions observed in human and chimpanzee (Table II), we can conclude that there were less substitutional changes on the human palindrome (219 and 234 for the left and right arm, respectively; relative rate test: P=0.51) than the chimpanzee palindrome (273 and 271 for the left and right arm, respectively; relative rate test: P=0.96). The mutational changes which occurred between the left arm and right arm were nearly equivalent both in human and chimpanzee, indicated evolutionary synchronization. The null hypothesis of equal branch lengths can not be rejected (likelihood ratio test: P>0.05). However, a GC pair has not often been equally replaced by an AT base pair on a single arm. There were excessive AT→GC (A or T to G or C) changes both in human and chimpanzee as shown in Table II. We subsequently inferred the frequencies that a GC base pair was replaced by an AT base pair and vice versa from the observed numbers of nucleotide changes. In addition, we also determined how these mutations affected the distribution of GC content of the palindromes in human and chimpanzee.

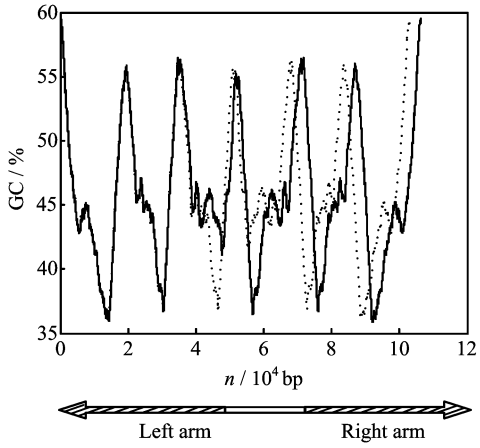

FIG. 1 GC-content curves for human (solid line) and chimpanzee (dotted line) palindromes. Below these curves, the corresponding arms of human palindrome are shown. There are two repeats on each arm, and the one in the spacer is identical to the right ones in human palindrome, whereas it is identical to the left ones in chimpanzee palindromes.

TABLE II Observed number of substitutions in the human and chimpanzee palindromes.

	Left-H	Right-H	Left-C	Right-C
G→A/C→T	52	56	68	71
G→T/C→A	14	15	41	14
G→C	21	26	20	41
A→G/T→C	98	102	97	95
A→C/T→G	24	24	27	28
A→T	10	11	21	15

Theoretically, the number of G/C→A/T mutations is expected to be equal to the number of A/T→G/C mutations if the composition is at equilibrium, irrespective of GC content. The expected u/v ratio at equilibrium decreases when GC content increases and becomes less than 1 when GC content is higher than 50% [19]. From the mutations of the human and chimpanzee palindromes shown in Table III, we conclude that the GC content in all systematic regions was not at equilibrium except in the left arm of chimpanzee (likelihood ratio test: P=0.36). There were much more A/T→G/C changes than G/C→A/T, thus led to an increasing trend in GC content.

C. The asymmetric mutation on both arms of the X-palindromes

If there is no change, the two arms of an ideal palindrome should be completely reverse and complementary. In order to further analyze the symmetry on the palindromes, we infer the mutational directions by comparing the left arm with the reverse complement sequence of right arm of the human and chimpanzee palindromes, respectively. The bases would be different if the symmetry was broken at the corresponding sites. Here, we just statistically considered the status about the right arm of palindromes using the left arm as the reference. The results were complementary on the left one when taking the right arm as the reference. For example, if the alignment pairs are A and G on the left arm and reverse complement of right arm, respectively, the direction of this change can be considered as T→C on the right arm. The real base is C, and the reverse

DOI:10.1088/1674-0068/22/04/401-405

TABLE III Mutations between nucleotides G/C and A/T in the human and chimpanzee palindromes.

Category	(G/C→A/T)/%	(A/T→G/C)/%	u/v	$(1-P_{GC})/P_{GC}$	P-value
Left-H	30.14	55.71	0.62	1.15	5.37×10^{-5}
Right-H	30.34	53.85	0.65	1.15	1.09×10^{-4}
Left-C	39.93	45.42	0.99	1.13	0.36
Right-C	31.37	45.39	0.78	1.13	0.01

