HAL
open science

Molecular Phylogeny and Functional Genomics of beta-Galactoside alpha2,6-Sialyltransferases That Explain Ubiquitous Expression of st6gal1 Gene in Amniotes.

Daniel Petit, Anne-Marie Mir, Jean-Michel Petit, Christine Thisse, Philippe Delannoy, Rafael Oriol, Bernard Thisse, Anne Harduin-Lepers

To cite this version:

Daniel Petit, Anne-Marie Mir, Jean-Michel Petit, Christine Thisse, Philippe Delannoy, et al.. Molecular Phylogeny and Functional Genomics of beta-Galactoside alpha2,6-Sialyltransferases That Explain Ubiquitous Expression of st6gal1 Gene in Amniotes.. Journal of Biological Chemistry, 2010, 285 (49), pp.38399-414. 10.1074/jbc.M110.163931 . hal-00521411

HAL Id: hal-00521411

https://hal.science/hal-00521411v1

Submitted on 30 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 285, NO. 49, pp. 38399–38414, December 3, 2010
© 2010 by The American Society for Biochemistry and Molecular Biology, Inc. Printed in the U.S.A.

Molecular Phylogeny and Functional Genomics of β-Galactoside α2,6-Sialyltransferases That Explain Ubiquitous Expression of *st6gal1* Gene in Amniotes*[S]

Received for publication, July 15, 2010, and in revised form, September 13, 2010 Published, JBC Papers in Press, September 20, 2010, DOI 10.1074/jbc.M110.163931

Daniel Petit[‡], Anne-Marie Mir[§], Jean-Michel Petit[‡], Christine Thisse[¶], Philippe Delannoy[§], Rafael Oriol[‖], Bernard Thisse[¶], and Anne Harduin-Lepers[§1]

From the [‡]Unité de Génétique Moléculaire Animale, Université de Limoges Faculté des Sciences et Techniques, INRA UMR 1061, 123 Avenue Albert Thomas, 87060 Limoges, France, the [§]Unité de Glycobiologie Structurale et Fonctionnelle, Université Lille Nord de France, Lille1, CNRS UMR 8576, IFR 147, 59655 Villeneuve d'Ascq, France, the [¶]Department of Cell Biology, School of Medicine, University of Virginia, Charlottesville, Virginia 22908, and the [‖]Unité de Microenvironnement et Physiologie de la Différenciation, Université de Paris Sud XI, INSERM U1004, 16 Avenue Paul Vaillant-Couturier, 94807 Villejuif, France

Downloaded from http://www.jbc.org/ at INRA Institut National de la Recherche Agronomique on June 14, 2018

Sialyltransferases are key enzymes in the biosynthesis of sialoglycoconjugates that catalyze the transfer of sialic residue from its activated form to an oligosaccharidic acceptor. β-Galactoside α2,6-sialyltransferases ST6Gal I and ST6Gal II are the two unique members of the ST6Gal family described in higher vertebrates. The availability of genome sequences enabled the identification of more distantly related invertebrates' *st6gal* gene sequences and allowed us to propose a scenario of their evolution. Using a phylogenomic approach, we present further evidence of an accelerated evolution of the *st6gal1* genes both in their genomic regulatory sequences and in their coding sequence in reptiles, birds, and mammals known as amniotes, whereas *st6gal2* genes conserve an ancestral profile of expression throughout vertebrate evolution.

Sialyltransferases described in higher vertebrates are glycosyltransferases that mediate the transfer of sialic acid residues from activated sugar donors (CMP-β-Neu5Ac,[2] CMP-β-Neu5Gc, and CMP-β-KDN) to terminal non-reducing positions of oligosaccharide chains of glycoproteins and glycolipids (reviewed in Refs. 1–3). Classically, the vertebrate sialyltransferase superfamily is divided into four families, namely the ST6Gal, ST3Gal, ST6GalNAc, and ST8Sia, depending on the glycosidic linkage formed and the monosaccharide acceptor used.[3] Members of the mammalian and avian ST6Gal family catalyze the transfer of sialic acid residues to the terminal galactose residues of the type 2 disaccharide (Gal(NAc)β1,4GlcNAc), resulting in the formation of an α2–6 glycosidic linkage (for reviews, see Refs. 3–10). Unlike the other sialyltransferase families, this family comprises only two paralogs in the human genome named *ST6GAL1* and *ST6GAL2*, respectively (1, 2). The human *ST6GAL1* gene is ubiquitously expressed in a broad variety of tissues, whereas the *ST6GAL2* gene is expressed in a tissue-specific (adult brain) and stage-specific (embryonic) manner. Mammalian *st6gal1* gene expression is regulated by multiple promoters governing the expression of several transcripts encoding identical polypeptide enzyme, and high levels of mRNA are detected in hematopoietic cells and in liver (11–13).

Sialylated α2,6-lactosaminyl structures (Neu5Acα2–6Galβ1–4GlcNAc; sia6LacNAc) found on *N*-glycosylproteins and also, to a lesser extent, on *O*-glycosylproteins, glycolipids, and free oligosaccharides (14) are involved in a highly specific recognition phenomenon (15). In the mammalian immune system, B cells highly express ST6Gal I (11, 16, 17), and sialylated α2,6-lactosaminyl structures generated on CD45 and immunoglobulin M (IgM) are the preferred ligands of CD22 (Siglec 2), a sialic acid-binding Ig-like lectin found exclusively on B-lymphocytes and involved in B cell immunologic activation and signaling as evidenced in KO mice (16, 18, 19). Overexpression of ST6Gal I has been reported in several human malignancies, and clinical and experimental studies suggest a positive correlation between high ST6Gal I levels and invasive behavior of cancer cells (14, 20). Integrin-mediated adhesion is based on protein interactions, and binding can be significantly modulated by sia6LacNAc structures on β1-integrin *in vivo* and *in vitro* in cancer cells, leading to enhanced cell motility and invasiveness (21–23). ST6Gal I plays a role in inflammation (24, 25), and in mammals, transient up-regulation occurs during acute phase reaction when the organism experiences trauma or infection (26, 27). Finally, in contrast to avian and other mammalian influenza viruses, human influenza virus A and B prefer the α2,6-linked sialic acid found in abundance in human upper airways over the α2–3-linked sialic acid (28–30). On the other hand, the ST6Gal II function remains unknown.

st6gal homologs have been cloned from several higher vertebrate species (1). Furthermore, a ST6Gal cDNA named DSiaT was cloned from *Drosophila melanogaster* (31), suggesting that the ST6Gal family was present in insects, although not much Neu5Ac could be detected (32–34). DSiaT is detected almost

* This work was supported in part by CNRS, Institut National de la Recherche Agronomique (INRA), INSERM, and the PPF-Bioinformatique de Lille.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. 1–5.

[1] To whom correspondence should be addressed. Tel.: 33-320-3362-46; Fax: 33-320-43-65-55; E-mail: anne.harduin@univ-lille1.fr.

[2] The abbreviations used are: Neu5Ac, *N*-acetylneuraminic acid; EST, expressed sequence tag; NJ, neighbor joining; ME, minimum evolution; ML, maximum likelihood; MYA, millions years ago; WGD, whole genome duplication; hpf, hours postfertilization; TSS, transcriptional start site; indel, insertion/deletion; PCA, principal component analysis; RACE, rapid amplification of cDNA ends; ISH, *in situ* hybridization; UT, untranslated.

[3] Sialyltransferase nomenclature is according to Tsuji *et al.* (123).

exclusively in central nervous system (CNS) neurons in the embryonic stage 17, in the optic lobe of third instar larva, and in adult head (35). Targeted disruption of the DSiaT gene results in a neurological phenotype, suggesting that DSiaT modulates the nervous system function of voltage-gated sodium channel (36). Because the mammalian *st6gal2* gene is detected mainly in CNS as well, it has been suggested that ST6Gal II might have conserved an ancestral function, whereas ST6Gal I would have developed new functions in vertebrates. Further understanding of the evolutionary history of *st6gal* genes through molecular phylogenetic analysis will shed light on the functions of these genes maintained during evolution.

In the era of genomics, we have developed the ability to investigate the genomic sequences of the sialyltransferase genes that modify glycans in different animal lineages, thus providing a powerful means of reconstructing the evolutionary history of sialylation, determining key genetic events in the establishment of glycan sialylation machinery (2, 37). In the present work, we address the fate of vertebrate *st6gal* genes. We take advantage of the wealth of data provided by complete genome projects to refine the molecular relationship of ST6Gal and to address *st6gal* gene evolutionary trends in terms of gene gain and loss and also translocation and mutation rate, those mechanisms that were instrumental in establishing modern functions of vertebrate ST6Gal I and ST6Gal II. We have traced the environment of these genes (*i.e.* the set of orthologous genes around *st6gal* gene loci). In parallel, we have compared the expression pattern of *st6gal* genes in the vertebrate lineage, through molecular cloning of bony fish (teleost) (*Danio rerio*) and amphibian (*Silurana tropicalis*) *st6gal*. Our phylogenetic and expression analysis provide valuable insights into *st6gal* gene evolution in vertebrates and a model of duplication events whereby the *st6gal1* genes have undergone neofunctionalization in higher vertebrates.

EXPERIMENTAL PROCEDURES

In Silico Sialyltransferase Retrieval—Only eukaryote sequences were considered for this study. Homologous *st6gal* sequences were searched through exploration of all genomic and expressed sequence tags (ESTs) available from general databases, such as NCBI (see the BLAST (Basic Local Alignment Search Tool) Web site) for the green lizard *Anolis carolinensis*, DDBJ, or Ensembl, or in specialized databases JGI for *Branchiostoma floridae*, the Genome Sequencing Center at the Washington University School of Medicine (St. Louis, MO) for the lamprey *Petromyzon marinus*, KEGG GENES (38–40), the Genome Sequencing Center at the Baylor College of Medicine for *Homo sapiens* and the sea urchin *Strongylocentrotus purpuratus*, and the Institute of Molecular and Cell Biology for the elephant shark *Callorhinchus milii* using BLASTN, TBLASTN, and PSI-BLAST (41) with default parameters (an *e*-value cut-off at 0.01 was used in all BLAST searches). Human and mouse sequences were used as first queries in the first round of search. The assignment of these sequences to ST6Gal was determined by the specific motifs that are hallmarks of this family (1, 42). All genomic sequences allowing generation of a complete catalytic domain were considered. Splice site prediction analysis was achieved at the Berkeley *Drosophila* Genome Project. The structure of the genes, in terms of exon/intron boundaries, was deduced from several non-exclusive strategies: (i) comparing the boundaries proposed by Genscan (MIT server), (ii) comparing EST from genomic assemblages (scaffolds or contigs), (iii) comparing the boundaries to those present in known genes.

