NRC Research Press

ARTICLE

Gene conversions are frequent but not under positive selection in the Siglec gene families of primates

Mouldi Zid and Guy Drouin

Abstract: Siglecs are cell surface proteins that belong to the immunoglobulin superfamily and which bind sialic acids. They are composed of two groups, the conserved Siglecs and the CD33-related Siglecs. Previous studies have reported the occurrence of gene conversions between human CD33-related Siglecs and suggested that these conversions are adaptive because they increase the diversity of these immunoglobulin-related genes. Here, we analyze the Siglec genes of five primate species and show that gene conversions are not observed between conserved Siglec genes but that they are frequent between primate CD33-related Siglecs. The gene conversions between CD33-related Siglec genes only occur between similar genes and equally frequently in sialic acid binding and nonbinding domains. Furthermore, dN/dS ratio tests show that most of the Ig-like V-type 1 and the Ig-like C2-type 1 domains of Siglec genes evolve either neutrally or under purifying selection and that gene conversions were not responsible for the positively selected regions detected in the Ig-like V-type1 domain of the human *SIGLEC7* and *SIGLEC9* genes. Our results suggest that the frequent gene conversions between CD33-related Siglec genes are simply a consequence of the high degree of sequence similarity of these genes and that they are not adaptive.

Key words: gene conversion, Siglec genes, positive selection, selective constraints, purifying selection.

Résumé : Les Siglecs sont des protéines qui lient les acides sialiques et appartiennent à la superfamille des immunoglobulines. Elles sont constituées de deux groupes, les Siglecs conservées et les Siglecs de type CD33. Des études antérieures ont suggéré que les conversions entres Siglecs de type CD33 des humains étaient adaptatives car elles augmentaient leur diversité. Ici, nous analysons les conversions géniques chez les gènes Siglec de cinq espèces de primates. Elles sont absentes entre les gènes Siglec conservés, mais elles sont fréquentes entre les gènes Siglec de type CD33. Ces conversions ne se produisent qu'entre gènes similaires et elles se produisent aussi souvent dans les régions qui lient les acides sialiques que dans celles qui ne les lient pas. Des tests de rapports dN/dS démontrent que la plupart des régions des gènes Siglec évoluent soit de façon neutre ou soit sous sélection purificatrice. Les gènes humains *SIGLEC7* et *SIGLEC9* contiennent des régions sélectionnées positivement, mais ces régions ne sont pas dues à des conversions géniques. Nos résultats suggèrent que la haute fréquence des conversions géniques entre les gènes Siglec de type CD33 est simplement une conséquence de la forte similarité de leurs séquences et que ces conversions ne sont pas adaptatives.

Mots-clés : conversion génique, gènes Siglec, sélection positive, contraintes sélectives, sélection purificatrice.

Introduction

Sialic acids binding immunoglobulin-like lectins (Siglecs) are a family of cell surface proteins belonging to the immunoglobulin superfamily. Siglecs are thought to be involved in cell–cell interactions and signalling functions such as self-recognition (Crocker 2002; Varki 2010). Siglecs are characterized by the presence of one or two N-terminal V-like immunoglobulin domain(s) followed by varying numbers of C2-set domains, a transmembrane domain, and a cytoplasmic domain (Crocker et al. 2007; Varki 2009; Fig. 1). Siglecs are subdivided into two groups based upon sequence similarity and evolutionary relatedness. The first group is called the conserved Siglecs because they are found in rodent and primate species. This group is composed of four genes: *SIGLEC1* (sialoadhesin, located on human chromosome 20p13), *SIGLEC2* (CD22, located on human chromosome 19q13.1), *SIGLEC4* (myelin-associated glycoprotein (MAG), located on human chromosome 19q13.1), and *SIGLEC15* (located on human chromosome 18q12.3, Fig. 1; Angata et al. 2004). Sialoadhesin is restricted to macrophages, CD22 to B cells and mast cells, MAG to glial cells, and *SIGLEC15* to macrophages, monocytes, and dendritic cells (Fig. 1). Members of this group are quite distantly related, sharing only about 25%–30% sequence similarity in their protein sequences (supplementary data, Table S1)[1]. The second group is called the CD33-related Siglecs (CD33rSiglecs) because they are similar to *SIGLEC3* (CD33). Members of this group are not conserved in all species (Angata et al. 2004). In humans, 11 hCD33rSiglec genes have been characterized. They are clustered on chromosome 19q13.3-13.4 and include *SIGLEC3*, *SIGLEC5*–*SIGLEC12*, *SIGLEC14*, and *SIGLEC16* (Fig. 2). In comparison, there are 12 CD33rSiglec genes in the chimpanzee genome and only 8 in the rhesus monkey genome (Fig. 2). Furthermore, *SIGLEC13* is present in the chimpanzee genome but not in the human genome (Angata et al. 2004). *SIGLEC12*, *SIGLEC14*, and *SIGLEC16* are partially pseudogenized in the human population (Angata et al. 2004; Cao et al. 2008; Wang et al. 2012). The protein sequences of human CD33rSiglecs share about 41%–88% identity,

Received 13 May 2014. Accepted 30 June 2014.

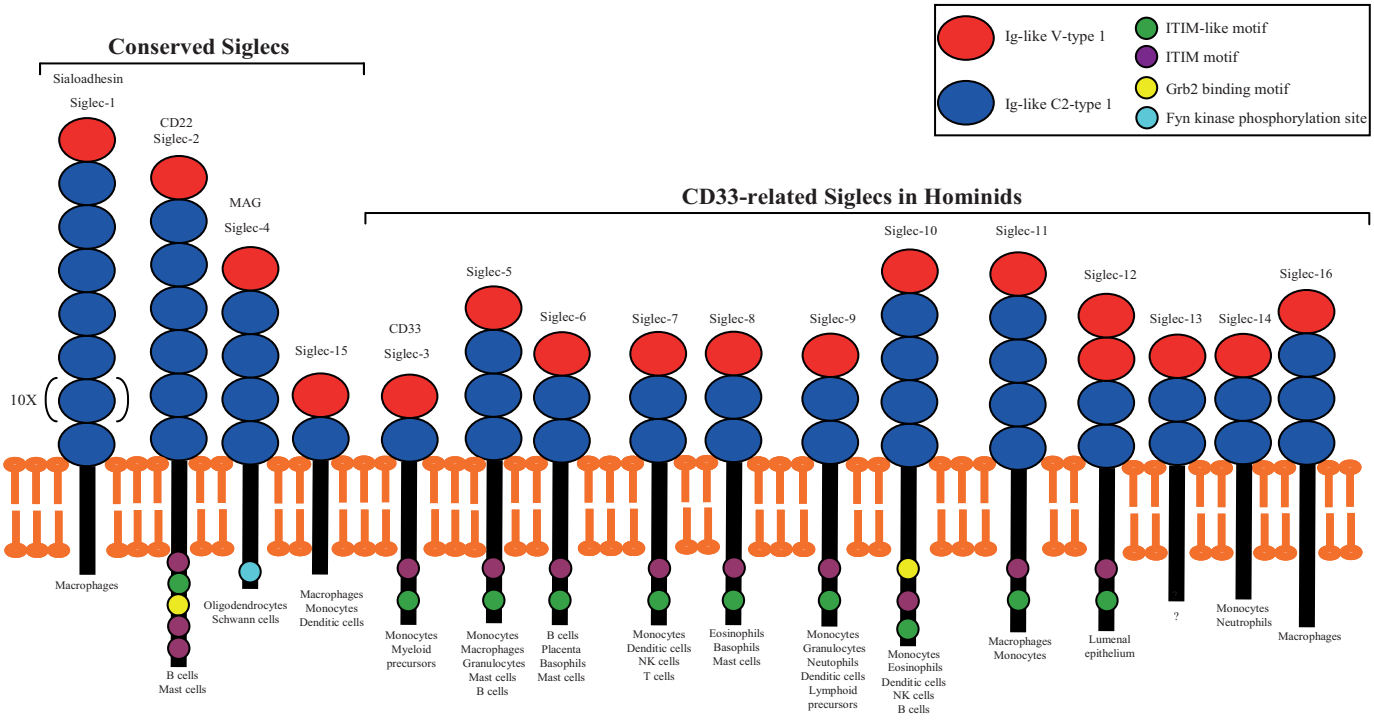
Corresponding Editor: B. Golding.