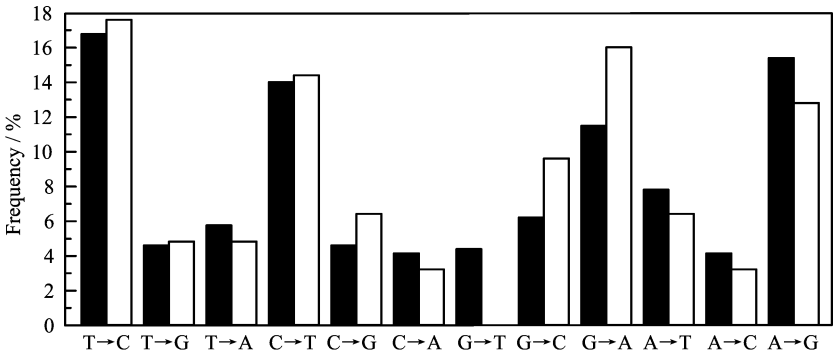

FIG. 2 Frequency of nucleotide changes among the palindromes. Black bar, the human palindrome; white bar, the chimpanzee palindrome. The frequencies of T→C, C→T, G→A, and A→G are higher than other changes in both palindromes.

complimentary base should be T on the right arm.

The overall frequencies of nucleotide changes are shown in Fig.2. Although the similar percents appeared in both human and chimpanzee palindromes, the total number of changes was 432, much higher than the 62 changes occurring in the right arm of the chimpanzee palindrome. Therefore, we can conclude that there were many more asymmetrical nucleotide changes on the human palindrome than that on the chimpanzee palindrome, and the symmetry is higher on the chimpanzee palindrome. In addition, we found that the mutational directions were easily affected by their neighboring bases, *e.g.*, if the neighboring base was purine/pyrimidine, the middle base would be likely to convert to purine/pyrimidine. Among these, mutations were prone to occur between bases with similar chemical structures.

IV. CONCLUSION

We examined the mutational patterns in an X-palindrome between human and chimpanzee. The different frequencies and patterns of nucleotide substitutions indicated that the compositional evolution was not at equilibrium in either of the two arms of human palindromes. Although the substitutional changes along the human palindrome lineage were fewer than along the chimpanzee lineage, we found that number of mutations between the arms of human palindrome was much more than that observed on the arms of the chimpanzee palindrome. The symmetrical analysis of the human and chimpanzee palindromes suggested that the mutations could occur easily between purines (A and G) or pyrimidines (C and T), which have similar chemical structures. Similar results could also be found in substitutional changes on human and chimpanzee palindromes when comparing them with orangutan. Therefore, the data indicate that symmetry breakage existed between the left and right arms, as well as the evolutional process. These rules observed in the palindromes may offer some new contributions for further analysis of symmetrical structures and evolutional process of other palindromes.

V. ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No.20173023 and No.90203012) and the Specialized Research Fund for the Doctoral Program of Higher Education of China (No.20020730006).

[1] H. Skaletsky, T. Kuroda-Kawaguchi, P. J. Minx, H. S. Cordum, L. Hillier, L. G. Brown, S. Repping, T. Pyntikova, J. R. Ali, T. Bieri, A. Chinwalla, A. Delehaunty, K. Delehaunty, H. Du, G. Fewell, L. Fulton, R. Fulton, T. Graves, S. Hou, P. Latrielle, S. Leonard, E. Mardis, R. Maupin, J. McPherson, T. Miner, W. Nash, C. Nguyen, P. Ozersky, K. Pepin, S. Rock, T. Rohlfing, K. Scott, B. Schultz, C. Strong, A. Tin-Wollam, S. Yang, R. H. Waterston, R. K. Wilson, S. Rozen, and D. C. Page, Nature **423**, 825 (2003).