Phylogenetic Analysis—The alignment of amino acid sequences was conducted using ClustalX software (43). The selection of informative sites was helped by G-BLOCKS (44) with the options of less stringent selection. Phylogeny trees were produced by maximum likelihood (ML) using PHYML, version 2.4.4 (45), with the Jones-Taylor-Thornton (JTT) model of amino acid substitution, neighbor joining (NJ), and minimum evolution (ME) using MEGA4.0 (46), and bootstrap percentages were calculated from 2000 replicates. The numbers of site changes in each branch were calculated with the Protpars program included in the PHYLIP Package (47), using 228 sites, under the constraint of the user tree produced by ME (see Refs. 48 and 49 for details).

The calibration used for dating the divergence between ST6Gal I and ST6Gal II in vertebrates was as follows: amphioxus/vertebrates, 650 MYA (50); lamprey/gnathostomata, 575 MYA; gnathostomata/osteichthyans, 460 MYA; osteichthyans/other vertebrates, 450 MYA; tetrapods/actinopterygians, 360 MYA; amniotes/other vertebrates, 310 MYA; genome duplication in teleosts (R3), 320 MYA (51). We calculated the regression equations between linearized branch Pearson's correlations, and associated probabilities were calculated with PAST version 2.01 (52).

Synteny Analysis and Paralogon Detection—Synteny between vertebrate *st6gal* and related genes in invertebrates was assessed by chromosomal walking and reciprocal BLAST searches of genes adjacent to *st6gal* loci in human (HSA), mouse (MMU), chicken (GGA), medaka (OLA), zebrafish (DRE), *Takifugu rubripes* (TRU), and amphioxus (BFL) genome databases (Ensembl). The detection of paralogous blocks (53) was done using the latest Ensembl data set (version 5.28). The Web site for these paralogons (see the Trinity College Dublin Web site) offers the possibility to carry out block detection in humans with self-defined parameters.

Expression Analysis—Unigene at the NCBI data base was used to quantify the number of ESTs identified for each tissue in the following species: *H. sapiens*, *Mus musculus*, *Gallus gallus*, *S. tropicalis*, and *D. rerio*. In order to homogenize the different overall values among organisms, we divided the number of ST6Gal ESTs by the total number of ESTs per tissue. Second, we removed the tissues for which only one organism was recorded. Third, the table containing 22 columns (tissues or developmental stages) and 5×2 (species $\times$ *st6gal1* and *st6gal2* genes) lines was submitted to a principal component analysis (PCA) using PAST 2.01 (52). According to the method described by Ermonval *et al.* (54), PCA allows projecting the data set onto a two-dimensional plan, each column factor represented by a vector according to pair-wise correlations; the higher the correlation between two factors, the more acute the angle between the vectors. In this plan, the *st6gal1* or *st6gal2* genes corresponding to a given species are projected in the direction of their greatest values. The EST ratios per tissue were multiplied by 10^6 and log-transformed to normalize the distri-

Downloaded from http://www.jbc.org/ at INRA Institut National de la Recherche Agronomique on June 14, 2018

bution and then submitted to a two-way clustering using Euclidean distance as measure of similarity, using PAST 2.01. The coloration intensity of each case in the table was in proportion to the values.

Animals and Maintenance—Zebrafish (*D. rerio*) and clawed frog (*S. tropicalis*) were maintained in our aquatic biology facility, as described previously (55, 56). All experimental procedures adhered to the CNRS guidelines for animals use.

Isolation of RNA, cDNA Synthesis, and PCR Analysis—Total RNA was extracted from various *S. tropicalis* and *D. rerio* tissues using the nucleospin RNA II kit (Macherey-Nagel, Hoerdt, France). A proteinase K digestion step (55 °C, 10 min) and phenol/chloroform extraction were inserted into the protocol after Dounce homogenization of the tissues and before column purification of total RNA. Cellular RNA was quantified using a NanoDrop® ND-1000 UV-visible spectrophotometer (NanoDrop Technologies, Wilmington, DE). RNA integrity was further assessed using the RNA 6000 Nano LabChip® kit on an Agilent Bioanalyzer (Agilent Technologies, Stratagene, La Jolla, CA). For subsequent PCR amplifications, first strand cDNA was synthesized from total RNA using an oligo(dT) primer and the AffinityScript Q-PCR cDNA synthesis kit according to the manufacturer's protocol (Agilent Technologies). Based on the nucleic acid sequences determined *in silico*, oligonucleotide primers were designed (Eurogentec, Herstal, Belgium) in the open reading frame (see supplemental Fig. 5). PCR amplifications were carried out with the Taq core kit DNA polymerase (Qiagen, Courtaboeuf, France) or Jena DNA polymerase (Jena Bioscience, Euromedex, Souffelweyersheim, France) using buffer solutions provided by the manufacturer. Annealing temperatures ranged from 48 to 55 °C, and amplified fragments were subjected to 2% agarose gel electrophoresis, visualized by ethidium bromide, gel-extracted, and subcloned in the pCR®2.1-TOPO vector (TOPO TA Cloning, Invitrogen, Cergy Pontoise, France). Nucleotide sequences were confirmed by sequencing (Genoscreen, Lille, France).

5′-Rapid Amplification of cDNA Ends (5′-RACE)—Amplification of the 5′-end of *S. tropicalis* and *D. rerio st6gal1* cDNA was achieved with the FirstChoice™ RLM-RACE kit (Ambion, Montrouge, France) according to the manufacturer's instructions. Total RNA (10 μg) from *S. tropicalis* liver and *D. rerio* eggs were treated with calf intestinal phosphatase and then with tobacco pyrophosphatase, leaving a 5′-monophosphate full-length mRNA. A 45-bp adaptor oligonucleotide was then ligated to the RNAs using T4 RNA ligase. A random-primed reverse transcription reaction was performed, followed by two consecutive PCRs with 200 μM dNTPs and 1 unit of AccuTaq DNA polymerase (Sigma) using two nested sets of primers (see supplemental Fig. 5). The 24-bp oligonucleotide sense-outer (5′-GCTGATGGCGATGAATGAACACTG-3′) and the gene-specific antisense oligonucleotide Reverse 2 or Reverse 3, for the amphibian and fish gene, respectively, were used in a first PCR at 96 °C for 2 min, followed by 38 cycles (96 °C for 45 s, 58 °C for 1 min and 68 °C for 1 min) and an extension step of 10 min at 68 °C. The 35-bp oligonucleotide sense-inner (5′-CGCGGATCCGAACACTGCGTTTGCTGGCTT-TGATG-3′) and the gene-specific antisense oligonucleotide Reverse 3 or Reverse 4 for the amphibian and fish gene, respectively, were used in a second PCR at 96 °C for 2 min, followed by 38 cycles (96 °C for 45 s, 58 °C for 90 s, and 68 °C for 1 min) and an extension step of 10 min at 68 °C. Amplification products were analyzed on a 1% (w/v) agarose gel with ethidium bromide staining, extracted from the gel, subcloned in TOPO^{R}II vector of the TOPO TA cloning kit (Invitrogen), and sequenced (Genoscreen).

Whole Mount mRNA in Situ Analysis—Both sense and antisense digoxigenin-labeled RNA probes were synthesized from PCR-amplified template using primers as described previously (55). Whole mount *in situ* hybridization was performed according to standard procedures (57–59).

RESULTS

Identification of ST6Gal-related Genes in Bilaterians

In order to identify putative genes encoding proteins with significant similarity to the known mammalian *st6gal* genes in animals with bilateral symmetry (bilaterians), we carried out a BLAST search in various invertebrate and vertebrate nucleotide databases using the known ST6Gal sequences. The search was based on the fact that the highly conserved sialylmotif peptide consensus sequences (L, S, III, and VS) are characteristic of all animal sialyltransferases and consequently serve as hallmarks for their identification.

A broad phylogenetic distribution of *st6gal* genes was observed in multicellular animals (metazoans). It should be noted that a short EST (NCBI, EST division: EC377350) from the sponge *Oscarella carmella* is attributable to ST6Gal. Despite an extensive examination of EST and whole genome shotgun sequences in data banks (JGI), no homologous *st6gal* gene was identified in the cnidaria *Nematostella vectensis*, the lophotrochozoa (polychete annelid *Capitella teleta* and mollusk *Lottia gigantia*), the hymenoptera insects *Apis mellifera* and *Nasonia vitripennis*, or in the nematoda *Caenorhabditis elegans* genome. It appears that among bilaterian animals developing first the mouth (protostomes), only one copy of *st6gal* gene sequence was retrieved from arthropods, like arachnida (*Ixodes scapularis* and *Varroa destructor*), crustacea (*Daphnia pulex* and *Calligus rogercresseyi*), and insects diptera (*D. melanogaster, Anopheles gambiae*, *Aedes aegypti*, and *Culex quinquefasciatus*), homoptera (*Acyrthosiphon pisum*), lepidoptera (*Bombyx mori*), phthiraptera (*Pediculus humanus corporis*), and coleoptera (*Tribolium castaneum*). Among bilaterian animals developing first the anus (deuterostomes), we found one copy of the *st6gal* gene in the hemichordata *Saccoglossus kowalevskii* and two copies in the amphioxus (*B. floridae*), but none was found in the sea urchin (*S. purpuratus*) or in the tunicates *Ciona intestinalis* and *Ciona savignyi*. In vertebrates, most examined genomes contain two members the *st6gal1* and *st6gal2* paralogous genes, except in the lamprey *Petromyzon marinus*, where three *st6gal* copies were found. In teleosts, we also describe three members named *st6gal1*, *st6gal2*, and *st6gal2-r* in the zebrafish (*D. rerio*) genome. In order to gain further insights into lower vertebrate *st6gal* genes, we carried out by RT-PCR molecular cloning of DNA clones encoding β-galactoside α2,6-sialyltransferases that were identified in the *D. rerio* and *S. tropicalis* genome. *D. rerio* and *S. tropicalis*

Downloaded from http://www.jbc.org/ at INRA Institut National de la Recherche Agronomique on June 14, 2018