M. Zid and G. Drouin. Département de biologie et centre de recherche avancée en génomique environnementale, Université d'Ottawa, Ottawa, ON K1N 6N5, Canada.

Corresponding author: Guy Drouin (e-mail: gdrouin@science.uottawa.ca).

[1]Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/gen-2014-0083.

Fig. 1. Structure and tissue expression of Hominid Siglecs. Most of the information contained in this figure was obtained from Crocker and Varki (2001) and Jandus et al. (2011).



with the greatest similarity being in their extracellular region (Fig. 3; supplementary data, Table S1).

The CD33rSiglecs are widely distributed throughout in various cell types in the immune system of primates (Angata et al. 2004; Fig. 1). *SIGLEC3* and *SIGLEC5*–*SIGLEC11* are expressed on monocytes, NK cells, granulocytes cells, mast cells, dendritic cells, T cells, B cells, and macrophages (Crocker et al. 2007). *SIGLEC14*, *SIGLEC15*, and *SIGLEC16* have been discovered only recently, and their functions are poorly understood. In humans, *SIGLEC12* has lost its lectin activity (Jandus et al. 2011).

Gene conversions are the nonreciprocal exchange of genetic information between two genes. They have been observed is a wide variety of eukaryotes, from yeasts to mammals, and usually occur between sequences sharing a high degree of sequence similarity (Benovoy et al. 2005; Chen et al. 2007; McGrath et al. 2009). Although some gene conversions are known to be responsible for some human genetic diseases, most deleterious gene conversions are eliminated by purifying selection (Chen et al. 2007, 2010; Petronella and Drouin 2011 and references therein). Therefore, most gene conversions are selectively neutral, i.e., they are not eliminated by purifying selection because they represent selectively neutral events.

Previous studies have shown that gene conversions occur between some human CD33rSiglec genes. For example, Hayakawa et al. (2005) showed that a 2 kb long region, including exons 1–5, of the human *SIGLEC11* gene was converted by the human *SIGLEC16* pseudogene and these authors argued that this gene conversion was evolutionarily significant. Here, to address the evolutionary significance of gene conversions between Siglec genes, we studied the presence, location, and characteristics of the gene conversions that occurred between the Siglec genes of five primate species (human, chimpanzee, Sumatran orangutan, northern white-cheeked gibbon, and rhesus monkey). We found 33 gene conversion events between CD33rSiglec genes but none between conserved Siglec genes. Furthermore, all the conversions occurring in the coding regions of the CD33rSiglec genes occur only in the extracellular domain of these proteins. However, they occur equally frequently in the sialic acid binding and nonbinding domains of the proteins encoded by these genes and the length of the conversions is strongly correlated with sequence similarity. Our results therefore suggest that the high frequency of gene conversions between CD33-related Siglec genes is simply a consequence of their high degree of sequence similarity and is not the result of positive or diversifying selection.

Materials and methods

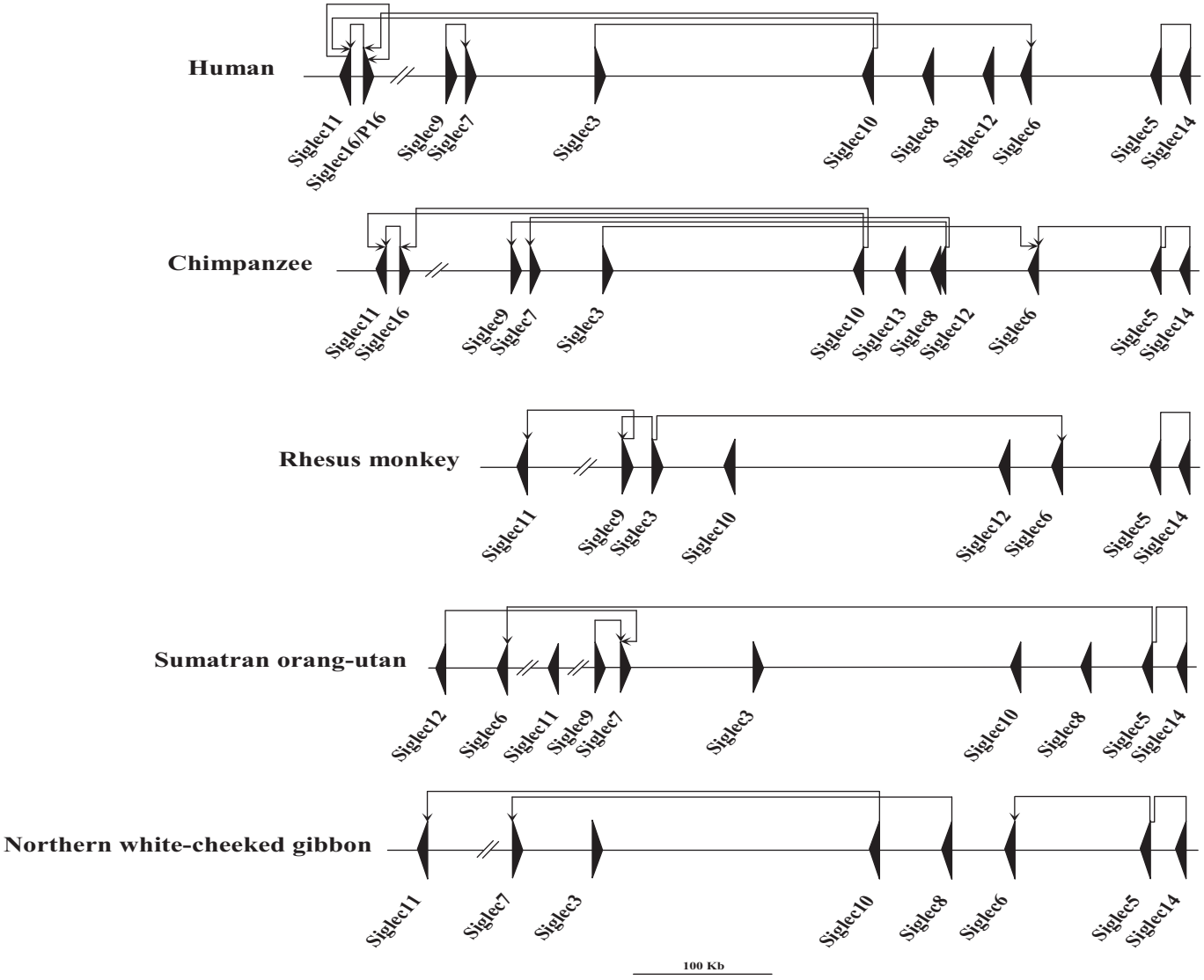
Gene sequences and their analyses

Sequences of five primate species were obtained from NCBI (http://www.ncbi.nlm.nih.gov/) and Ensembl (http://www.ensembl.org/). The list of sequences used, and their accession numbers, is shown in the supplementary data, Table S2. A zipped (*.zip) file containing the FASTA formatted sequences used is also available as supplementary data. Sequence alignments were performed using MUSCLE v3.8.31 (Edgar 2004) and then verified and refined using BioEdit v7.0.5.3 (Hall 1999). Sequence similarities were calculated using MEGA5 (Tamura et al. 2011). GC contents were calculated using Seqool (http://www.biossc.de). Standard deviations and *t* test were calculated using Excel 2010. Spearman rank correlations tests were calculated using R version 2.15 (R Development Core Team 2006).

