Negative correlation of G+C content at silent substitution sites between orthologous human and mouse protein-coding sequences

Type of paper: Short Communication

Negative correlation of G+C content at silent substitution sites between orthologous human and mouse protein-coding sequences

Naoki Takahashi and Hiroshi Nakashima*

Department of Clinical Laboratory Science, Graduate Course of Medical Science and Technology, Division of Health Science, Kanazawa University, 5-11-80 Kodatsuno, Kanazawa 920-0942, Japan

Running title: G+C characteristics of silent codon substitutions

*Corresponding author.

Abstract

We conducted a genome-wide analysis of variations in guanine plus cytosine (G+C) content at the third codon position at silent substitution sites of orthologous human and mouse protein-coding nucleotide sequences. Alignments of 3,776 human protein-coding DNA sequences with mouse orthologs having more than 50 synonymous codons were analyzed, and nucleotide substitutions were counted by comparing sequences in the alignments extracted from gap-free regions. The G+C content at silent sites in these pairs of genes showed a strong negative correlation (r = -0.93). Some gene pairs showed significant differences in G+C content at the third codon position at silent substitution sites. For example, human thymine-DNA glycosylase was A+T-rich at the silent substitution sites, while the orthologous mouse sequence was G+C-rich at the corresponding sites. In contrast, human matrix metalloproteinase 23B was G+C-rich at silent substitution sites, while the mouse ortholog was A+T-rich. We discuss possible implications of this significant negative correlation of G+C content at silent sites.

Key words: G+C content variation, human-mouse orthologs, nucleotide substitutions in synonymous codons.

The availability of complete mammalian genome sequences[1-4] provides an opportunity to characterize nucleotide substitution patterns among mammalian genomes by comparative sequence analysis.[5-11] The G+C content in bacterial genomes varies among species from 25% to 75%, but it is relatively homogeneous for genes within a given bacterial genome.[12-13] However, the G+C content of genes in a given mammalian genome varies considerably, because mammalian genomes are a mosaic of long (over hundreds of kilobases) DNA segments known as isochores.[14] Some G+C-poor isochores have G+C contents as low as 35%, while G+C-rich isochores have G+C contents as high as 60%. It is known that the genes within a given isochore are fairly homogeneous in G+C content. It has been reported that some homologous mammalian genes that occupy different chromosomal positions differ considerably in their base composition and codon usage.[14-16] Human α- and β-globin genes are an example of this position-dependent variation. The α-globin gene cluster occupies a G+C-rich region, while the β-globin gene resides in a G+C-poor region.[14]

In comparing alignments of orthologous human and mouse sequences, we noted that silent substitution sites at the third codon position were biased toward G+C-rich or A+T-rich nucleotides. For example, human thymine-DNA glycosylase was A+T-rich at silent substitution sites, while the orthologous mouse sequence was G+C-rich at the corresponding sites. However, human matrix metalloproteinase 23B was G+C-rich at silent substitution sites, whereas the mouse ortholog was A+T-rich at those sites. Since complete human and mouse genome sequences are now available, we conducted a comparative genome-wide analysis of G+C content variation at silent substitution sites in orthologous human and mouse sequences.

1. Correlation of G+C content at the third codon position at silent substitution sites

The G+C content at the third codon position in synonymous codons, i.e., silent substitutions of the same amino acid, was determined for both human and mouse sequences. For simplicity, only synonymous codons that had an identical nucleotide at the first codon position were considered. The G+C contents at the third codon position at silent substitution sites in the 3,776 pairs of human and mouse genes are plotted in Figure 1, showing a high negative correlation (r = -0.93). The plot indicates distinct variations in G+C content at mutual silent sites in many human and mouse orthologs.

--------- Figure 1 here --------

In two-fold degenerate codons, the equivalent third position nucleotides are either two purines (A/G) or two pyrimidines (C/T). Therefore, their silent substitutions always result in G+C content variation. However, not every silent substitution in four-fold degenerate codons yields G+C content variation. To further examine G+C content correlation at silent substitution sites, the G+C contents in four-fold degenerate codons were analyzed. The G+C contents at silent substitution sites were determined for the eight sets of four-fold degenerate codons (Ala: GCN, Arg: CGN, Gly: GGN, Leu: CUN, Pro: CCN, Ser: UCN, Thr: ACN, and Val: GUN) in human and mouse sequences. The plot of the G+C contents in four-fold degenerate sites between 2,084 human and mouse orthologous sequences having more than 50 four-fold degenerate codons had a high negative correlation coefficient (r = -0.82; data not shown).