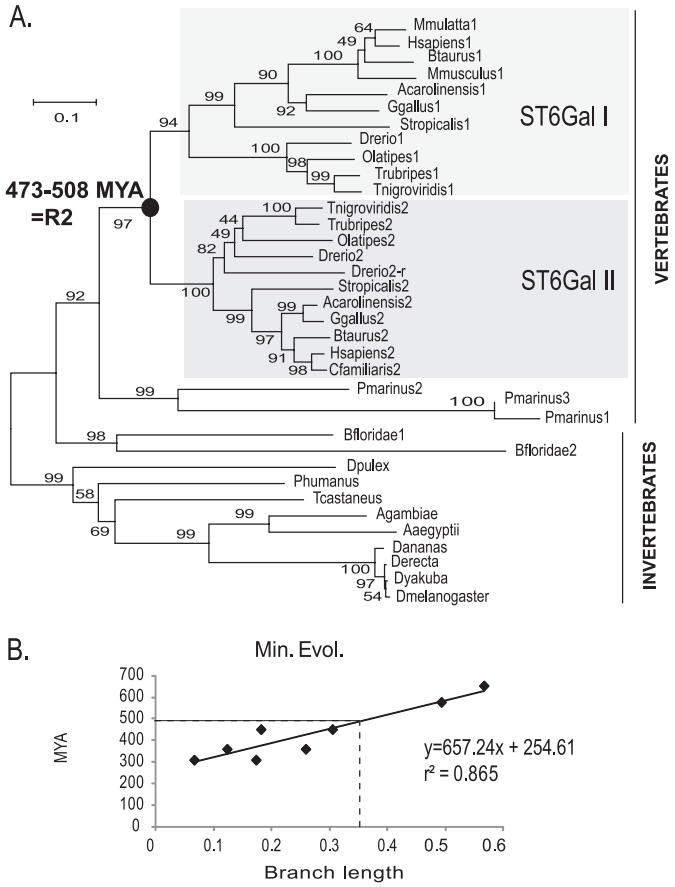


FIGURE 1. *A*, ML phylogenetic tree of 36 sialyltransferases of the ST6Gal family. A maximum likelihood phylogenetic tree was inferred by the Phyml, JTT model of amino acid substitution: 36 ST6Gal sequences, 11 vertebrate ST6Gal I, 11 vertebrate ST6Gal II. G-BLOCKS selected 200 positions. Bootstrap values were calculated from 500 replicates, and values of >50% are reported at the *left* of each divergence point. The tree is rooted with the invertebrate arthropod sequences as the outgroup. *B*, calculation of divergence time between ST6Gal I and ST6Gal II. The lengths of each horizontal branch were calculated from the linearized tree obtained with ME (*Min. Evol.*) using MEGA4.0.

ST6Gal I deduced protein sequences are 484 and 474 amino acids long, respectively, and show little overall sequence identity (40 and 36%) with their human counterpart (406 amino acids). On the other hand, ST6Gal II and ST6Gal II-r deduced protein sequences of *D. rerio* (514 and 453 amino acids, respectively) and *S. tropicalis* ST6Gal II have a higher level of sequence identity, 53, 44, and 60% compared with human ST6Gal II. The accession numbers of all *st6gal* sequences identified and analyzed are gathered in supplemental Fig. 1.

Molecular Phylogeny Analysis

Catalytic Domain—As a first step in the analysis, we assessed the orthology of the catalytic domain of vertebrate and invertebrate ST6Gal-related protein sequences by multiple sequence alignments with ClustalW (supplemental Fig. 2). The G-BLOCKS server evidenced 200 informative sites to construct the phylogenetic trees. The three tested methods to infer ST6Gal phylogeny (NJ, ME, and ML, using JTT as transition matrix) gave the same topology (Fig. 1). We found that bony fishes, such as the zebrafish *D. rerio*, the medaka *Oryzias latipes*, the three-spined stickleback *Gasterosteus aculeatus*, the tetraodonte *Tetraodon nigroviridis*, and the fugu *T. rubripes* have orthologs of the two mammalian ST6Gal subfamilies. Moreover, a new subfamily is present in *D. rerio* and is named ST6Gal II-related (ST6Gal II-r) because it has a clear sequence relationship to the ST6Gal II subfamily. This new subfamily has disappeared from the other fish genomes. The three copies in the lamprey *P. marinus* and two copies in the amphioxus *B. floridae* are sister sequences to both ST6Gal I and ST6Gal II vertebrate subfamilies because they branch out from the phylogenetic tree before the split into two subfamilies. These ST6Gal sequences result from one and two duplication events, respectively, limited to these organisms that occurred after divergence of the amphioxus and lamprey lineages, respectively.

TABLE 1

Number of site changes in the catalytic part (228 sites) during transitions of vertebrate evolution

	ST6Gal I	ST6Gal II	χ^2	Significance[a]
Amniotes[b]/mammals	34	8	16.10	***
Tetrapods[c]/amniotes	28	13	5.49	*
Tetrapods/*Xenopus*	58	30	8.91	**
Osteichthyans[d]/tetrapods	25	23	0.08	NS
Osteichthyans/teleosts[e]	48	17	14.78	***
R2/vertebrates	18	32	3.92	*

[a] *, $p < 5\%$; **, $p < 1\%$; ***, $p < 0.1\%$; NS, not significant.
[b] Tetrapod vertebrates with amnios.
[c] Four-legged vertebrates.
[d] Bony fishes plus tetrapods.
[e] Bony fishes with mobile maxilla and premaxilla.

In order to estimate the time of divergence of the vertebrate *st6gal* gene subfamilies, we reconstructed linearized trees for duplicate genes under the assumption of a molecular clock using MEGA4.0 (60). The results obtained with NJ, ME, and ML are given in Fig. 1 and give an estimated divergence time in the range of 473 MYA by ME (Fig. 1*B*), 499 MYA by NJ, and 508 MYA according to ML.

We also observed on the phylogenetic tree that the branch lengths in the vertebrate ST6Gal I clade were longer than in the vertebrate ST6Gal II clade (Fig. 1*A*). We thus tested the significance of these differences for each internal branch (*e.g.* from the ancestor of osteichthyans to the ancestor of teleosts). For each branch, we counted the numbers of site changes using the parsimony program Protpars in PHYLIP in the ST6Gal I and ST6Gal II sequences of the catalytic domain (Table 1). The χ^2 tests show that there is a highly significant accumulation of mutations in the ST6Gal I branches leading to mammals and to teleosts and to a lesser extent to amphibians, relative to ST6Gal II branches. In contrast, we observe an accumulation of substitutions in the branch leading to osteichthyans ST6Gal II compared with the ST6Gal I counterpart.

In order to better understand the significance of these changes, we also compared the substitution numbers in the conserved motifs between ST6Gal sequences for each branch (Table 2). The greatest amounts were observed in the sialylmotifs L and S and in the family motif b (1), with a regular excess of changes found in ST6Gal I sequences; it concerns the sialylmotifs L and family motif b in the transition amniotes-mammals, the sialylmotif L in the transition tetrapods-amniotes, and the family motif b and sialylmotif S in the transition osteichthyans-teleosts.

N Terminus of ST6Gal—The length of the protein sequences encoded by the first exon of vertebrate *st6gal* genes, encom-

Downloaded from http://www.jbc.org/ at INRA Institut National de la Recherche Agronomique on June 14, 2018

passing the cytoplasmic and transmembrane domains and the stem region, varies from 201 amino acid residues in the human ST6Gal I sequence to 329 amino acids in the fugu ST6Gal II sequence. Multiple sequence alignments of this region using ClustalW revealed weak sequence conservation upstream of the tryptophan residue Trp^{96} and Trp^{208} in human ST6Gal I and ST6Gal II protein sequences, respectively, and among tetrapod ST6Gal I protein sequences (QVW-KDP) (61). Local alignments performed by ClustalX allowed refinement of the correspondences between the amino acid sequences (supplemental Fig. 3). These alignments revealed several insertion events, such as a poly(E) in *T. nigroviridis*, *D. rerio*, *O. latipes*, and *T. rubripes* ST6Gal II sequences and a poly(QLEREK) in the amphibian *S. tropicalis* ST6Gal I sequence of unknown biological relevance. Altogether, these observations suggest that the ancestral *st6gal1* and *st6gal2* genes have undergone small insertion/deletion (indel) events during vertebrate evolution that led to changes in the reading frame.

At the gene level, we pointed out previously overall gene organization conservation in five coding exons of the *st6gal1* vertebrate genes with the notable exception of fish *st6gal1* genes, which exhibit additional coding exons in their 5′ region (1, 2) or, alternatively, two additional intron sequences. The position of the teleost second intron is inside a relatively well conserved protein sequence, downstream to the amino acid corresponding to the human Trp^{96}. These results do not support the exon shuffling hypothesis.

We then tested if the indel events in *st6gal* genes could be linked to evolutionary change amounts in the catalytic regions. We took into account the events encompassing at least three codons retrieved in the sequences coding the stem region but absent in the sequences coding the catalytic domain. We also considered the two introns located in the region encoding the stem region of the teleost *st6gal1* gene, which could be interpreted as insertions. Except in this last case, most indels could be considered as deletions compared with arthropod sequences. The largest deletion, denoted *ID6* in Fig. 2*A*, concerns tetrapod ST6Gal I and comprises around 70 codons. A 17-codon-long deletion (*ID5*) characterizes vertebrate ST6Gal I sequences. A 15-codon-long insertion is only shared by *T. rubripes* and *T. nigroviridis* ST6Gal II sequences (supplemental Fig. 3). Three indels remain ambiguous and may correspond to insertions in the ancestor to vertebrate ST6Gal I and ST6Gal II or to deletions in arthropod ST6Gal I/II (Fig. 2*A*). Interestingly, most indel events are clearly hallmarks of ST6Gal subfamilies in different subsets of vertebrates (Fig. 2*B*). We tested if these indel events were linked to the length of corresponding branches in the phylogeny tree constructed from the comparison of the catalytic part of the protein. Because the branch

TABLE 2

Number of site changes in the conserved motifs during transitions of vertebrate evolution

	Sialylmotif L		Family motif b		Sialylmotif S	
	ST6Gal I	ST6Gal II	ST6Gal I	ST6Gal II	ST6Gal I	ST6Gal II
Amniotes/mammals	**5**	1	**4**	1	2	1
Tetrapodes/amniotes	4	0	2	2	**3**	**1**
Osteichthyans/tetrapodes	7	6	2	2	1	2
Osteichthyans/teleosts	4	5	**9**	**1**	**4**	**0**
R2/osteichthyans	4	4	1	1	4	3