Detection of gene conversion events

The GENECONV v.1.81 (Sawyer 1989) computer program was used to identify gene conversions. This method was chosen because it has one of the highest probabilities of correctly inferring gene conversions when they are present (Posada and Crandall 2001). The GENECONV program computes global and pairwise p values and allows mismatches within converted regions. Global and pairwise p values are calculated using two methods. The first method is based on 10 000 permutations of the original data, and the second is based on a method similar to that used by the BLAST database searching algorithm. Here, we only used p values from permutations because they are more conservative (Sawyer 1989). We also only considered p values ($p < 0.05$) from global inner fragments

Fig. 2. Organisation of CD33rSiglec genes and gene conversions in five primate species. Arrows indicate gene conversions and their direction. Note that, in humans, the *Siglec16* locus contains either a functional gene (*Siglec16*) or a pseudogene (*Siglec16P*).



because their *p* values are corrected for multiple comparisons, whereas the *p* values of pairwise fragments are not corrected for multiple comparisons. Given that gene conversions occur between the sequences of a given species, analyses were performed independently on the coding sequences of each species. These analyses were also performed using the g2 parameter, to allow mismatches within converted fragments and to take into account that substitutions do occur after conversions have occurred (Drouin 2002).

The directions of the conversions, i.e., from the donor gene to the acceptor gene, were determined based on the patterns of nucleotide variation inside and outside of the converted regions when compared with nonconverted sequences from closely related species. For example, if there was a gene conversion from nucleotides 4 to 7 between genes 1 and 2 of species A, with sequences CATCGGCTGT and AGCCGGCTAG, respectively, and that a related gene from a closely related species has the sequence CATCGGCTGT, or AGCTATATAG, then it is gene 1 of species A that converted gene 2 of species A and not the reverse.

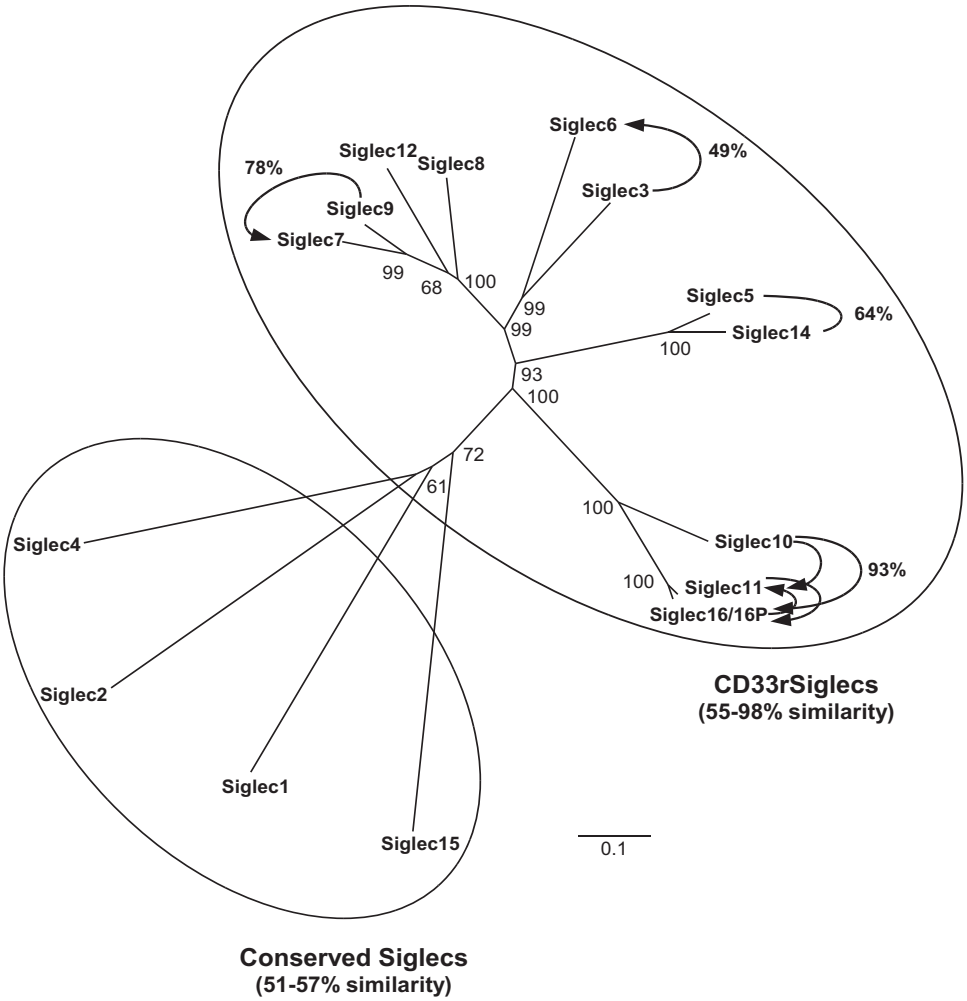
Phylogenetic analyses

Phylogenetic trees were built using the maximum likelihood method implemented in the PhyML v3.0 program, using the Blosum62 model of amino acid substitutions (Guindon and Gascuel 2003). Trees were visualized using the TreeView program (Page 1996). These trees were used to visualise the relationships between the human SIGLEC sequences and to map the gene conversion we detected. Since some Siglecs genes undergo frequent gene conversions (see below), the branch length of these trees do not necessarily reflect the actual divergence times between some of these duplicated genes. However, they do represent the degree of sequence similarity between them.

Tests of selection

To measure the selective pressures acting on different gene regions, the number of nonsynonymous substitutions per nonsynonymous site (dN) and the number of synonymous substitutions per synonymous site (dS) of Ig-like V-type 1 and Ig-like C2-type 1 domains were calculated using the maximum likelihood method implemented in the codeml program of the PAML software package version 4.5 (Yang 2007). These values were calculated using the options seqtype = 1, runmode = –2, CodonFreq = 2, and fix_omega = 0 in the codeml.ctl files of this package. We calculated dN and dS values between human, chimpanzee, and Sumatran orangutan sequences because these three genomes share the largest number of orthologous genes and that such orthologous genes are necessary to obtain meaningful dN and dS values.

Fig. 3. Phylogenetic tree of human Siglecs proteins. The scale bar represents 10% difference. Percent (%) values indicate the percentage sequence identity of the DNA sequences between the genes indicated by arrows. Arrows indicate gene conversions and their direction. Numbers at the nodes are bootstrap values.



We tested whether the dN/dS ratios we obtained were significantly different from 1 by calculating the likelihood of these ratios being equal to 1 and performing a likelihood ratio test against a χ^2 distribution with one degree of freedom. The values used for these likelihood ratio tests were calculated as twice the difference between the log likelihood of the calculated dN/dS ratios and the log likelihood of these ratios being equal to 1.

Results

Number, lengths, and polarity of gene conversions

We detected 33 statistically significant gene conversion events in the five species we analyzed: 11 in human, 9 in chimpanzee, 4 in Sumatran orangutan, 5 in northern white-cheeked gibbon, and 4 in rhesus monkey (Table 1). FASTA formatted sequence files of these converted regions are available as supplementary data. All of these conversions are between CD33rSiglec genes and none are between conserved Siglec genes. A few of these conversions are found in more than one species. For example, a conversion of about 200 bp is found in the region coding for the C2-type 1 domain of the human, chimpanzee, and rhesus monkey *SIGLEC3* and *SIGLEC6* genes. In all three cases, *SIGLEC3* converted *SIGLEC6* (Table 1; Fig. 2). Furthermore, a conversion ranging from 503 to 1060 bp, and starting in exon 1, is found between the *SIGLEC5* and *SIGLEC14* genes of all five species (Table 1). Unfortunately, the high degree of sequence similarity of these two genes in these regions, both within and between species, makes it impossible to determine the polarity of these conversions (Table 1; Fig. 2; further results not shown). We also analyzed whether there were gene conversions between the 15 human Siglec pseudogenes (*SIGLEC16P*–*SIGLEC30P*) or between these pseudogenes and functional human Siglec genes. The fact that we did not detect any such conversion is likely due to the fact that these pseudogene sequences are very different from one another and from functional Siglec genes (results not shown).