2. Classification of orthologous sequences according to G+C content at the third codon position at silent substitution sites

Human and mouse orthologous sequence pairs were divided into three groups according to the G+C content at the third codon position in synonymous substitution sites. In group (a),

the human gene had a much lower G+C content than that in the mouse gene, in group (b) the human gene had a much higher G+C content than that in the mouse gene, and in group (c) the human gene had a G+C content similar to that in the mouse gene. The number of genes in the three groups varied according to the criterion of the classification. Using a cutoff of 30% lower G+C content in the human gene than in the mouse, 25.4% (960/3776) of the orthologous sequence pairs were classified in group (a). Based on 30% higher G+C content in the human gene than in the mouse gene, group (b) contained 17.2% (648/3776) of the orthologous sequence pairs. Group (c) contained the remaining 57.4% (2168/3776) of sequence pairs, which showed deviations between -30% and +30% in G+C content when human and mouse orthologous sequences were compared. Table 1 lists the proteins encoded by 10 well-characterized genes in each of the three groups. Most of the gene pairs within groups (a), (b), and (c) had different chromosomal locations. However, some genes within the same group had the same chromosomal locations. For example, in group (a), human potassium channel tetramerization domain containing protein 3 and human ribosomal protein S6 kinase are both located on human chromosome 1q41, and both of the corresponding mouse genes are on mouse chromosome 1H6. In group (b), human agrin, human calcium binding protein Cab45 precursor, and human matrix metalloproteinase 23B are all located on human chromosome 1p36.33, and the corresponding mouse genes are on mouse chromosome 4E2. These findings suggest that the G+C content at some silent substitution sites might be determined by their chromosomal locations. This finding is consistent with the report by Bernardi et al.[14]

--------- Table 1 here --------

3. Frequencies of substitution

The parts of the alignments for thymine-DNA glycosylase (human gene: NM_003211.3 and mouse gene: NM_011561.1) and for matrix metalloproteinase 23B (human gene: NM_006983.1and mouse gene: NM_011985.1) are shown in Figure 2(a) and (b), respectively. In Figure 2, nucleotides A and T at silent substitution sites are red, nucleotides G and C at silent substitution sites are blue, and other nucleotides are shown in yellow. Amino acids are shown along the DNA sequences.

--------- Figure 2 (a) and (b) here --------

Nucleotide substitutions were observed in 14% (738,506/5,401,758) of all of the total nucleotides contained within the 3,776 pairs of orthologous human and mouse genes. Substitution frequencies in codons were 18% at the first position, 12% at the second, and 70% at the third. Silent nucleotide substitutions with an identical nucleotide at the first codon position accounted for 58% (425,945/738,506) of the total substitutions. The substitution frequency of transitions (purine-purine and pyrimidine-pyrimidine substitutions) was 66.1%, and that of transversions (purine-pyrimidine and pyrimidine-purine substitutions) was 33.9%. Transitions accounted for 72.1%, and transversions for 27.9%, of the total silent nucleotide substitutions. Transitions were more frequent than transversions at silent substitutions because transitions at the third codon position are essentially silent.

When a silent substitution was observed at an alignment site, a silent nucleotide substitution was assumed to occur once in either branch at a synonymous codon site since the divergence of human and mouse lineage. Silent substitutions at synonymous codon sites between human and mouse sequences were estimated. Nucleotide substitutions were considered from the human sequence. The frequencies of the four nucleotides A, T, C, and G at the third codon position at silent substitution sites in human sequence were expressed as f(A), f(T), f(C), and f(G). Let α and β be the transition and the transversion substitution rate

per year per site. *T* indicates the divergence time between human and mouse. Then, the nucleotide substitution frequencies at silent sites from human to mouse were calculated as shown in Table 2. Substitution frequency at G or C nucleotide in human silent sites is $2T(\alpha+2\beta)\cdot(f(G)+f(C))$, and that in mouse silent sites is $2T(\alpha+\beta)\cdot(f(A)+f(T))+ 2T\beta\cdot(f(G)+f(C))$, which is equivalent to $2T(\alpha+\beta)-2T\alpha\cdot(f(G)+f(C))$. The above nucleotide substitution frequencies were expressed as X and Y, respectively.

$$X = 2T(\alpha+2\beta)\cdot(f(G)+f(C))$$

$$Y = 2T(\alpha+\beta)-2T\alpha\cdot(f(G)+f(C))$$

The relationship between X and Y is

$$Y =2T (\alpha+\beta)-\{\alpha/(\alpha+2\beta)\}\cdot X$$

This equation indicates that Y increases when X decreases and Y decreases when X increases. This result indicated a negative correlation in the variation of G+C content at silent sites between the two DNA sequences.