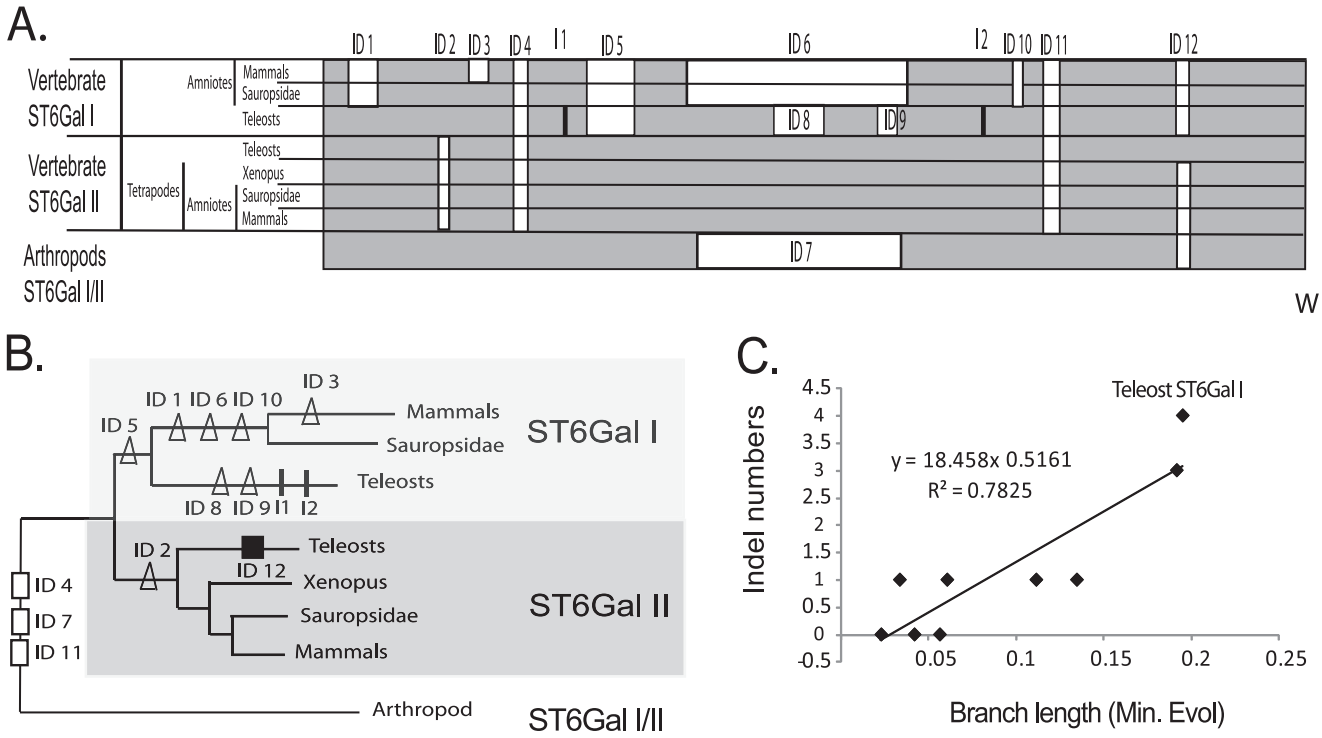


FIGURE 2. **Indel insertion/deletion events in the ST6Gal I and ST6Gal II protein sequences.** *A*, alignments of protein sequences without gap are schematized in *gray*; indels are indicated by *white boxes*. Intron insertions are indicated by *black bars* in the teleost *st6gal1* sequence. *B*, simplified phylogenetic tree and insertion/deletion distribution. The tree was constructed with 200 positions within the catalytic domain, using ME. *Filled square*, insertion; *open squares*, indel; *filled rectangles*, intron insertion; *open triangles*, deletion. *C*, correlation between Indel numbers and branch lengths. The tree was constructed as in *B*. The regression equation was given by PAST 2.01.

Downloaded from http://www.jbc.org/ at INRA Institut National de la Recherche Agronomique on June 14, 2018

lengths vary upon the algorithms, we considered the values given by NJ, ME, and parsimony (*i.e.* the number of site changes using the topology obtained with ME) (Fig. 2*C*). The Pearson's *r* values between branch lengths and indel events are summarized in Table 3. Whatever the reconstruction algorithm, it appears that there is a significant and positive correlation between the branch length and the number of indel events.

Chromosomal Location of st6gal Genes; Synteny and Paralogy around st6gal Genes

In order to investigate the dynamic of *st6gal* gene evolution across vertebrate genomes and to explain the appearance of the two vertebrate *st6gal* gene subfamilies, we first analyzed the evolutionary history of *st6gal* in the context of the two rounds of whole genome duplications (WGD), also known as the 2R hypothesis (62). We assessed the paralogy and synteny relationships of the identified *st6gal* genomic loci in various vertebrate genomes. The presence of two or more orthologous gene pairs on two distinct chromosomes in a single species can define paralogons issued from WGD events R1 and R2. In the human genome, using the Paralogon program (53), we found a statistically significant ($sm > 3$) block limited to three genes (data not shown). We then studied a larger segment around both *st6gal* genes, using Ensembl and found a set of 11 putative paralogous genes on HSA 3q27 and HSA 2q11.3 (Fig. 3) emphasizing the involvement of a genome doubling event. Taken together, these approaches support the hypothesis of WGD as a cause of *st6gal* gene duplication in vertebrates.

Next, we examined the two *st6gal* loci and their neighbors in the genome of various vertebrate species using the Synteny Database (63) (available on the World Wide Web). A conserved synteny refers to the existence of two or more orthologous genes that are co-localized on the same chromosome in two or more animal species, although their gene order on each chromosome can be different (64). The synteny including the *st6gal2* gene is simple because the synteny data base site gave a set of 10 genes common to human HSA2q12 and zebrafish DRE9 (Fig. 4*A*). In the other examined teleost genomes (medaka and fugu), only one chromosome bears the *st6ga12* synteny. In the *S. tropicalis* genome, a series of four scaffolds corresponds to this synteny, suggesting their colinearity (supplemental Fig. 4). In addition, a paralogon of four genes, including the *st6gal2* gene, was found in the zebrafish genome on DRE9 and DRE6 (Fig. 4*A*), suggesting a genome doubling event in teleosts (WGD R3).

For the synteny around the *st6gal1* gene, the situation appears to be more complex because two different sets of genes can be defined in teleosts and in amniotes (Fig. 4*B*), both well conserved within these two vertebrate groups. On one hand, HSA3 and GGA9 share 261 orthologous genes, among which 21 are present on *S. tropicalis* scaffold 55 (supplemental Fig. 4). On the other hand, the fish chromosomes DRE21, GAC7, and TNI7 share six genes (sclc6a7, trpc2, ca4, pura, *st6gal1*, rhogb), but only one gene, *st6gal1*, is common to both groups of vertebrate genes (Fig. 4*B*). Further analysis performed in the synteny data base revealed seven genes shared by GGA9 and DRE21, including *st6gal1*, 15 shared by GGA9 and DRE15, and 18 shared by GGA9 and DRE2 (supplemental Fig. 4). Five genes (rbp2, itm2c, clsn2, crbp2, and atp1b) have paralogs on DRE2 and DRE15, indicating that these segments result from a WGD R3 event that occurred at the base of teleost radiation, ~350 MYA (65–67). In summary, we can infer that in teleosts, a block of at least seven genes has been translocated to the equivalent chromosome of DRE21, from the protochromosome DRE15–2 of their common ancestor. Interestingly, there are two paralogous genes on DRE21 and DRE15 (neu2 and gpcr-rhod) that are absent from DRE2, suggesting that the seven-gene block has been translocated from the DRE2 ancestral chromosome, after the WGD R3 event (Fig. 4*C*). In addition, several genes around *st6gal1/2* in the *B. floridae* genome (scaf V2 104q) are retrieved around both the *st6gal1* and *st6gal2* genes in the human and chicken genomes (Fig. 4*D*), further suggesting conservation of synteny for *st6gal* genes from cephalochordates to mammals and a disruption of *st6gal1* synteny in teleosts.

TABLE 3

Correlations between indel event number and branch lengths obtained with different methods ($n = 8$)

	ME	NJ	Site change numbers
Pearson's *r*	0.885	0.833	0.785
p	0.15%	0.53%	1.21%

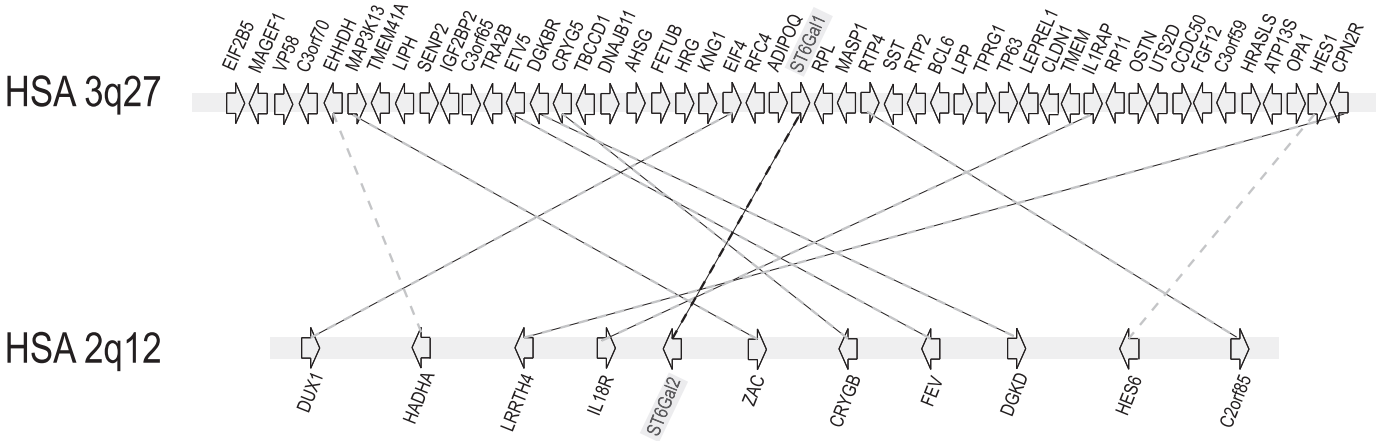


FIGURE 3. **Genomic organization of human *st6gal1*/*st6gal2* cluster paralogon and putative orthologous counterpart is indicative of a WGD event.** Eleven putative paralogous genes spanning regions on human chromosomes HSA2q11.3 and HSA3q27 were found. Genes represented were chosen from analysis performed at Ensembl by chromosomal walking and reciprocal TBLASTN searches of genes adjacent to *st6gal* loci.