The average length (± standard deviation) of all conversions is 590.21 ± 429.34 nucleotides and they range from 100 to 2286 nucleotides long. Of these 33 gene conversions, 12 started at an exon and ended at another exon (average size = 822.17 ± 671.55 nucleotides), 7 started at an intron and ended at another intron (average size = 185.43 ± 53.79 nucleotides), and 14 started at an intron and ended at exon, or vice versa (average size = 593.78 ± 279.07 nucleotides).

As shown in Fig. 2, of the 11 CD33rSiglec genes involved in gene conversion events, 4 are donor only (*SIGLEC3*, *SIGLEC8*, *SIGLEC10*, and *SIGLEC12*), 2 are acceptor only (*SIGLEC6* and *SIGLEC7*), and 4 are both acceptor and donor (*SIGLEC9*, *SIGLEC11*, *SIGLEC16*, and *SIGLEC16P*). The *SIGLEC13* gene is only found in the chimpanzee genome and it was never involved in gene conversions. Interestingly, gene conversions occur between both adjacent genes and distant genes (Fig. 2). In fact, in humans, there is no correlation between the distance between genes and the frequency of conversions (Spear-

Table 1. Gene conversions, and their location, in the CD33rSiglec genes of five primate species.

			Location			
			Genomic DNA (bp)		Protein (domains)	
Species	Donor (D) and acceptor (A) genes	Length (bp)	From	To	From	To
Homo sapiens (Human)	Siglec-11 (A) /Siglec-16P* (D)	2286	Exon1 (134)	Exon7 (2420)	Signal peptide	C2-type 3
	Siglec-11 (D) /Siglec-16 (A)	1628	Exon1 (92)	Exon5 (1720)	Signal peptide	C2-type 2
	Siglec-10 (D) /Siglec-16P* (A)	847	Exon2 (708)	Intron4 (1555)	V-type 1	C2-type 1
		694	Intron4 (1684)	Exon6 (2378)	C2-type 3	C2-type 3
	Siglec-10 (D) /Siglec-16 (A)	235	Exon2 (703)	Intron2 (938)	V-type 1	V-type 1
		427	Exon5 (1819)	Intron6 (2246)	C2-type 2	C2-type 3
	Siglec-11 (A) /Siglec-10 (D)	852	Exon2 (611)	Intron4 (1463)	C2-type 1	C2-type 1
		1236	Exon5 (1599)	Exon8 (2835)	C2-type 3	C2-type 3
	Siglec-14 (?) /Siglec-5 (?)	1060	Exon1 (121)	Intron3 (1181)	Signal peptide	C2-type 1
	Siglec-6 (A) /Siglec-3 (D)	208	Exon3 (960)	Exon3 (1168)	C2-type 1	C2-type 1
	Siglec-7 (A) /Siglec-9 (D)	145	Intron1 (1985)	Exon2 (2130)	C2-type 1	C2-type 1
Pan troglodytes (Chimpanzee)	Siglec-11 (A) /Siglec-16 (D)	2168	Exon1 (281)	Exon7 (2449)	Signal peptide	C2-type 3
	Siglec-10 (D) /Siglec-16 (A)	644	Exon6 (2085)	Exon8 (2729)	C2-type 2	C2-type 3
	Siglec-11 (A) /Siglec-10 (D)	685	Exon5 (1611)	Exon7 (2296)	C2-type 2	C2-type 3
		417	Exon7 (2461)	Intron8 (2878)	C2-type 3	C2-type 3
	Siglec-6 (A) /Siglec-3 (D)	192	Exon3 (1017)	Intron3 (1209)	C2-type 1	C2-type 1
	Siglec-9 (A) /Siglec-12 (D)	200	Intron5 (4691)	Intron5 (4891)	NA	NA
	Siglec-7 (A) /Siglec-12 (D)	131	Intron3 (4274)	Intron3 (4405)	NA	NA
	Siglec-5 (?) /Siglec-14 (?)	503	Exon1 (126)	Intron2 (629)	Signal peptide	V-type 1/C2-type 1
	Siglec-6 (A) /Siglec-5 (D)	199	Exon5 (1905)	Exon5 (2104)	C2-type 2	C2-type 2
Macaca mulatta (Rhesus monkey)	Siglec-9 (D) /Siglec-11 (A)	137	Intron6 (6623)	Intron6 (6760)	NA	NA
	Siglec-6 (A) /Siglec-3 (D)	235	Exon4 (1034)	Exon4 (1269)	C2-type 1	C2-type 1
	Siglec-14 (?) /Siglec-5 (?)	1022	Exon1 (21)	Intron5 (1043)	Signal peptide	V-type 1/C2-type 1
	Siglec-9 (A) /Siglec-3 (D)	100	Intron6 (6447)	Intron6 (6547)	NA	NA
Pongo abelii (Sumatran orangutan)	Siglec-14 (D) /Siglec-5 (A)	1029	Exon1 (115)	Intron3 (1144)	Signal peptide	C2-type 1
	Siglec-6 (A) /Siglec-5 (D)	116	Exon5 (1864)	Exon5 (1980)	C2-type 2	C2-type 2
	Siglec-7 (A) /Siglec-9 (D)	209	Intron1 (2451)	Intron1 (2660)	NA	NA
	Siglec-7 (A) /Siglec-12 (D)	268	Intron4 (10243)	Intron4 (10511)	NA	NA
Nomascus leucogenys (Northern white-cheeked gibbon)	Siglec-10 (D) /Siglec-11 (A)	296	Exon2 (544)	Exon3 (840)	C2-type 1	C2-type 1
		284	Exon5 (1543)	Intron5 (1827)	C2-type 2	C2-type 3
	Siglec-5 (?) /Siglec-14 (?)	606	Exon1 (245)	Intron2 (851)	Signal peptide	C2-type 1
	Siglec-5 (D) /Siglec-6 (A)	165	Exon4 (2348)	Exon4 (2513)	C2-type 2	C2-type 2
	Siglec-7 (A) /Siglec-8 (D)	253	Intron5 (4808)	Intron5 (5061)	NA	NA

Note: NA, not applicable. Asterisks (*) indicate that this sequence is a pseudogene. Interrogation points (?) indicate conversions for which the polarity could not be determined.

man's rank correlation test, $\rho = -0.16$, $p = 0.23$). The same is true for the other four species (results not shown). However, there is a weak, but significant, correlation between the distance between genes and the frequency of conversions when considering the data from all species (Spearman's rank correlation test, $\rho = -0.15$, $p = 0.03$). The proximity of genes therefore only explains some 2% of the variation in gene conversion frequency.

Sequence similarities and GC content of converted and nonconverted regions

For the primate CD33rSiglec genes, there is a strong and significant correlation between the length of the converted regions and the sequence similarity of the converted regions ($r = 0.52$, $p = 0.002$). Furthermore, the average similarity (± standard deviation) between the converted regions (94.85% ± 3.27%) is also significantly higher than the average similarity between the nonconverted regions (61.09% ± 13.74%, *t* test, $p = 8.68 \times 10^{-16}$). Moreover, there is also a significant difference in GC content between converted and nonconverted regions. The average GC content (± standard deviation) of the converted regions is 61.4% ± 3.18% whereas that of the nonconverted regions is 50% ± 4.17% (*t* test, $p = 2.08 \times 10^{-18}$).