--------- Table 2 here --------

4. The implications of substitutions at silent sites

Because substitutions at silent sites in codons do not change amino acids, no effect on proteins would be expected, and these substitutions are commonly thought of as being evolutionarily neutral. However, substitutions at silent sites do alter codon usage. Grantham reported that synonymous codons are used differently by different organisms,[18] and Ikemura found a strong positive correlation between codon usage and tRNA content in unicellular organisms.[19] The codon-choice patterns of genes are often very different among multicellular eukaryotes, and codon usage in mammals is known to have dramatic effects on the translation rate.[20] Our findings on the differential codon usage between human and mouse genes suggest

the possibility of different expression patterns.

Evidence indicates that genes with a high G+C content at the third codon position are usually surrounded by long G+C-rich genomic sequences, while those with a low G+C content at the third position are usually surrounded by long A+T-rich sequences.[19,21] Human-mouse genome sequence comparisons demonstrated a large number of rearrangements of conserved syntenic segments. [2,22] Since human and mouse genomes exhibit large variations in G+C content (e.g., isochores),[14] the rearrangements might produce a large deviation in G+C content between human and mouse genes by changing the surrounding sequences. The gene pairs classified into groups (a) and (b), which showed a large variation in G+C content at silent substitution sites, are considered to be products of the rearrangements of syntenic segments. The genes located in an identical syntenic segment exhibited similar G+C content variation. Gene rearrangements could be the cause of large variations in G+C content at silent substitution sites, and might lead to a significant negative correlation.

Nucleotide substitutions between human and mouse sequences have accumulated during evolution ever since their divergence from a common ancestor. It is generally assumed that the substitutions occurred independently in the two species, and there would seem to be no connection between the silent nucleotide substitutions in humans and mice. The results of this study raise the question of whether correlated substitutions at silent sites might have some possible evolutionary function. Further study is needed to address this issue.

Acknowledgments: We are grateful to one anonymous referee and an editor for constructive suggestions regarding negative correlation.

References

1. International Human Genome Sequencing Consortium. 2001, Initial sequencing and analysis of the human genome, *Nature*, **409**, 860-921.
2. Mouse Genome Sequencing consortium. 2002, Initial sequencing and comparative analysis of the mouse genome, *Nature*, **420**, 520-562.
3. Rat Genome Sequencing Project Consortium. 2004, Genome sequence of the Brown Norway Rat yields insights into mammalian evolution, *Nature*, **428**, 493-521.
4. The Chimpanzee Sequencing and Analysis Consortium. 2005, Initial sequence of the chimpanzee genome and comparison with the human genome, *Nature*, **437**, 69-87.
5. Castresana, J. 2002, Genes on human chromosome 19 show extreme divergence from the mouse orthologs and a high GC content, *Nucleic Acids Res*., **30**, 1751-1756.
6. Hardison, R.C., Roskin, K.M., Yang, S., et al. 2003, Covariation in frequencies of substitution, deletion, transposition, and recombination during eutherian evolution, *Genome Res*., **13**, 13-26.
7. Yang, S., Smit, A.F., Schwartz, S., et al. 2004, Patterns of insertions and their covariation with substitutions in the rat, mouse, and human genomes, *Genome Res*., **14**, 517-527.
8. Cooper, G.M., Brudno, M., Stone, E.A., Dubchak, I., Batzoglou, S., and Sidow, A. 2004, Characterization of evolutionary rates and constraints in three mammalian genomes, *Genome Res*., **14**, 539-548.
9. Zhang, Z. and Gerstein, M. 2003, Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes, *Nucleic Acids Res*., **31**, 5338-5348.
10. Subramanian, S. and Kumar, S. 2003, Neutral substitutions occur at a faster rate in exons

than in noncoding DNA in primate genomes, *Genome Res.*, **13**, 838-844.

11. Clark, A.G., Glanowski, S., Nielsen, R., et al. 2003. Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios, *Science*, **302**, 1960-1963.
12. Muto, A. and Osawa, S. 1987, The guanine and cytosine content of genomic DNA and bacterial evolution, *Proc. Natl. Acad. Sci. USA,* **84**, 166-169.
13. Lawrence, J.G. and Ochman, H. 1997, Amelioration of bacterial genomes: rates of change and exchange, *J. Mol. Evol.*, **44**, 383-397.
14. Bernardi, G., Olofsson, B., Filipski, J., et al. 1985, The mosaic genome of warm-blooded vertebrates, *Science,* **228**, 953-958.
15. Nadeau, J.H. and Taylor, B.A. 1984, Lengths of chromosomal segments conserved since divergence of man and mouse, *Proc. Natl. Acad. Sci. USA*, **81**, 814-818.
16. Mouchiroud, D. and Gautier, C. 1988, High codon-usage changes in mammalian genes, *Mol. Biol. Evol.*, **5**, 192-194.
17. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990, Basic local alignment search tool, *J. Mol. Biol.*, **215,** 403-410.
18. Grantham, R. 1980, Workings of the genetic code, *Trend Biochem. Sci.*, **5**, 327-331.
19. Ikemura, T. 1985, Codon usage and tRNA content in unicellular and multicellular organisms, *Mol. Biol. Evol.*, **2**, 13-34.
20. Zolotukhin, S., Potter, M., Hauswirth, W.W., Guy, J., and Muzyczka, N. 1996, A "humanized" green fluorescent protein cDNA adapted for high-level expression in mammalian cells, *J. Virol.*, **70**, 4646-4654.
21. Ikemura, T. and Wada, K. 1991, Evident diversity of codon usage patterns of human genes with respect to chromosome banding patterns and chromosome numbers; relation between nucleotide sequence data and cytogenetic data, *Nucleic Acids Res.*, **19**,

4333-4339.