Downloaded from http://www.jbc.org/ at INRA Institut National de la Recherche Agronomique on June 14, 2018

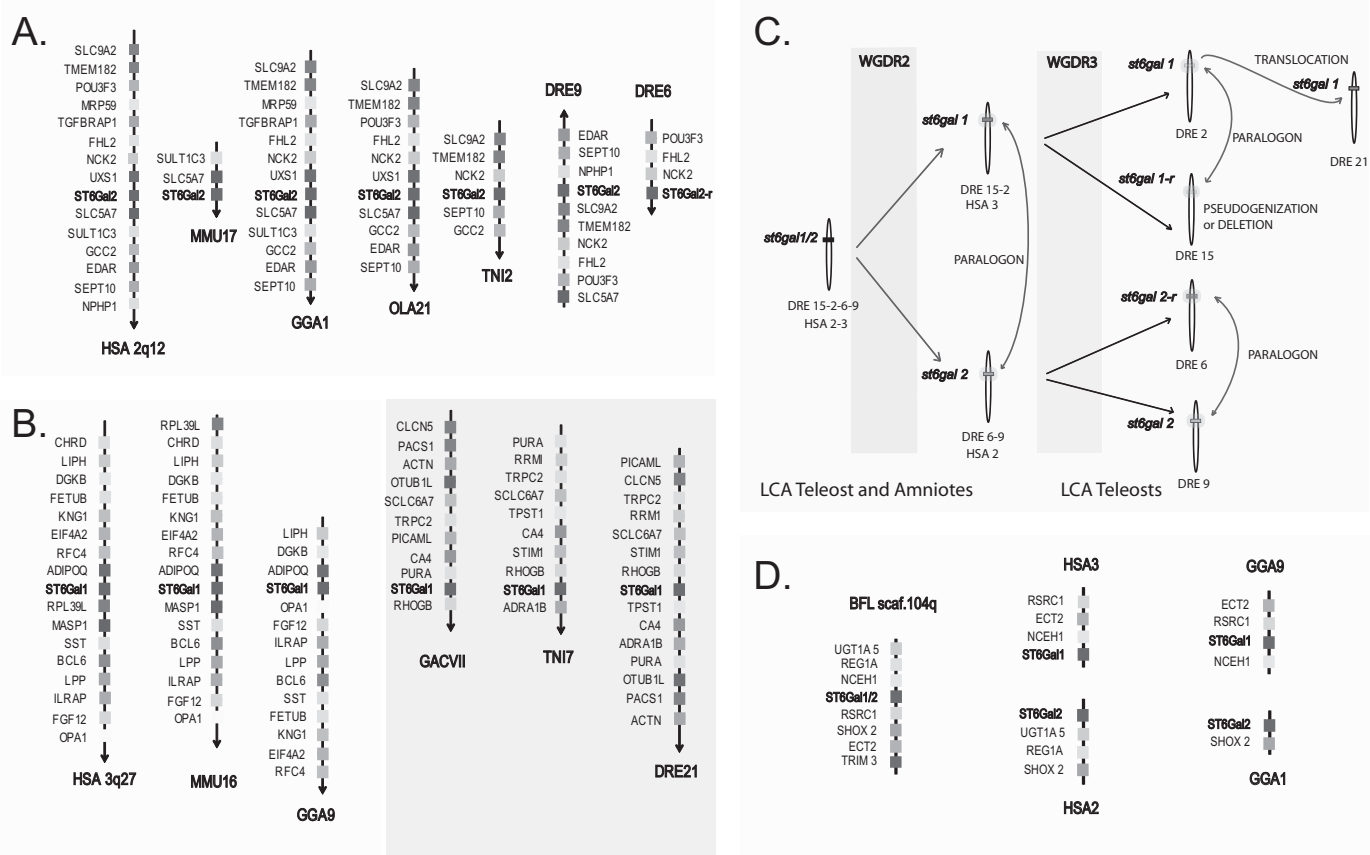


FIGURE 4. **Syntenic relationships of the *st6gal* gene loci in vertebrate and amphioxus genomes.** Physically mapped genomes of human (*HSA*), mouse (*MMU*), chicken (*GGA*), *O. latipes* (*OLA*), *T. nigroviridis* (*TNI*), *G. aculeatus* (*GAC*), and *D. rerio* (*DRE*) in Ensembl were used to identify conserved gene neighbors of the *st6gal* genes. *A*, conserved syntenic blocks found around the *st6gal2* gene; *B*, disrupted synteny around the *st6gal1* gene found in amniotes and in teleost genomes. *C*, schematic diagram depicting the genetic events that occurred early in the jawed vertebrate lineage and led to modern *st6gal* genes in fish. The last common ancestor (*LCA*) of teleost fishes and amniotes around 500 MYA bore an ancestral *st6gal* gene designated the *st6gal1/2* gene, which was located on the protochromosome DRE 15-2-6-9 HSA 2-3; WGD R2 (~475 MYA) gave rise to *st6gal1* and *st6gal2* genes found on HSA3 and HSA2, respectively, and on the protochromosome DRE 15-2 and DRE 6-9 of the teleost last common ancestor. The DRE 2/DRE 15 and DRE 6/DRE 9 pairs of chromosomes derived from these single common protochromosomes DRE 15-2 and DRE 6-9 after WGD R3. Rearrangement of a block of genes including *st6gal1* occurred in teleosts, resulting in the translocation of *st6gal1* on DRE21, whereas the other duplicate, named *st6gal1-r*, disappeared from DRE 15 through pseudogenization or deletion. *st6gal2* and *st6gal2-r* were maintained on DRE 6 and DRE 9. *D*, conserved synteny observed between the amphioxus genomic region hosting the *st6gal*-like gene (scaf V2 104q) and the two sets of human paralogons described in the legend to Fig. 3.

ST6Gal Gene Expression in Vertebrates and EST Analysis

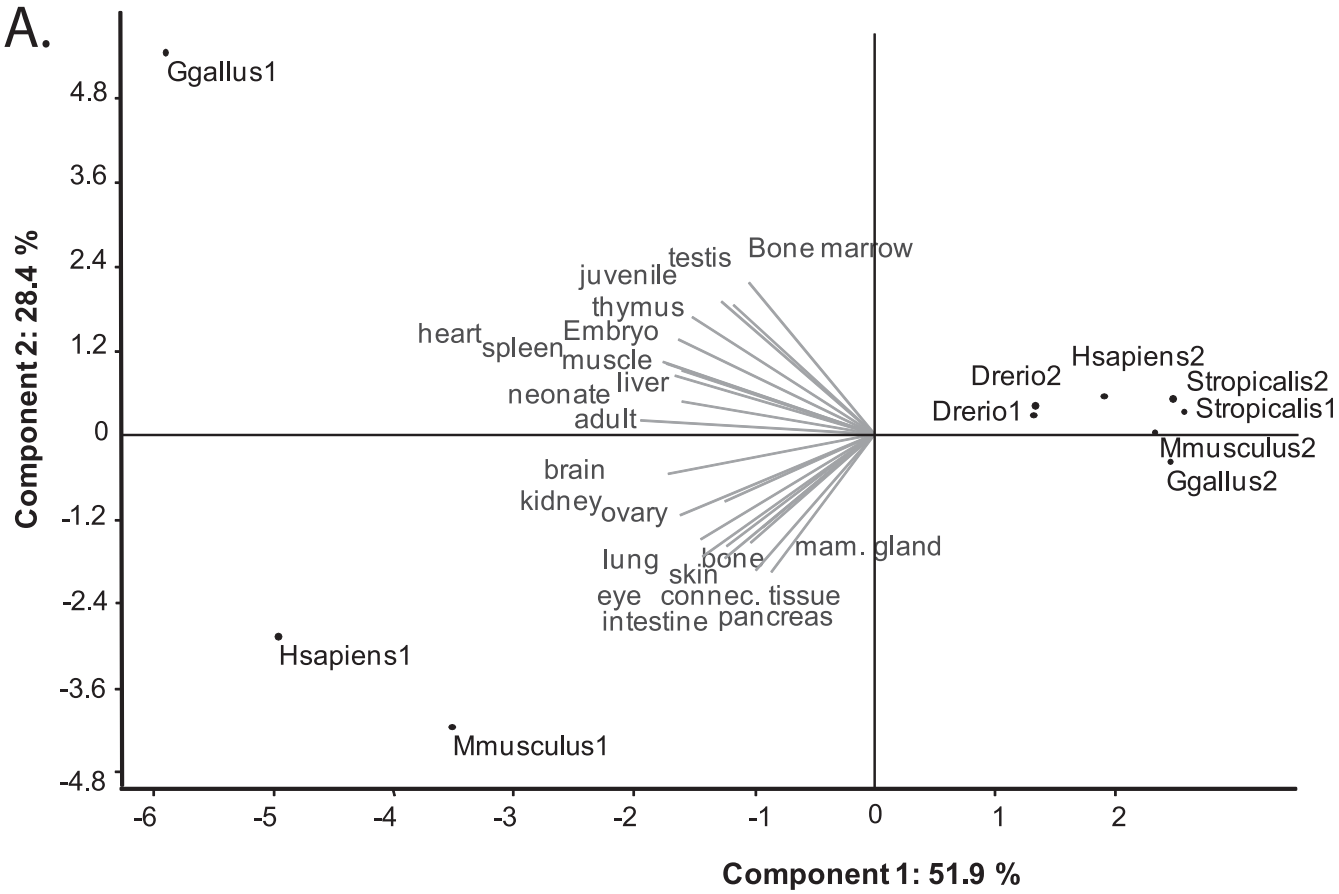
Several studies have noted the differential expression pattern of α2,6-sialylation and *st6gal* genes in various mammal species (37, 68–71). To estimate the breadth of *st6gal* gene expression, we looked at various tissue EST libraries from several representative animal species. We statistically analyzed the expression profiles of ESTs from the information retrieved on the Unigene site of NCBI. Tissue-dependent expression patterns were inferred from the EST profile accessible from the Unigene data base. The multivariate approach of PCA gave a quite satisfactory result. The plan defined by the two first axes takes into account about 80% of information of the data set (Fig. 5*A*). The first axis of PCA expresses nearly 52% of variance, whereas the second axis represents more than 28% of variance. It appears that the projections of most gene expression profiles are gathered on the *right side* of the plane, whereas mammal and bird *st6gal1* appears apart on the *left side*. This observation suggests that the expression profile of amniote *st6gal1* genes is almost ubiquitous, whereas teleost and amphibian *st6gal1* genes have a more similar profile of expression compared with vertebrate *st6gal2* genes. Furthermore, direction of the vectors corresponding to each tissue indicates preponderant expression of the pointed gene. As an example, the avian *st6gal1* gene is more expressed in thymus, testis, or muscle compared with its mammalian counterpart, which is predominantly expressed in lung, kidney, or brain. The heat map (Fig. 5*B*) constructed using PAST 2.01 with log-transformed values illustrates the sububiquitous expression of the *st6gal1* gene in mammals and bird and indicates that testis, brain, kidney, and embryo tissues frequently express the *st6gal2* gene.

RT-PCR in Adult Fish and Amphibian Tissues

To substantiate these observations and gain further insights into the expression of lower vertebrate *st6gal* genes, we designed oligonucleotides primers in the amphibian *S. tropicalis* and fish *D. rerio st6gal* genes (supplemental Fig. 5). We analyzed their expression patterns in various adult tissues by means of RT-PCR (Fig. 6). The three zebrafish and the two amphibian *st6gal* genes were differentially transcribed in various *D. rerio* and *S. tropicalis* adult tissues. Interestingly, the *st6gal1* gene is not ubiquitously expressed in fish or in amphibian adult tissues like in mouse, human, or bovine tissues, but its expression is

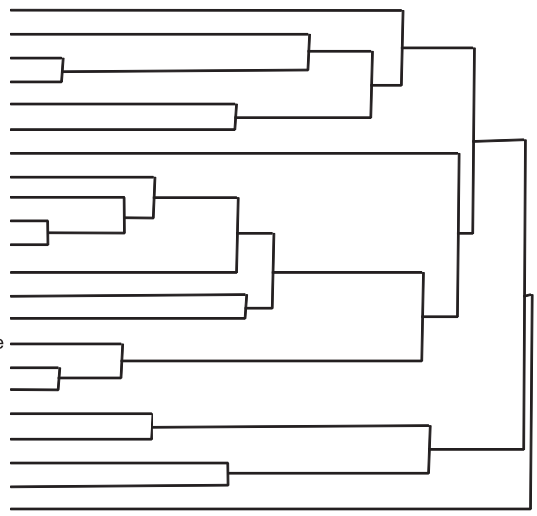
Downloaded from http://www.jbc.org/ at INRA Institut National de la Recherche Agronomique on June 14, 2018

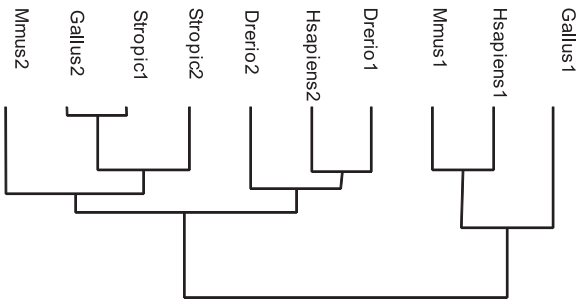
A.