Biased distribution of converted regions

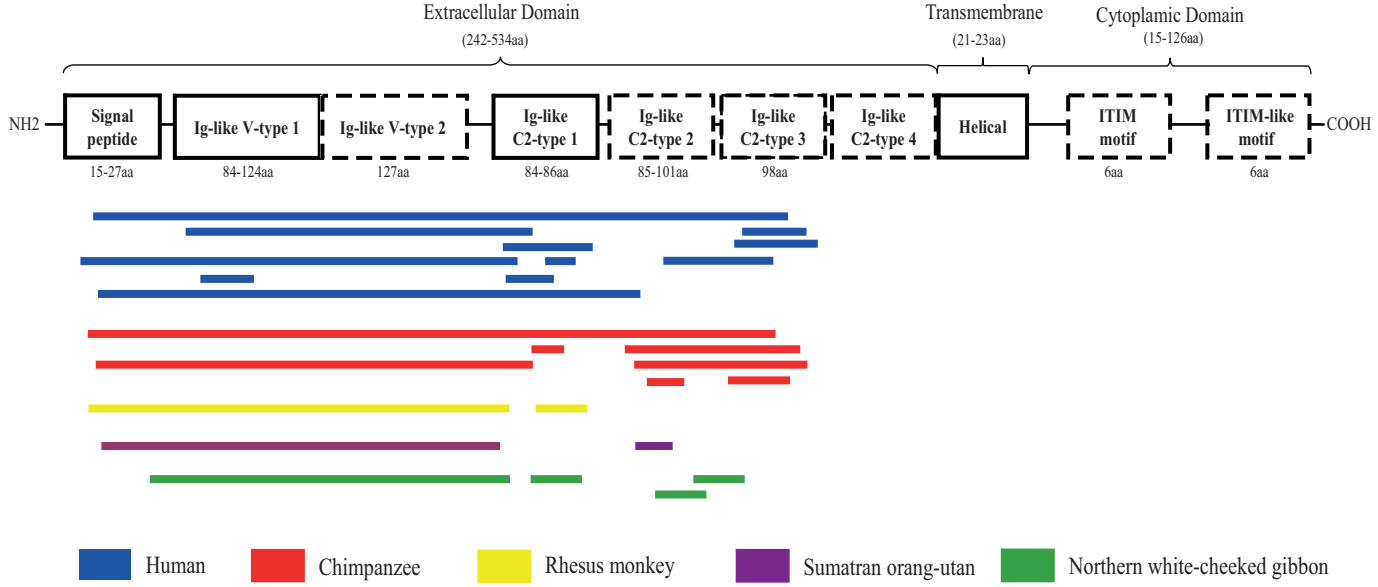
Except for *SIGLEC12*, all CD33rSiglec genes contain a single Ig-like V-type 1 domain followed by 1–16 Ig-like C2-type domains (Fig. 1). These two types of domains constitute the extracellular portion of these proteins, the rest being made up of a transmembrane and a cytoplasmic domain. Interestingly, all of the conversions affecting the protein coding regions are only found in the gene regions coding for the extracellular part of these proteins (Table 1; Fig. 4). This suggests that the conversions that are present in the extracellular part of these proteins are not deleterious. The absence of conversions in the regions coding for the transmembrane and cytoplasmic domains of these proteins is likely due to the low degree of sequence similarity in these regions (see below).

Tests of selection

Table 2 shows the dN and dS values in the Ig-like V-type 1 domain and Ig-like C2-type 1 domain in the conserved Siglec genes and in the CD33rSiglecs genes. The unusually large dN and dS values observed for the comparisons involving the chimpanzee *SIGLEC4* sequence are likely the result of the fact that this GenBank sequence contains sequencing or annotation mistakes.

All but two of the statistically significant dN/dS ratios of the Ig-like V-type 1 and Ig-like C2-type 1 domains are smaller than 1 (Table 2). This suggests that, in most Siglecs, both these domains evolve either neutrally (when dN/dS ratios are not significantly different from 1) or by purifying (negative) selection (when dN/dS ratios are significantly smaller than 1). The two ratios that are significantly greater than 1, that of the human and chimpanzee *SIGLEC7* Ig-like V-type 1 domains and that of the human and orangutan *SIGLEC9* Ig-like V-type 1 domains, suggest that they evolve by positive selection. However, these two positively selected regions

Fig. 4. Schematic representation of the structure of CD33rSiglec proteins and location of the gene conversions detected between the CD33rSiglec genes of five primate species. The different domains of the proteins are indicated by boxes. Boxes with a full line represent domains found in all Siglec proteins whereas those with a dashed line represent domains found only in some Siglec proteins (see Fig. 1). The range of lengths of these different domains, in amino acids, is indicated below each box. Only conversions affecting protein coding regions are indicated.



were not generated by gene conversions because gene conversions did not occur in their Ig-like V-type 1 domains. The human *SIGLEC7* conversion occurred in the Ig-like C2-type 1 domain region, and the V-type 1 domain of the five primate *SIGLEC9* genes was never converted (Table 1).

Discussion

Our results show that gene conversions are frequent between CD33rSiglecs genes but that they are not observed between conserved Siglec genes (Table 1). This suggests that gene conversions do not occur between the conserved genes or that, if they do occur, they are removed by purifying selection. Of these two possibilities, the former is likely responsible for the absence of gene conversions between conserved Siglec genes. As can be seen in Fig. 3, the sequences of the conserved Siglecs are much more divergent from one another than those of CD33rSiglecs. In fact, at the protein level, the conserved Siglecs share only an average of 21% similarity between one another (supplementary data, Table S1). At the DNA level, the conserved Siglecs share only an average of 53% similarity between one another (Fig. 3). Since gene conversions are more frequent between more similar genes, and that gene conversions are very infrequent between genes sharing less than 80% nucleotide sequence similarity, the low similarity between conserved Siglec genes, and between conserved Siglec and CD33rSiglec genes, is likely responsible for the absence of gene conversions in conserved Siglec genes (Benovoy et al. 2005; Chen et al. 2007; Benovoy and Drouin 2009). This suggestion is also consistent with the pattern of gene conversions observed between CD33rSiglec genes where gene conversions occur almost exclusively between related genes. For example, in humans, the eight conversions between CD33rSiglec genes occur exclusively between closely related genes, i.e., the *SIGLEC7* and *SIGLEC9* genes, the *SIGLEC3* and *SIGLEC6* genes, the *SIGLEC5* and *SIGLEC14* genes, and the *SIGLEC10*, *SIGLEC11*, and *SIGLEC16* genes (Table 1; Fig. 3). Apart from the 199-bp conversion between the *SIGLEC5* and *SIGLEC6* genes of chimpanzee, the same pattern is observed for the gene conversions between chimpanzee genes; gene conversions occur exclusively between closely related genes. Therefore, gene conversion are limited to very similar genes that are part of the same CD33rSiglec subfamilies. This is supported by the fact that there is a strong positive correlation between the length of the conversions and the similarity of the converted regions ($r = 0.52$, $p = 0.002$).