22. Bourque, G.., Pevzner, P.A., and Tesler, G.. 2004, Reconstructing the genomic architecture of ancestral mammals: lessons from human, mouse, and rat genomes, *Genome Res.*, **14**, 507-516.

Table 1. Variations in G+C content at the third codon position in synonymous substitution sites (G+C III) for protein-coding genes. The length, the percentage of nucleotide sequence identity, the G+C content of the entire alignment, and chromosomal locations for human and mouse are shown.

(a) G+C III content: human << mouse

No. gene product	entire alignment						synonymous codons		
	length	identity	G+C (%)		chromosomal location		G+C III (%)		
	nucleotides	(%)	human	mouse	human	mouse	No.	human	mouse
1. cytidine monophosphate N-acetylneuraminic acid synthetase	1203	88.0	43.1	49.9	12p12.1	6G2	106	19.8	84.0
2. DNA helicase HEL308	849	80.8	40.6	50.1	4q21.23	5E4	83	18.1	80.7
3. DNA primase large subunit, 58kDa	1515	84.4	41.4	48.6	6p12	1B	136	19.9	80.2
4. dolichyl phosphate glucosyltransferase	900	87.2	41.2	48.0	13q13.3	3C	71	15.5	81.7
5. geranylgeranyltransferase type I, beta subunit	1131	87.1	44.4	51.6	5q22.3	18C	119	18.5	79.0
6. phosphatidylinositol glycan, class B mannose transferase	1482	80.2	38.7	50.0	15q21	9D	126	12.7	86.5
7. phosphoinositide-3-kinase, regulatory subunit 4, p150	3477	85.3	42.5	51.8	3q21.3	9F1	389	15.2	82.5
8. potassium channel tetramerization domain containing protein	1815	83.3	43.4	54.8	1q41	1H6	225	12.4	83.1
9. ribosomal protein S6 kinase, 52kDa, polypeptide 1	990	80.6	44.0	56.1	1q41	1H6	88	14.8	85.2
10. thymine-DNA glycosylase	915	85.8	40.2	48.5	12q24.1	10C1	102	19.6	85.3

(b) G+C III content: human >> mouse

No. gene product	length	identity	entire alignment G+C (%)		chromosomal location		synonymous codons G+C III (%)		
	nucleotides	(%)	human	mouse	human	mouse	No.	human	mouse
1. agrin (neuronal aggregation factor)	2448	80.2	69.0	58.7	1p36.33	4E2	270	81.9	23.0
2. calcium binding protein Cab45 precursor	1074	82.8	60.6	51.5	1p36.33	4E2	103	85.4	24.3
3. galactose-3-O-sulfotransferase 2	657	74.1	69.0	54.8	2q37.3	1D	54	96.3	31.5
4. alpha-L-iduronidase precursor	1827	80.2	67.9	56.7	4p16.3	5E5	164	87.2	21.3
5. matrix metalloproteinase 23B	996	81.6	68.7	56.1	1p36.33	4E2	104	93.3	13.5
6. methyl-CpG binding domain protein 3	801	89.1	65.2	58.8	19p13.3	10C1	73	86.3	20.6
7. N-methyl D-aspartate receptor 1 isoform NR1-3	2814	90.7	60.8	55.8	9q34.3	2A3	235	81.3	24.7
8. plakophilin 3	2391	86.4	67.3	60.5	11p15	7F5	216	83.8	21.3
9. prostaglandin E receptor 1 subtype EP1	642	80.7	76.6	65.6	19p13.1	8C2	75	89.3	29.3
10. renin binding protein	1200	85.2	61.5	53.8	Xq28	XA7.3	92	82.6	22.8