B.

	Mmus2	Gallus2	Stropic1	Stropic2	Drerio2	Hsapiens2	Drerio1	Mmus1	Hsapiens1	Gallus1
testis	0	0	0	3.201	?	3.695	?	2.886	2.95	4.475
ovary	0	0	0	0	0	3.409	3.671	4.303	4.77	3.806
thymus	0	0	0	0	?	2.589	?	2.859	4.601	5.658
kidney	0	?	0	0	3.681	2.343	3.012	3.226	4.815	?
brain	3.902	0	0	0	3.53	3.43	3.139	4.541	4.356	4.338
adult	2.279	0	0	1.664	2.728	2.373	3.44	4	4.198	4.423
heart	0	0	0	0	0	0	0	0	3.533	4.149
skin	0	?	0	0	0	0	0	5.082	4.78	?
spleen	0	0	0	0	?	0	?	3.763	3.638	4.616
mammary gland	0	?	?	?	?	0	?	4.658	3.685	?
lung	0	?	0	0	?	0	?	4.799	3.937	?
liver	0	0	0	0	0	0	0	2.95	5.185	5.51
muscle	0	0	0	0	2.856	0	0	3.621	2.327	4.117
bone marrow	0	0	?	?	?	0	?	2.729	3.062	6.024
connective tissue	0	0	?	?	?	2.039	?	3.934	4.311	0
intestine	0	0	0	0	0	2.25	0	3.19	3.865	0
bone	0	?	0	0	0	2.703	0	3.425	3.756	?
juvenile	3.213	0	?	?	0	0	0	3.65	4.004	5.991
neonate	3.864	0	?	?	0	0	0	3.864	4.577	4.709
embryo	2.427	2.816	2.207	0	?	3.51	?	?	3.857	4.156
eye	3.318	?	?	?	0	2.347	0	4.771	3.97	?
pancreas	3.372	0	0	0	?	0	?	3.869	3.512	0





Downloaded from http://www.jbc.org/ at INRA Institut National de la Recherche Agronomique on June 14, 2018

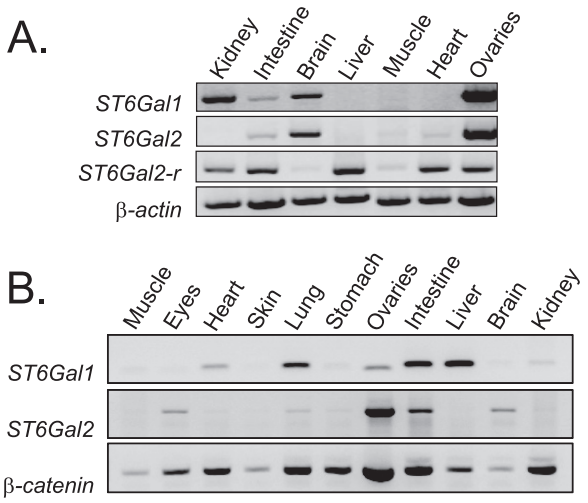


FIGURE 6. **Expression pattern of the zebrafish and amphibian *st6gal* genes in various adult tissues using RT-PCR.** *A*, fish *st6gal* genes. Relative expression levels of zebrafish *st6gal* and β-actin mRNA were evaluated by RT-PCR as described under "Experimental Procedures," among various adult tissues. Oligonucleotide primer sequences specific to fish *st6gal1*, *st6gal2*, and *st6gal2-r* are given in supplemental Fig. 5 and yield 896-, 628-, and 299-bp PCR fragments, respectively. *Lane 1*, kidney; *lane 2*, intestine; *lane 3*, brain; *lane 4*, liver; *lane 5*, muscle; *lane 6*, heart; *lane 7*, eggs; *lane 8*, oocytes. The zebrafish β-actin (378 bp) was amplified as a control of cDNA synthesis and purity. *B*, amphibian *st6gal* genes. Oligonucleotides specific to *S. tropicalis st6gal1* and *st6gal2* are indicated in supplemental Fig. 5 and yield 236- and 465-bp PCR fragments, respectively. *Lane 1*, muscle; *lane 2*, eyes; *lane 3*, heart; *lane 4*, skin; *lane 5*, lung; *lane 6*, stomach; *lane 7*, ovaries; *lane 8*, intestine; *lane 9*, liver; *lane 10*, brain; *lane 11*, kidney. The amphibian β-catenin (327 bp) was amplified for 35 cycles as a control of cDNA synthesis and purity.

restricted to intestine, kidney, and ovaries. It is also expressed in liver at a low level in fish and to a larger extent in frogs. Conversely, the *st6gal1* gene is largely expressed in adult fish brain, whereas it is almost not detected in frog brain. Altogether, both *st6gal* genes have a similar expression profile, and they are notably not detected in muscle and heart. The amphibian and zebrafish *st6gal2* gene expression is maintained in adult brain, ovaries, and intestine with overlapping territories of expression for *st6gal1*.

In Situ Hybridization (ISH) during Zebrafish Development

To examine the expression pattern among *st6gal* paralogs during zebrafish embryonic development, we performed whole-mount RNA ISH with zebrafish embryos (Fig. 7). *st6gal2* and *st6gal2-r* gene expression was detected from gastrulation until larva stage (5 days postfecundation), whereas *st6gal1* gene expression was not detected before 24 h postfecundation (hpf) or after hatching (48 hpf). Our ISH analysis indicated that at embryonic developmental stage 48 hpf, *st6gal1* and *st6gal2* genes are expressed in overlapping brain territories of zebrafish. We found a continuous expression of the two *st6gal2*-related genes during development, from egg to larva stages. Both genes are detected in hatching gland cells. As for the *st6gal2-r* gene, the overall level of expression is rather low, and we noticed an increased expression during late stages of development. The highest level of expression was found in the brain and in non-neuronal territories, such as the proctodeum, gall bladder, and intestinal bulb. *st6gal2* is expressed in the marginal zone of the CNS, stronger in anterior diencephalon and in lateral anterior hindbrain, and in the ganglion cell layer of retina, except in the proliferative zone.

Regulatory Evolution of st6gal1 Genes; 5′-RLM-RACE

The transcriptional start site(s) (TSS) and complete 5′-untranslated region (5′-UTR) were determined by 5′-RLM-RACE in lower vertebrate *st6gal1* genes of the zebrafish *D. rerio* and the frog *S. tropicalis* using total RNA extracted from zebrafish eggs and intestine tissues or frog liver and intestine tissues, respectively (data not shown). Unique 5′-RACE amplification products of about 160 bp in zebrafish tissues and of about 1060 bp in frog tissues were obtained and subcloned in TOPO TA pCRII vector, and several clones were fully sequenced. The results demonstrated the existence of a unique TSS for zebrafish and frog *st6gal1* genes in these tissues. Comparison of these cDNA sequences with genomic databases indicated that these unique zebrafish and frog transcripts show either one or two additional 5′-UT exons, respectively (Fig. 8), located far upstream the first coding exon. In contrast to the higher vertebrates, where a complex 5′-UTR with multiple upstream non-coding exons and multiple start sites has been described for the *st6gal1* genes, there is a unique *st6gal1* transcript in lower vertebrates showing a simple 5′-UTR with one or two non-coding exons.

DISCUSSION

Because many biological processes are governed by carbohydrate-protein interactions involving sialic acids, the evolutionary approach to gain further insights into the biological relevance of sialyltransferases is of particular interest (72, 73). The β-galactoside α2,6-sialyltransferases ST6Gal I and ST6Gal II mainly described in mammals mediate the addition of α2,6-linked sialic acid to Galβ1–4GlcNAc and GalNAcβ1–4GlcNAc disaccharides, respectively (8). Our phylogenetic and gene expression studies provide insights into the regulation and function of these conserved genes as well as important clues to the evolutionary events and functional changes that have occurred in different animal species. To date, such results on phylogenetic relationships and expression patterns of a glycosyltransferase family are quite unique (1, 37, 55).

st6gal Gene Sequences Appeared Early in Metazoans—The mRNA fragment identified from *O. carmella*, a sponge with chemical conduction, epithelial-like cells, and sensory-like cells from the porifera phylum (74), suggests that an ancestral *st6gal1/2* gene was already present in the earliest metazoans. This gene could be orthologous to the one present in the siliceous sponge *Geodia cydonium*, in which Muller *et al.* (75)

FIGURE 5. ***st6gal* EST expression profile analysis in vertebrates using PCA and heat map.** *A*, the PCA plot obtained with PAST 2.01 illustrates the analysis of 6420950 ESTs from *H. sapiens*, 4432921 from *M. musculus*, 1033498 from *Xenopus laevis*, and 1294007 from *D. rerio*. The first two principal components are used as the *horizontal* and *vertical axis*, respectively (the cumulative proportion is 80.3%). PC1 represents 51.9%, and PC2 is 28.4%. *B*, heat map diagram of differential *st6gal* gene expression in vertebrate tissues. Each *column* represents a single vertebrate *st6gal* gene, and each *row* represents a single tissue. The two gene clusters were as follows: Cluster 1, Mmus2 Ggallus2 Stropic1, Stropic2, Drerio2, Hsapiens2, and Drerio1; Cluster 2, Mmus1, Hsapiens1, and Ggallus1. Expression levels are directly indicated by *numbers*. 0 indicates no EST found, and a *question mark* indicates no investigation in this tissue.