DOI:10.1088/1674-0068/22/04/401-405

[2] S. Rozen, H. Skaletsky, J. D. Marszalek, P. J. Minx, H. S. Cordum, R. H. Waterston, R. K. Wilson, and D. C. Page, Nature **423**, 873 (2003).
[3] P. E. Warburton, J. Giordano, F. Cheung, Y. Gelfand, and G. Benson, Genome Res. **14**, 1861 (2004).
[4] M. J. Lercher, A. O. Urrutia, and L. D. Hurst, Mol. Biol. Evol. **20**, 1113 (2003).
[5] A. J. Zendman, D. J. Ruiter, and G. N. Van Muijen, J. Cell Physiol. **194**, 272 (2003).
[6] W. Y. Qiu, *Knot Theory, DNA Topology, and Molecular Symmetry Breaking*, *Chemical Topology-Applications and Techniques, Mathematical Chemistry Series*, Vol.6, D. Bonchev and D. H. Rouvray, Eds., Amsterdam: Gordon and Breach Science Publishers, 175 (2000).
[7] S. Larionov, A. Loskutov, and E. Ryadchenko, Chaos **18**, 013105 (2008).
[8] A. Sheari, M. Kargar, A. Katanforoush, S. Arab, M. Sadeghi, H. Pezeshk, C. Eslahchi, and S. A. Marashi, BMC Bioinformatics **9**, 274 (2008).
[9] N. Z. Jin, W. Y. Qiu, Z. X. Liu, and Y. J. Qi, Chin. J. Chem. Phys. **22**, 255 (2009).
[10] N. Z. Jin, Z. X. Liu, and W. Y. Qiu, Chin. J. Chem. Phys. **22**, 27 (2009).
[11] T. Gojobori, K. Ishii, and M. Nei, J. Mol. Evol. **18**, 414 (1982).
[12] J. Meunier and L. Duret, Mol. Biol. Evol. **21**, 984 (2004).
[13] M. T. Webster, N. G. Smith, L. Hultin-Rosenberg, P. F. Arndt, and H. Ellegren, Mol. Biol. Evol. **22**, 1468 (2005).
[14] L. Duret and P. F. Arndt, PLoS Genet. **4**, e1000071 (2008).
[15] E. L. Sonnhammer and R. Durbin, Gene **167**, GC1 (1995).
[16] J. D. Thompson, T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins, Nucleic Acids Res. **25**, 4876 (1997).
[17] M. Brudno, R. Steinkamp,and B. Morgenstern, Nucleic Acids Res. **32**, W41 (2004).
[18] N. Sueoka, Proc. Natl. Acad. Sci. USA **48**, 582 (1962).
[19] C. Z. Jiang and Z. M. Zhao, Genomics **88**, 527 (2006).
[20] A. P. Bird, Nucleic Acids Res. **8**, 1499 (1980).
[21] A. Khelifi, J. Meunier, L. Duret, and D. Mouchiroud, J. Mol. Evol. **62**, 745 (2006).
[22] I. Ebersberger and M. Meyer, Mol. Biol. Evol. **22**, 1240 (2005).
[23] D. Takai and P. A. Jones, In Silico Biol. **3**, 235 (2003).
[24] M. Gardiner-Garden and M. Frommer, J. Mol. Biol. **196**, 261 (1987).
[25] A. Eyre-walker and L. D. Hurst, Nat. Rev. Genet. **2**, 549 (2001).
[26] D. Mouchiroud, G. D'Onofrio, B. Aissani, G. Macaya, C. Gautier, and G. Bernardi, Gene **100**, 181 (1991).
[27] S. Saccone, A. De Sario, J. Wiegant, A. K. Raap, V. G. Della, and G. Bernardi, Proc. Natl. Acad. Sci. USA **90**, 11929 (1993).
[28] H. Caron, B. van Schaik, M. van der Mee, F. Baas, G. Riggins, P. van Sluis, M. C. Hermus, R. van Asperen, K Boon, P. A. Voûte, S. Heisterkamp, A. van Kampen, and R. Versteeg, Science **291**, 1289 (2001).
[29] S. M. Fullerton, A. B. Carvalho, and A. G. Clark, Mol. Biol. Evol. **18**, 1139 (2001).
[30] A. Kong, D. F. Gudbjartsson, J. Sainz, G. M. Jonsdottir, S. A. Gudjonsson, B. Richardsson, S. Sigurdardottir, J. Barnard, B. Hallbeck, G. Masson, A. Shlien, S. T. Palsson, M. L. Frigge, T. E. Thorgeirsson, J. R. Gulcher, and K. Stefansson, Nat. Genet. **31**, 241 (2002).
[31] M. J. Lercher, A. O. Urrutia, A. Pavlícek, and L. D. Hurst, Hum. Mol. Genet. **12**, 2411 (2003).
[32] M. T. Webster, N. G. Smith, and H. Ellegren, Mol. Biol. Evol. **20**, 278 (2003).
[33] J. E. Karro, M. Peifer, R. C. Hardison, M. Kollmann, and H. H. von Grünberg, Mol. Biol. Evol. **25**, 362 (2008).
[34] L. Duret, M. Semon, G. Piganeau, D. Mouchiroud, and N. Galtier, Genetics **162**, 1837 (2002).
[35] N. Galtier, G. Piganeau, D. Mouchiroud, and L. Duret, Genetics **159**, 907 (2001).
[36] J. I. Montoya-Burgos, P. Boursot, and N. Galtier, Trends Genet. **19**, 128 (2003).
[37] U. Pozzoli, G. Menozzi, M. Fumagalli, M. Cereda, G. P. Comi, R. Cagliani, N. Bresolin, and M. Sironi, BMC Evol. Biol. **8**, 99 (2008).
[38] M. T. Webster, E. Axelsson, and H. Ellegren, Mol. Biol. Evol. **23**, 1203 (2006).

DOI:10.1088/1674-0068/22/04/401-405