(c) G+C III content: human ≈ mouse

No. gene product	length	identity	entire alignment G+C (%)		chromosomal location		synonymous codons G+C III (%)		
	nucleotides	(%)	human	mouse	human	mouse	No.	human	mouse
1. apolipoprotein B	5394	82.3	46.8	46.9	2p24	12A1	365	49.0	47.7
2. dihydrolipoamide S-succinyltransferase	1203	88.2	51.3	51.6	14q24.3	12D3	88	47.7	47.7
3. 5'-3' exoribonuclease 2	2736	91.9	43.9	44.2	20p11.2	2H1	146	45.9	45.2
4. glutathione synthetase	1422	88.3	56.4	55.3	20q11.2	2H1	87	50.6	50.6
5. 5-hydroxytryptamine (seretonin) receptor 3B	1263	80.1	50.5	50.9	11q23.1	9B	92	50.0	48.9
6. ornithine carbamoyltransferase	1062	88.9	44.1	44.9	Xp21.1	XA1	70	48.6	48.6
7. polymerase (DNA directed), alpha	3270	90.8	41.7	40.5	Xp22.1	XC	154	47.4	48.1
8. transaldolase 1	1011	88.1	54.9	55.0	11p15.5	7F3	81	54.3	55.6
9. triosephosphate isomerase 1	747	88.6	56.6	55.7	12p13	6F2	68	52.9	52.9
10. ubiquitin protein ligase E3C	3249	87.3	45.9	46.1	7q36.3	5B1	265	50.2	49.8

Table 2. Rates of nucleotide substitution.

	mouse nucleotide			
human nucleotide	A	T	C	G
A		$2T\beta f(A)$	$2T\beta f(A)$	$2T\alpha f(A)$
T	$2T\beta f(T)$		$2T\alpha f(T)$	$2T\beta f(T)$
C	$2T\beta f(C)$	$2T\alpha f(C)$		$2T\beta f(C)$
G	$2T\alpha f(G)$	$2T\beta f(G)$	$2T\beta f(G)$	

Figure legends

Figure 1.

G+C contents at silent sites in 3,776 human protein-coding sequences versus G+C contents at silent sites in 3,776 corresponding mouse sequences. The human and mouse cDNA sequences were obtained from Reference Sequence Release 11 from the U.S. National Center for Biotechnology Information (ftp://ftp.ncbi.nih.gov/refseq/). The protein-encoding nucleotide sequences were selected according to the feature table for the data. The amino acid sequences of 28,893 human cDNAs and 25,298 mouse cDNAs were obtained by translation. Orthologs were identified by the two-directional best hit approach using BLASTP.[17] Pairs of a given sequence were selected if they showed greater than 30% amino acid identity over three-fourths of the total length. To avoid bias, proteins showing greater than 30% sequence identity with other proteins in the same species were excluded. Gap-free alignment regions longer than 100 amino acid residues and the corresponding DNA sequences were analyzed. Based on this criterion, 3,776 pairwise alignments of human and mouse sequences that had more than 50 synonymous codons were chosen for analysis in this study.

Figure 2

Panels (a) and (b) show the alignment between sequences of human and mouse thymine-DNA glycosylase, and alignment of human and mouse sequences of matrix metalloproteinase 23B.