Another interesting result is that, in CD33rSiglec genes, all gene conversions that span exon sequences are found exclusively in the extracellular region of the protein coded by these genes, and never in their transmembrane or cytoplasmic regions (Table 1; Fig. 4). Again, this bias is most readily explained by the fact that gene conversions only occur between sequences sharing at least 80% similarity (see above). Since the similarity (and standard deviation) of the transmembrane and cytoplasmic regions (35.7% ± 19.6%) is significantly smaller than that of extracellular regions (56.4% ± 13.4%, *t* test, $p = 4.2 \times 10^{-9}$), gene conversions are unlikely to occur in transmembrane and cytoplasmic regions (supplementary data, Table S1). Furthermore, the frequent conversions between the extracellular region of the *SIGLEC5* and *SIGLEC14* genes, and the long conversions between the extracellular regions of the *SIGLEC11* and *SIGLEC16* genes, are likely simply the result of the fact that the extracellular regions of these two pairs of genes are 96% identical at the protein level, and 95% similar at the DNA level (supplementary data, Table S1; Table 1; Fig. 3). The effect of sequence similarity on the occurrence of gene conversions is also evident when one compares the human CD33rSiglec genes that converted one another with those that did not. At the protein level, the average similarity (± standard deviation) between the extracellular regions of the genes that did convert one another (i.e., the *SIGLEC3* and *SIGLEC6*, *SIGLEC5* and *SIGLEC14*, *SIGLEC7* and *SIGLEC9*, *SIGLEC10* and *SIGLEC11*, *SIGLEC10* and *SIGLEC16*, and *SIGLEC11* and *SIGLEC16* gene pairs, 83.7% ± 11.2%) is significantly higher than the average similarity of the extracellular regions between the genes that did not convert one another (i.e., all other gene pairs, 54.4% ± 11.3%, *t* test, $p = 0.0007$; supplementary data, Table S1).

Our data also suggest that the higher similarity caused by conversions favours the occurrence of more conversions. This suggestion, that gene conversions have occurred repeatedly during the evolution of the extracellular domain of CD33rSiglecs, is supported by the fact the GC content of the converted regions is

Table 2. dN and dS values for Ig-like V-type 1 and Ig-like C2-type 1 domains.

		Ig-like V-type 1 domain			Ig-like C2-type 1 domain		
Genes	Pairwise species	dN	dS	dN/dS	dN	dS	dN/dS
Conserved Siglecs							
Siglec1	Human/Chimpanzee	0.0068	0.0217	0.31	0.0045	0.0200	0.28
	Human/Orangutan	0.0216	0.1332	0.16**	0.0138	0.0198	0.70
	Chimpanzee/Orangutan	0.0212	0.1116	0.19*	0.0092	0.0396	0.24
Siglec2	Human/Chimpanzee	0.0138	0.0001	138	0	0	0
	Human/Orangutan	0.0188	0.0371	0.51	0.0371	0.0132	2.81
	Chimpanzee/Orangutan	0.0181	0.0423	0.43	0.0371	0.0132	2.81
Siglec4	Human/Chimpanzee	0.1565	0.4614	0.34*	0.99	93.4478	0.01***
	Human/Orangutan	0.0045	0.1238	0.03***	0.0041	0.2132	0.02***
	Chimpanzee/Orangutan	0.1618	0.6644	0.24***	1.0124	87.4077	0.01***
Siglec15	Human/Chimpanzee	0.0068	0.0001	68	0.0101	0.0409	0.25
	Human/Orangutan	0.0067	0.04	0.17	0.0203	0.1167	0.17*
	Chimpanzee/Orangutan	0	0.0479	0**	0.0204	0.1189	0.17*
CD33rSiglecs							
Siglec3	Human/Chimpanzee	0.0324	0.0109	2.97	0.0106	0.0167	0.63
	Human/Orangutan	0.0414	0.0432	0.96	0.0326	0.0166	1.96
	Chimpanzee/Orangutan	0.0544	0.0319	1.70	0.0213	0.0002	106.5
Siglec5	Human/Chimpanzee	0.0394	0.0116	3.40	0.0158	0.0171	0.92
	Human/Orangutan	0.1003	0.0686	1.46	0.0209	0.0562	0.37
	Chimpanzee/Orangutan	0.0913	0.0828	1.10	0.0156	0.0373	0.42
Siglec6	Human/Chimpanzee	0.0053	0.0001	53	0.005	0.0674	0.07*
	Human/Orangutan	0.0649	0.0752	0.86	0.0422	0.1076	0.39
	Chimpanzee/Orangutan	0.0707	0.0743	0.95	0.0367	0.0825	0.44
Siglec7	Human/Chimpanzee	0.0277	0.0148	1.87***	0	0.0224	0.001
	Human/Orangutan	0.0279	0.0295	0.95	0	0.0224	0.001
	Chimpanzee/Orangutan	0.0276	0.0456	0.61***	0	0	0.06
Siglec8	Human/Chimpanzee	0.0056	0.0001	56	0.0210	0.0353	0.59
	Human/Orangutan	0.0155	0.0193	0.80	0.0322	0.0919	0.35
	Chimpanzee/Orangutan	0.0212	0.0181	1.17	0.0318	0.0543	0.59
Siglec9	Human/Chimpanzee	0.0285	0.0003	95	0.0060	0.0623	0.10
	Human/Orangutan	0.1146	0.0162	7.07*	0.0220	0.0840	0.26
	Chimpanzee/Orangutan	0.0867	0.0169	5.13	0.0291	0.0780	0.37
Siglec10	Human/Chimpanzee	0.0047	0.0247	0.19	0.0146	0.0207	0.71
	Human/Orangutan	0.0172	0.0764	0.23***	0.0151	0.0384	0.39
	Chimpanzee/Orangutan	0.0218	0.0752	0.29***	0.0100	0.0184	0.55
Siglec11	Human/Chimpanzee	0.0119	0.037	0.32	0.005	0.0566	0.09*
	Human/Orangutan	0.0357	0.0561	0.64	0.0155	0.0559	0.28
	Chimpanzee/Orangutan	0.043	0.0477	0.90	0.0204	0.0382	0.53
Siglec12	Human/Chimpanzee	0.0116	0.0001	116	0.0199	0.0665	0.30
	Human/Orangutan	0.0418	0.0551	0.76	0.0253	0.1569	0.16**
	Chimpanzee/Orangutan	0.0382	0.0533	0.72	0.0257	0.1264	0.20*
Siglec14	Human/Chimpanzee	0.0394	0.0116	3.40	0.0104	0.0175	0.60
	Human/Orangutan	0.0924	0.081	1.14	0.0103	0.077	0.13*
	Chimpanzee/Orangutan	0.0834	0.0949	0.88	0.0104	0.0934	0.11**
Siglec16	Human/Chimpanzee	0.0073	0.0587	0.12**	0.0101	0.0374	0.27

Note: *Siglec13* is not present in the human genome. *Siglec16* is only present in the human and chimpanzee genomes. *, $p < 0.5$; **, $p < 0.1$; ***, $p < 0.01$.

significantly higher than that of nonconverted regions. Several studies have shown that increased recombination, such as gene conversions, leads to increases in GC content (Galtier et al. 2001; Birdsell 2002; Meunier and Duret 2004; Benovoy et al. 2005). Therefore, the fact that converted regions have significantly higher GC contents than nonconverted regions (61.4% versus 49.4%, $p = 3.95 \times 10^{-18}$), suggests that repeated recombination events occurred in the converted regions.

Our results show that gene conversions are indeed very frequent between CD33rSiglecs. Given that we observed 11 conversions in humans, 9 in chimpanzees, 4 in Sumatran orangutan, 4 in rhesus monkey, and 5 in northern white-cheeked gibbon (Table 1) and that these species have, respectively, 11, 12, 10, 7, and 8 CD33rSiglec genes (Fig. 2), the frequency of gene conversions between the CD33rSiglec genes of these respective species, calculated as number of conversions per number of gene pairs compared, is therefore 20%, 14%, 9%, 19%, and 18%. This is much higher than the average frequency of gene conversion events between human genes, which we calculated to be 0.88% using the same methodology as we used in this study (Benovoy and Drouin 2009).