Downloaded from http://www.jbc.org/ at INRA Institut National de la Recherche Agronomique on June 14, 2018

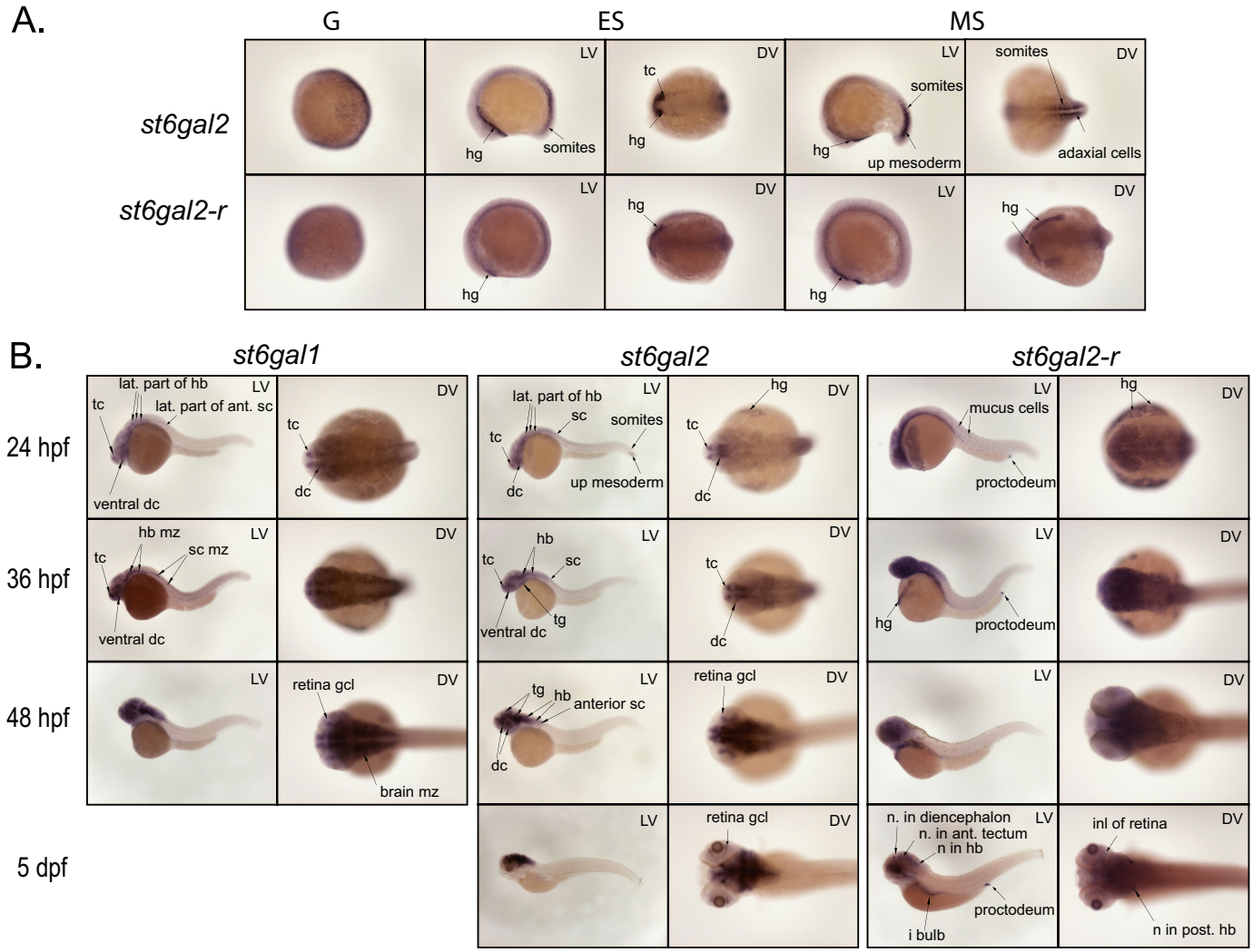


FIGURE 7. **Expression pattern of the zebrafish *st6gal* genes during zebrafish embryonic development using whole mount ISH at early developmental stages (*A*) and later developmental stages (*B*).** The *st6gal1* gene shows no labeling at early developmental stages (gastrula (*G*), early somitogenesis (*ES*), and middle somitogenesis (*MS*)). *st6gal1* is found in the ventricular part of telencephalon, in the ventral part of diencephalon, in the lateral part of the hindbrain (discontinuous patches of cells), and in the lateral part of the spinal cord at 24 and 36 hpf. *st6gal1* gene expression is detected in the ganglion cell layer of the retina (except in the proliferative zone), cranial ganglia, and brain marginal zone at 48 hpf and is not detected later. The *st6gal2* gene is expressed at the gastrula stage in the prechordal plate mesendoderm and at early somitogenesis in polster, anterior border of neural plate, adaxial cells, formed somites, and also weakly in segmental plates. At middle somitogenesis, it is expressed in polster, anterior head epidermis, caudal adaxial cells, and the last formed somites. At 24 hpf, the *st6gal2* gene is expressed in tail somites and adaxial cells and weakly in the CNS; later, at 36 hpf, a stronger signal is found in the marginal zone, and no labeling is detected in tectum. At 48 hpf and 5 days postfertilization, the *st6gal2* transcript is detected in anterior diencephalon and in the ganglion cell layer of retina but not in the proliferative zone. The *st6gal2-r* gene is weakly expressed at the gastrula stage and shows also a weak basal level of expression at early somitogenesis and middle somitogenesis and more labeling in hatching glands. At later developmental stages, from pharyngula to larva stages (24 hpf to 5 days postfertilization (*dpf*)), the *st6gal2-r* gene is detected in hatching glands, mucus cells, proctodeum, intestinal bulb, gall bladder, inner nuclear cell layer of retina (except in the proliferative zone), nucleus in diencephalon, nucleus in anterior tectum, and nucleus in hindbrain at the larval stage. *hg*, hatching gland; *inl of retina*, inner nuclear layer of retina; *i bulb*, intestinal bulb; *n in diencephalon*, nucleus in diencephalon; *n in ant tectum*, nucleus in anterior tectum; *n in hb*, nucleus in hindbrain; *up mesoderm*, unsegmented paraxial mesoderm; *tg*, tegmentum; *ventral dc*, ventral diencephalon. *LV*, lateral view; *DV*, dorsal view.

detected a sialyltransferase activity at the cell surface involved in cell-cell recognition. Although the relationships between all of the sialyltransferase families are not yet established, the *st6gal* gene family could constitute the most ancient sialyltransferase family described in animals (2). Because this *st6gal1/2* gene is retrieved from most studied arthropod and deuterostome genomes, we can deduce that it has disappeared independently in several lineages, as in the cnidarians, the lophotrochozoa (mollusks and annelids), the hymenoptera insects *A. mellifera* and *N. vitripennis*, nematodes such as *C. elegans* (76), the sea urchin *S. purpuratus*, and the tunicates *C. intestinalis* and *C. savignyi*. The reason for *st6gal1/2* gene loss in these taxa must be related to the primary function of this gene product. Given the small number of invertebrate genomes explored so far, the information available in protostomes and deuterostomes is quite fragmentary and has been mainly documented in *Drosophila* and vertebrates. Sialylation in insects has long been controversial (32, 35, 77), and recently, DSiaT, a unique *st6gal* gene, has been characterized in *Drosophila* (31). It is exclusively expressed in a subset of neurons in late embryonic stage 17, in the optic lobe of third instar larva and in the region of olfactory projection neurons in adult head (35). The encoded enzyme was found to be involved in the function of a voltage-gated sodium channel and neuromuscular junction and appears to be

Downloaded from http://www.jbc.org/ at INRA Institut National de la Recherche Agronomique on June 14, 2018

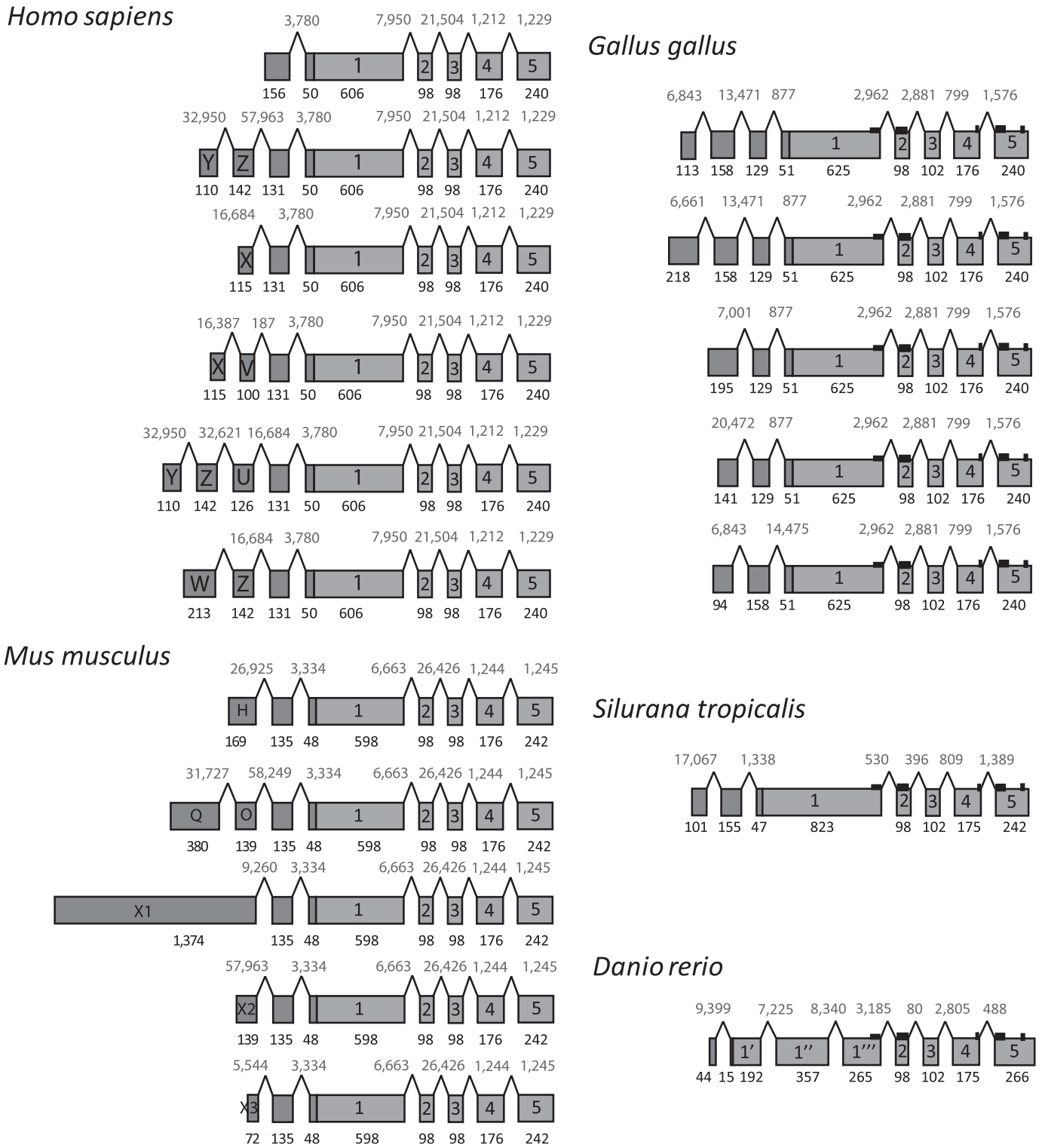


FIGURE 8. **Schematic diagram of the *st6gal1* transcripts described in various vertebrate species.** Depicted is the genomic organization of the *st6gal1* gene in *H. sapiens*, *M. musculus*, *G. gallus*, *S. tropicalis*, and *D. rerio*. The pattern of splicing is conserved during evolution in the 3′-end of *st6gal1* genes. In higher vertebrates, the *st6gal1* gene shows multiple and variable 5′-UT exons spanning large genomic distances that give rise to multiple transcripts due to the use of multiple transcriptional start sites. In lower vertebrates, the *st6gal1* gene shows one or two 5′-UT exons and a unique transcriptional start site.