Fig. 1

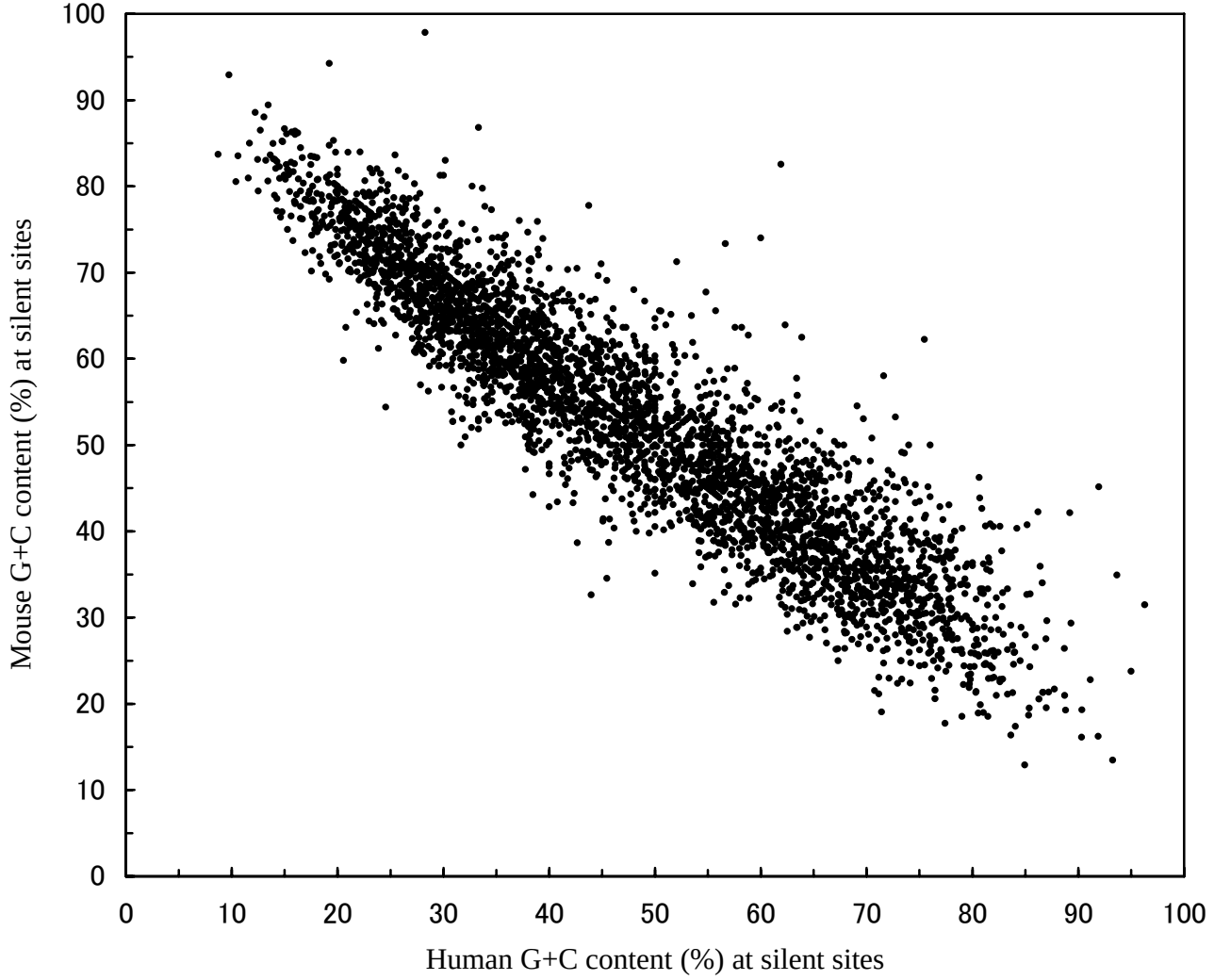