As mentioned above, Hayakawa et al. (2005) previously observed the ~2 kb long gene conversion present between the human *SIGLEC11* gene and the *SIGLEC16P* pseudogene and they suggested that this conversion was potentially adaptive (Table 1). Note that this conversion is not present in other primate species because the *SIGLEC16P* pseudogene (allele) is only found in human genomes (Fig. 2). Wang et al. (2012) also recently suggested that this conversion is evolutionary significant, but they suggested that this conversion was in fact made up of two tandem gene conversions. Our results support both the presence of a conversion between these two sequences and the fact that it might be composed of two tandem gene conversions (Table 1; further results not shown). In fact, although the genomic DNA (i.e., including exons and introns) of the coding regions of these two sequences are only 70% similar, the 2286 bp contained in the converted region are 99% similar and this region is followed by a region of 490 bp with 88% similarity

between these two sequences. This second region of high similarity might therefore represent the remnants of an older conversion. Our results also complement those of previous studies by showing that gene conversions are not limited to those between the human *SIGLEC11* and *SIGLEC16P* sequences. In fact, in both humans and chimpanzees, gene conversions also occurred between the *SIGLEC10* and *SIGLEC11*, the *SIGLEC10* and *SIGLEC16*, and the *SIGLEC11* and *SIGLEC16* gene pairs (Table 1).

The conversion between the 5′-end of the *SIGLEC5* and *SIGLEC14* genes has previously been detected in five primate species (Angata et al. 2006). We extend this finding by showing that it is also present in the northern white-cheeked gibbon (Table 1; Fig. 2). Again, given that this conversion is found in the same region in all species, it likely occurred in their common ancestor. However, given that the increased similarity brought about by this initial gene conversion event facilitates further conversion events, and, as discussed above, that converted regions have significantly higher GC content, suggests that multiple subsequent conversion events occurred independently in the same region in some or all these species. The conversion between the *SIGLEC10* and *SIGLEC11* genes of human and chimpanzee was also previously reported in humans (Angata et al. 2002; Cao and Crocker 2011; Table 1).

Positive selection

Are the frequent gene conversions present in the extracellular region of CD33rSiglec genes positively selected (adaptive) or not? Since the extracellular region of CD33rSiglecs is exposed outside the cells, previous studies have suggested that the gene conversions occurring in this region were subject to positive selection (Angata et al. 2004; Crocker et al. 2007; Varki 2010). Furthermore, since the V-type 1 Ig domain binds sialic acid, it has been suggested that positive selection should act on V-type 1 Ig domains and not the adjacent C2-type 1 Ig domains (Altheide et al. 2006; Fig. 1). This prediction has been convincingly confirmed for the *SIGLEC9* gene (Sonnenburg et al. 2004). Therefore, if the gene conversions observed between CD33rSiglec genes were selected to increase sequence diversity, we would expect them to occur between the most divergent gene sequences. We would also expect them to be more frequent within V-type 1 domains than within the adjacent C2-type 1 domains. As discussed above, our results show that there is a strong correlation between sequence similarity and the length of the converted regions. In other words, gene conversions only occur between very similar sequences. Furthermore, gene conversions are not more frequent within V-type 1 Ig domains than within the adjacent C2-type 1 Ig domains. In fact, if we consider the data for all primate species, there are 10 gene conversions within V-type 1 domains and 12 within the adjacent C2-type 1 domains (Table 1; Fig. 4). These observations therefore do not support the hypothesis that diversifying selection is responsible for gene conversions observed between CD33rSiglec genes. Since we cannot reject the neutral hypothesis that gene conversions have no selective impact on the evolution of CD33rSiglec genes, our results suggest that the gene conversions observed between them are selectively neutral.

The fact that gene conversions occur equally frequently in the Ig-like V-type 1 and the Ig-like C2-type 1 domains of CD33rSiglec genes suggests that both these regions evolve under similar selective constraints. We therefore tested the claim that the Ig-like V-type 1 domains of Siglec genes evolve under positive selection and that the Ig-like C2-type 1 domains do not. Our results clearly show that, except for the Ig-like V-type 1 domain of the *SIGLEC7* and *SIGLEC9* genes, both the Ig-like V-type 1 and the Ig-like C2-type 1 domains of Siglec genes do not evolve under positive selection because the dN/dS ratios of these domains are either not significantly different than 1 or are significantly smaller than 1 (Table 2). Furthermore, the conversion between the human *SIGLEC7* and *SIGLEC9* genes involved the conversion of the C2-type 1 domain of the *SIGLEC7* gene by the C2-type 1 domain of the *SIGLEC9* gene (Table 1). This conversion was therefore not involved in generating the positively selected V-type 1 domain of the human and chimpanzee *SIGLEC7* genes (Table 1). Similarly, the positively selected V-type 1 domain of *SIGLEC9* was also not generated by a gene conversion because the *SIGLEC9* gene was not converted by any other gene (Tables 1 and 2). This suggests that the regions that were converted evolve neutrally or under purifying selection. These results are contrary to those of previous studies that suggested that gene conversions between Siglec genes are adaptive (Angata et al. 2004; Altheide et al. 2006; Varki 2010; Jandus et al. 2011). This discrepancy is mainly the result of the fact that previous studies did not perform detailed gene conversion analyses to address whether gene conversions where creating positively selected regions. Previous studies also sometimes did not assess the statistical significance of the dN/dS ratios they calculated, used inappropriate methodologies to calculate these ratios, concatenated numerous sequences, and mistakenly interpreted higher dN/dS ratios smaller than 1 as representing positive selection rather than relaxed purifying selection.

Although our results suggest that gene conversions are not generating positively selected regions, we cannot rule out that a few sites within these regions are under positive selection. Since our dN/dS ratio analyses are performed on domains ranging in size from 84 to 124 codons (Fig. 4), these analyses would not detect positive selection of a few sites within a region composed mainly of sites evolving under purifying selection. However, since most of the gene conversions we detected span all or most of the V-type 1 and C2-type 1 domains our analyses should have detected positive selection generated by gene conversions because these conversions would have affected all or most sites within these regions (Table 1; Fig. 4).

In conclusion, our results suggest that the evolution of Siglec genes is different from the currently accepted view, which posits that gene conversions between Siglec genes are adaptive (Angata et al. 2004; Altheide et al. 2006; Varki 2010; Jandus et al. 2011). If the gene conversion we detected were adaptive, one would expect that there would be conversions between the conserved Siglec genes. The fact that there are none suggests that it is not adaptive. We suggest that the absence of gene conversions is simply the result of the fact that these genes are too different from one another, and from the CD33rSiglec genes, to convert them or be converted by them. Conversely, the fact that conversions are frequent between closely related CD33rSilec genes is likely the result of the fact that they are very similar to one another. This suggestion is supported by the fact that conversions occur almost exclusively between CD33rSiglec genes having similar sequences (supplementary data, Table S1; Table 1; Fig. 3). It is also supported by the fact that gene conversions only occur in highly similar regions of CD33rSiglecs (supplementary data, Table S1; Table 1; Fig. 3). Furthermore, the fact that gene conversions are not more frequent within V-type 1 Ig domains than within the adjacent C2-type 1 Ig domains suggest that these conversions are not subject to positive selection. Finally, the positive selection we detected in the Ig-like V-type1 domains of the human *SIGLEC7* and *SIGLEC9* genes (Table 2) was not generated by gene conversions because the conversions between these genes occurred outside the Ig-like V-type1 domains (Table 1). Therefore, the frequent conversions observed between primate CD33rSiglec genes likely represent neutral events that are not selected against. This conclusion is consistent with our recent analyses of the carcinoembryonic antigen family of five primate species (Zid and Drouin 2013). This study reported that the frequent conversions between these genes of the immunoglobulin superfamily also did not generate positively selected regions and were not involved in increasing sequence diversity. They therefore also likely represent selectively neutral events between genes having similar sequences and functions.

Acknowledgements

We thank the three anonymous referees for their constructive comments. We also thank Stéphane Aris-Brosou (Biology Department, University of Ottawa) for his help with the PAML software. This work was supported by a discovery grant from the Natural Science and Engineering Research Council of Canada to G.D. M.Z. also received a tuition fee exemption scholarship from the Ministry of Higher Education, Scientific Research and Technology of the Republic of Tunisia.