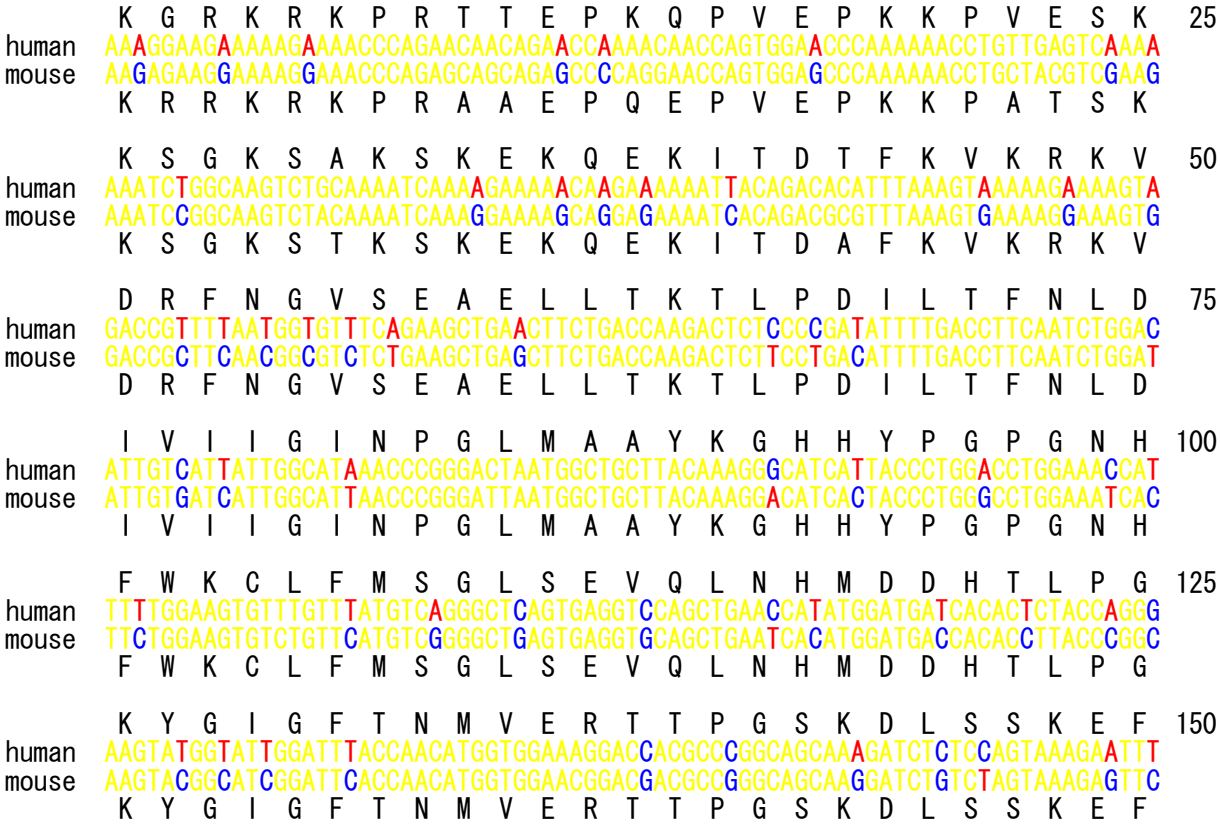
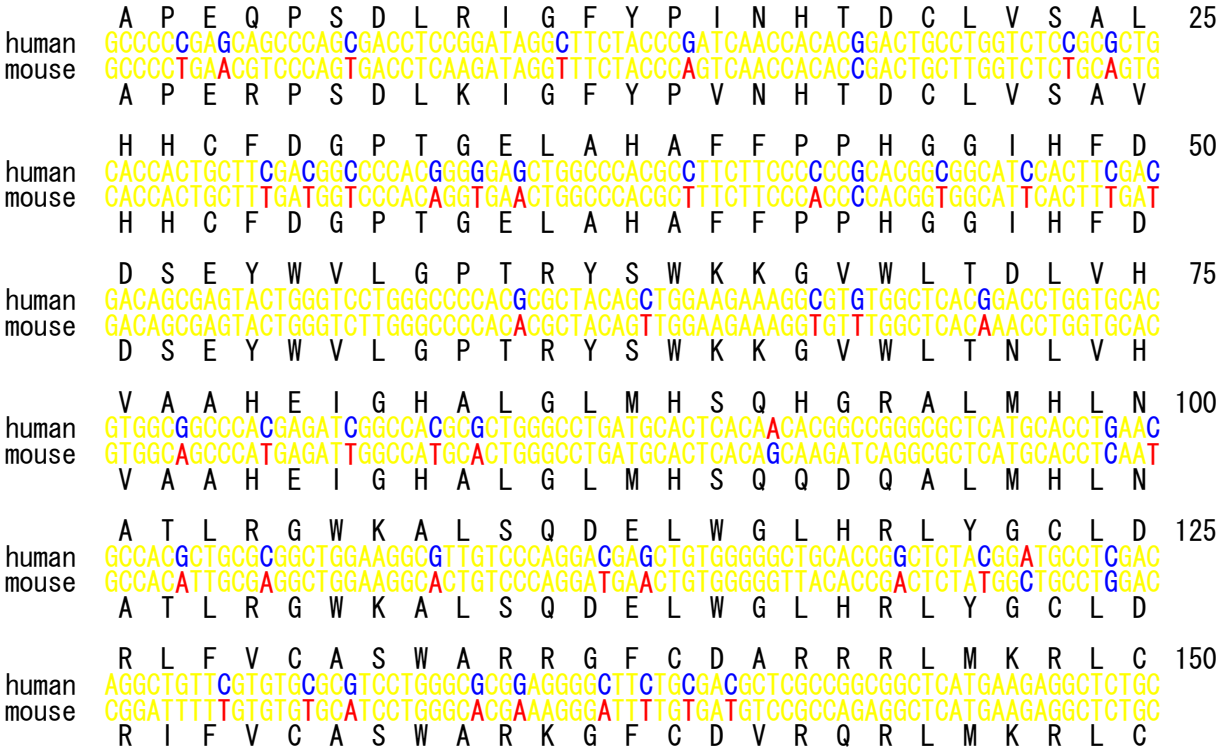

Mouse G+C content (%) at silent sites
Human G+C content (%) at silent sites
0
10
20
30
40
50
60
70
80
90
100

Fig. 2(a), (b)


K G R K R K P R T T E P K Q P V E P K K P V E S K 25
human AAAGGAAGAAAAAGAAAACCCAGAACAACAGAACCAAAACAACCAGTGGAACCCAAAAAACCTGTTGAGTCAAAA
mouse AAGAGAAGGAAAAGGAAACCCAGAGCAGCAGAGCCCCAGGAACCAGTGGAGCCCAAAAAACCTGCTACGTCGAAG
K R R K R K P R A A E P Q E P V E P K K P A T S K
K S G K S A K S K E K Q E K I T D T F K V K R K V 50
human AAATCTGGCAAGTCTGCAAAATCAAAAGAAAAACAAGAAAAAATTACAGACACATTTAAAGTAAAAAGAAAAGTA
mouse AAATCGGGCAAGTCTACAAAATCAAAGGAAAAGCAGGAGAAAATCACAGACGCGTTTAAAGTGAAAAGGAAAGTG
K S G K S T K S K E K Q E K I T D A F K V K R K V
D R F N G V S E A E L L T K T L P D I L T F N L D 75
human GACCGTTTTAATGGTGTTTCAGAAGCTGAACTTCTGACCAAGACTCTCCCCGATATTTTGACCTTCAATCTGGAC
mouse GACCGCTTCAACGGCGTCTCTGAAGCTGAGCTTCTGACCAAGACTCTTCCTGACATTTTGACCTTCAATCTGGAT
D R F N G V S E A E L L T K T L P D I L T F N L D
I V I I G I N P G L M A A Y K G H H Y P G P G N H 100
human ATTGTCATTATTGGCATAAACCCGGGACTAATGGCTGCTTACAAAGGGCATCATTACCCTGGACCTGGAAACCAT
mouse ATTGTGATCATTGGCATTAACCCGGGATTAATGGCTGCTTACAAAGGACATCACTACCCTGGGCCTGGAAATCAC
I V I I G I N P G L M A A Y K G H H Y P G P G N H
F W K C L F M S G L S E V Q L N H M D D H T L P G 125
human TTTTGGAAGTGTTTGTTTATGTCAGGGCTCAGTGAGGTCCAGCTGAACCATATGGATGATCACACTCTACCAGGG
mouse TTCTGGAAGTGTCTGTTCATGTCGGGGCTGAGTGAGGTGCAGCTGAATCACATGGATGACCACACCTTACCCGGC
F W K C L F M S G L S E V Q L N H M D D H T L P G
K Y G I G F T N M V E R T T P G S K D L S S K E F 150
human AAGTATGGTATTGGATTTACCAACATGGTGGAAAGGACCACGCCCGGCAGCAAAGATCTCTCCAGTAAAGAATTT
mouse AAGTACGGCATCGGATTCACCAACATGGTGGAACGGACGACGCCGGGCAGCAAGGATCTGTCTAGTAAAGAGTTC
K Y G I G F T N M V E R T T P G S K D L S S K E F


A P E Q P S D L R I G F Y P I N H T D C L V S A L 25
human GCCCCCGAGCAGCCCAGCGACCTCCGGATAGGCTTCTACCCGATCAACCACACGGACTGCCTGGTCTCCGCGCTG
mouse GCCCCTGAACGTCCCAGTGACCTCAAGATAGGTTTCTACCCAGTCAACCACACCGACTGCTTGGTCTCTGCAGTG
A P E R P S D L K I G F Y P V N H T D C L V S A V
H H C F D G P T G E L A H A F F P P H G G I H F D 50
human CACCACTGCTTCGACGGCCCCACGGGGAGCTGGCCCACGCCTTCTTCCCCCCGCACGGCGGCATCCACTTCGAC
mouse CACCACTGCTTTGATGGTCCCCACAGGTGAACTGGCCCACGCTTTCTTCCCACCCCACGGTGGCATTCACTTTGAT
H H C F D G P T G E L A H A F F P P H G G I H F D
D S E Y W V L G P T R Y S W K K G V W L T D L V H 75
human GACAGCGAGTACTGGGTCCTGGGCCCCACGCGCTACAGCTGGAAGAAAGGCGTGTGGCTCACGGACCTGGTGCAC
mouse GACAGCGAGTACTGGGTCTTGGGCCCCACACGCTACAGTTGGAAGAAAGGTGTTTGGCTCACAAACCTGGTGCAC
D S E Y W V L G P T R Y S W K K G V W L T N L V H
V A A H E I G H A L G L M H S Q H G R A L M H L N 100
human GTGGCGGCCCACGAGATCGGCCACGCGCTGGGCCTGATGCACTCACAACACGGCCGGGCGCTCATGCACCTGAAC
mouse GTGGCAGCCCATGAGATTGGCCATGCACTGGGCCTGATGCACTCACAGCAAGATCAGGCGCTCATGCACCTCAAT
V A A H E I G H A L G L M H S Q Q D Q A L M H L N
A T L R G W K A L S Q D E L W G L H R L Y G C L D 125
human GCCACGCTGCGCGGCTGGAAGGCGTTGTCCCAGGACGAGCTGTGGGGGCTGCACCGGCTCTACGGATGCCTCGAC
mouse GCCACATTGCGAGGCTGGAAGGCACTGTCCCAGGATGAACTGTGGGGGTTACACCGACTCTATGGCTGCCTGGAC
A T L R G W K A L S Q D E L W G L H R L Y G C L D
R L F V C A S W A R R G F C D A R R R L M K R L C 150
human AGGCTGTTCGTGTGCGCGTCCTGGGCGCGGAGGGGCTTCTGCGACGCTCGCCGGCGGCTCATGAAGAGGCTCTGC
mouse CGGATTTTTGTGTGTGCATCCTGGGCACGAAAGGGATTTTGTGATGTCCGCCAGAGGCTCATGAAGAGGCTCTGC
R I F V C A S W A R K G F C D V R Q R L M K R L C