References

Altheide, T.K., Hayakawa, T., Mikkelsen, T.S., Diaz, S., Varki, N., and Varki, A. 2006. System-wide genomic and biochemical comparisons of sialic acid biology among primates and rodents: evidence for two modes of rapid evolution. J. Biol. Chem. **281**: 25689–25702. doi:10.1074/jbc.M604221200. PMID:16769723.

Angata, T., Kerr, S.C., Greaves, D.R., Varki, N.M., Crocker, P.R., and Varki, A. 2002. Cloning and characterization of human Siglec-11. A recently evolved signaling molecule that can interact with SHP-1 and SHP-2 and is expressed by tissue macrophages, including brain microglia. J. Biol. Chem. **277**: 24466–24474. PMID:11986327.

Angata, T., Margulies, E.H., Green, E.D., and Varki, A. 2004. Large-scale sequencing of the CD33-related Siglec gene cluster in five mammalian species reveals rapid evolution by multiple mechanisms. Proc. Natl. Acad. Sci. U.S.A. **101**: 13251–13256. doi:10.1073/pnas.0404833101. PMID:15331780.

Angata, T., Hayakawa, T., Yamanaka, M., Varki, A., and Nakamura, M. 2006. Discovery of Siglec-14, a novel sialic acid receptor undergoing concerted evolution with Siglec-5 in primates. FASEB J. **20**: 1964–1973. doi:10.1096/fj.06-5800com. PMID:17012248.

Benovoy, D., and Drouin, G. 2009. Ectopic gene conversions in the human genome. Genomics, **93**: 27–32. doi:10.1016/j.ygeno.2008.09.007. PMID:18848875.

Benovoy, D., Morris, R.T., Morin, A., and Drouin, G. 2005. Ectopic gene conversions increase the G + C content of duplicated yeast and *Arabidopsis* genes. Mol. Biol. Evol. **22**: 1865–1868. doi:10.1093/molbev/msi176. PMID:15917495.

Birdsell, J.A. 2002. Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. Mol. Biol. Evol. **19**: 1181–1197. doi:10.1093/oxfordjournals.molbev.a004176. PMID:12082137.

Cao, H., and Crocker, P.R. 2011. Evolution of CD33-related Siglecs: regulating host immune functions and escaping pathogen exploitation? Immunology, **132**: 18–26. doi:10.1111/j.1365-2567.2010.03368.x. PMID:21070233.

Cao, H., Lakner, U., de Bono, B., Traherne, J.A., Trowsdale, J., and Barrow, A.D. 2008. *SIGLEC16* encodes a DAP12-associated receptor expressed in macrophages that evolved from its inhibitory counterpart *SIGLEC11* and has functional and non-functional alleles in humans. Eur. J. Immunol. **38**: 2303–2315. doi:10.1002/eji.200738078. PMID:18629938.

Chen, J.-M., Cooper, D.N., Chuzhanova, N., Férec, C., and Patrinos, G.P. 2007. Gene conversion: mechanisms, evolution and human disease. Nat. Rev. Genet. **8**: 762–775. doi:10.1038/nrg2193. PMID:17846636.

Chen, J.-M., Férec, C., and Cooper, D.N. 2010. Gene conversion in human genetic disease. Genes, **1**: 550–563. doi:10.3390/genes1030550. PMID:24710102.

Crocker, P.R. 2002. Siglecs: sialic-acid-binding immunoglobulin-like lectins in cell-cell interactions and signalling. Curr. Opin. Struct. Biol. **12**: 609–615. doi:10.1016/S0959-440X(02)00375-5. PMID:12464312.

Crocker, P.R., and Varki, A. 2001. Siglecs, sialic acids and innate immunity. Trends Immunol. **22**: 337–342. doi:10.1016/S1471-4906(01)01930-5. PMID:11377294.

Crocker, P.R., Paulson, J.C., and Varki, A. 2007. Siglecs and their roles in the immune system. Nat. Rev. Immunol. **7**: 255–266. doi:10.1038/nri2056. PMID:17380156.

Drouin, G. 2002. Characterization of the gene conversions between the multigene family members of the yeast genome. J. Mol. Evol. **55**: 14–23. doi:10.1007/s00239-001-0085-y. PMID:12165839.

Edgar, R. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics, **5**: 113. doi:10.1186/1471-2105-5-113. PMID:15318951.

Galtier, N., Piganeau, G., Mouchiroud, D., and Duret, L. 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. Genetics, **159**: 907–911. PMID:11693127.

Guindon, S., and Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. **52**: 696–704. doi:10.1080/10635150390235520. PMID:14530136.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acid. Symp. Ser. **41**: 95–98.

Hayakawa, T., Angata, T., Lewis, A.L., Mikkelsen, T.S., Varki, N.M., and Varki, A. 2005. A human-specific gene in microglia. Science, **309**: 1693. doi:10.1126/science.1114321. PMID:16151003.

Jandus, C., Simon, H.U., and von, and Gunten, S. 2011. Targeting Siglecs: a novel pharmacological strategy for immuno- and glycotherapy. Biochem. Pharmacol. **82**: 323–332. doi:10.1016/j.bcp.2011.05.018. PMID:21658374.

McGrath, C.L., Casola, C., and Hahn, M.W. 2009 Minimal effect of ectopic gene conversion among recent duplicates in four mammalian genomes. Genetics, **182**: 615–622. doi:10.1534/genetics.109.101428. PMID:19307604.

Meunier, J., and Duret, L. 2004. Recombination drives the evolution of GC-content in the human genome. Mol. Biol. Evol. **21**: 984–990. doi:10.1093/molbev/msh070. PMID:14963104.

Page, R.D.M. 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. **12**: 357–358. PMID:8902363.

Petronella, N., and Drouin, G. 2011. Gene conversions in the growth hormone gene family of primates: stronger homogenizing effects in the Hominidae lineage. Genomics, **98**: 173–181. doi:10.1016/j.ygeno.2011.06.001. PMID:21683133.

Posada, D., and Crandall, K.A. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci. U.S.A. **98**: 13757–13762. doi:10.1073/pnas.241370698. PMID:11717435.

R Development Core Team. 2006. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Sawyer, S. 1989. Statistical tests for detecting gene conversion. Mol. Biol. Evol. **6**: 526–538. PMID:2677599.

Sonnenburg, J.L., Altheide, T.K., and Varki, A. 2004. A uniquely human consequence of domain-specific functional adaptation in a sialic acid-binding receptor. Glycobiology, **14**: 339–346. doi:10.1093/glycob/cwh039. PMID:14693915.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28**: 2731–2739. doi:10.1093/molbev/msr121. PMID:21546353.

Varki, A. 2009. Natural ligands for CD33-related Siglecs? Glycobiology, **19**: 810–812. doi:10.1093/glycob/cwp063. PMID:19429924.

Varki, A. 2010. Uniquely human evolution of sialic acid genetics and biology. Proc. Natl. Acad. Sci. U.S.A. **107**: 8939–8946. doi:10.1073/pnas.0914634107. PMID:20445087.

Wang, X., Mitra, N., Cruz, P., Deng, L., NISC Comparative Sequencing Program, Varki, N., et al. 2012. Evolution of Siglec-11 and Siglec-16 genes in hominins. Mol. Biol. Evol. **29**: 2073–2086. doi:10.1093/molbev/mss077. PMID:22383531.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. **24**: 1586–1591. doi:10.1093/molbev/msm088. PMID:17483113.

Zid, M., and Drouin, G. 2013. Gene conversions are under purifying selection in the carcinoembryonic antigen immunoglobulin gene families of primates. Genomics, **102**: 301–309. doi:10.1016/j.ygeno.2013.07.003. PMID:23867109.