essential for the regulation of nervous system function (36). Moreover, this *Drosophila* protein exhibits notable preferred enzymatic activity toward LacdiNAc substrates over LacNAc termini in *in vitro* assays (31), despite the fact that no evidence for the presence of LacdiNAc or LacNAc could be established *in vivo* (34). Mammalian *st6gal1* and *st6gal2* genes described previously have counterparts in all vertebrates examined, except for the lampreys, the living representatives of jawless vertebrates (agnatha), in which the three *st6gal* gene sequences form a sister group to all other vertebrate sequences. Two of

Downloaded from http://www.jbc.org/ at INRA Institut National de la Recherche Agronomique on June 14, 2018

these three *st6gal* genes were amplified by PCR (supplemental Fig. 5) from a 6–10-day embryonic cDNA library kindly provided by Prof. J. Langeland (78), indicating their expression during embryogenesis, whereas the third one appears to be absent (data not shown). Our phylogenetic analysis indicates that the single *st6gal* gene found in arthropods, the two copies found in amphioxus, and the three copies found in lampreys are orthologous to all vertebrate *st6gal* genes.

In order to explain the origin of *st6gal1* and *st6gal2* gene duplication in vertebrates, we compared the environment of each identified *st6gal* gene locus and determined their paralogy or orthology relationships. We pointed out a disruption in the conserved synteny of *st6gal1* loci in teleost fishes further suggesting a chromosomal rearrangement. These translocation events are known to occur at higher rates in fish genomes compared with tetrapod genomes (79). On the other hand, *st6gal2* synteny was maintained during vertebrate evolution. Moreover, intraspecific comparisons of chromosome segments inside vertebrates revealed that blocks of paralogous genes, named paralogons, can be identified (EPGD (80), CHSMiner (81)). Large sets of paralogons have been interpreted as a result of two rounds of genome duplications that occurred early in vertebrate evolution. The first round R1 probably occurred around 550 MYA, before the separation of lampreys from jawed vertebrates (gnathostomata). The second round R2 dates to about 474 MYA, after the emergence of lamprey and before cartilaginous fishes (chondrichthyan) divergence (82, 83). Identification of paralogons in the vertebrate genome and our calculations indicate that the *st6gal1* and *st6gal2* split dates back to this period and lead us to assume that one of the *st6gal* genes duplicated from R1 was lost. Subsequently, a third WGD R3 occurred ~350 MYA in the ray-finned fish lineage, after emergence of lobe-finned fishes (65, 79, 84–86), leading to the paralogon pair including *st6gal2* and *st6gal2-r* genes found in the zebrafish genome. The *st6gal2-r* gene was maintained in zebrafish but lost over time in other fish lineages, probably due to functional redundancy because both genes show similar patterns of expression during development.

A Scenario of Tissue Expression Evolution in Vertebrates—Our EST analysis using PCA highlighted another differential profile of expression of lower vertebrate *st6gal1* genes compared with higher vertebrates. *st6gal1* genes from fishes and amphibians form a cluster with all of the vertebrate *st6gal2* genes, whereas mammalian and avian *st6gal1* genes are found apart. This suggests an evolutionary change of the expression profile of *st6gal1* gene in amniotes. Using ISH in embryonic zebrafish tissues, we found overlapping territories of expression of *st6gal1* and *st6gal2* genes maintained in the adult brain in several vertebrate species (1). Surprisingly, *st6gal2* and *st6gal2-r* genes exhibit differential patterns of expression. Both genes are expressed at early developmental stages, and the gastrula stage marks their onset of expression. The *st6gal2-r* gene is primarily detected in hatching gland cells, which produce metalloprotease choriolytic enzymes HCE and LCE digesting egg envelope (chorion) at the time of embryo hatching (87). This suggests a role in the process of hatching gland differentiation (same time as differentiation of notochord and paraxial mesoderm), in mucous cells and proctodeum.

We next analyzed adult tissue distribution of *st6gal* genes using RT-PCR in lower vertebrates. We observed that adult *D. rerio* and *S. tropicalis* express the *st6gal2* gene mainly in the brain, as previously reported for the mammalian *st6gal2* gene (5, 9). It is also highly expressed in ovaries and to a lesser extent in intestine. Such slight variations in the *st6gal2* expression profile have been reported for the bovine gene, which is significantly amplified from lung and intestine adult tissues (88). Both organisms also express the *st6gal1* gene in ovaries and intestine, but their expression profile is more heterogeneous in lower vertebrates, which is in sharp contrast to the ubiquitous mammalian *st6gal1* gene expression profile. Interestingly, in *D. rerio*, *st6gal1* is found in kidney and is notably absent in liver, whereas in *S. tropicalis*, it is amplified in liver tissue and is almost not detected in kidney. Analysis of the EST profile of the chicken (Gga.1148) provided by the GenBank™ data base illustrates expression of the *st6gal1* gene in several adult tissues, such as brain, liver, thymus, muscle, ovary, or bursa of Fabricius. Because it is expressed in the zebrafish kidney, the frog liver and the bird bursa of Fabricius, which are the chief organs of B-cell development corresponding to the mammalian bone marrow (89), we hypothesize that the *st6gal1* gene product would have gained a progressive function in lymphoid organs during evolution. Indeed, genetically *st6gal1*-altered mice provided fragmentary insights into *st6gal1* biological function, showing that the enzyme is implicated in immune system function (16, 90). We could also predict that this gene is expressed in the thymus of teleosts and amphibians.

In summary, we observe a relative conservation of the *st6gal2* expression profile in vertebrates, suggesting that it could be involved in molecular mechanisms that support neurogenesis (91) and thus would have conserved this role in the CNS already recorded in *Drosophila*. However, its expression is also maintained in ovaries and intestine in lower vertebrates and mammals, further suggesting that the *st6gal2* gene might have evolved new functions acquired within the 75 million years that elapsed between the R2 and the osteichthyans radiation because during that period, the *st6gal2* gene evolved more rapidly than the *st6gal1* gene, as illustrated by its longer branch lengths (Fig. 1). Up to now, its physiological function in vertebrates remains unknown, although it has been shown to be implicated in apoptosis (92).

Analysis of 5′-End of st6gal Genes—In order to better understand the pattern of *st6gal* gene expression diversification in vertebrates, we assessed which factors might have influenced their expression at the genomic level. BLAST searches of zebrafish and amphibian EST resources of the NCBI data base tend to demonstrate conservation over vertebrate evolution of the number of TSS for the *st6gal2* gene already described in mammals (88, 93), correlating with their conserved pattern of expression (data not shown). We thus focused on the *st6gal1* gene, which is ubiquitously detected in mammalian adult tissues, strongly expressed by the human liver, and transiently up-regulated during inflammation and in several cancers, due to multiple promoter-driven 5′-UT exons (1, 94). Our 5′-RACE analysis of fish and frog *st6gal1* genes in several adult tissues clearly demonstrated the use of a unique TSS and the presence of one 5′-UT exon in *D. rerio* and two 5′-UT exons in *S. tropi-*

Downloaded from http://www.jbc.org/ at INRA Institut National de la Recherche Agronomique on June 14, 2018

calis tissues. Our data further suggest that major changes have occurred at the level of regulatory *cis*-acting sequences and point to a still hypothetical rapid evolution of their regulatory genomic sequences that might be due to greater relaxation of evolutionary constraints often considered to be the driving force in the evolution of genetic networks (95). This rapid complication of the genetic/epigenetic regulation of expression of the *st6gal1* gene has led to a diversification of the tissue distribution and also of function in higher vertebrates. Indeed, phenotypic variation in α2,6-sialylation of *N*-glycosylproteins has been observed in various animals and in particular in mammals despite genetic conservation of their translated gene sequences (24). The patterns of tissue α2,6-sialylation of *N*-glycosylproteins differ widely among mammals, even among closely related taxa, such as mice and humans, which diverged only 96 MYA (68, 96–98), or great apes and humans, which diverged 13–14 MYA (28, 96). We suggest a still on-going evolution and neofunctionalization of *st6gal1* genes in mammals, which could explain differences in influenza virus infection of airway epithelial cells (24).

In the context of the rapid evolution of functions of *st6gal* genes in vertebrates, our data further suggest that the *st6gal2* genes might have maintained an ancestral function due to their localized expression in vertebrate CNS and similar biochemical activity compared with DSiaT (31). Mammalian recombinant ST6Gal I and ST6Gal II enzymes produced in heterologous systems like *Spodoptera frugiperda* (Sf-9) mediate the addition of α2,6-linked sialic acid to Galβ1–4GlcNAc (LacNAc) and to GalNAcβ1–4GlcNAc (LacDiNAc) disaccharides, respectively (7, 8). It has been previously shown that the ST6Gal I/II enzyme from *D. melanogaster* prefers LacdiNAc-bearing substrates over LacNAc (31), an enzymatic characteristic that was maintained in mammalian ST6Gal II enzymes (8, 10). Although information on these enzymes is lacking for lower vertebrates, we postulate that these biochemical properties extend to all of the vertebrate ST6Gal enzymes. Interestingly, the distribution of LacdiNAc in mammals is very limited, and LacdiNAc might be substituted by 4-*O*-sulfated-, α1,3-fucosylated, or α2,6-sialylated derivatives (99). As indicated by these authors, the glycans bearing LacdiNAc are notably recorded in pituitary glycoprotein hormones and tenascin-R produced by oligodendrocytes and small interneurons in the hippocampus and cerebellum. Other glycoproteins concerned are glycodelin, with potent immunosuppressive and contraceptive activities in humans, and zona pellucida glycoproteins from murine eggs. We suggest that the ancestral ST6Gal I/II accept GalNAc substrates better than Gal substrates and that the new properties of amniote ST6Gal I toward Gal substrates may help to evade pathogens, as suggested previously (98).

Variations in the ST6Gal Sequences during Evolution—Comparison of cumulate numbers of amino acid substitutions in the catalytic part of the ST6Gal enzymes between corresponding branches in the ST6Gal I and ST6Gal II trees raised intriguing points that deserve discussion. Within the ST6Gal II tree, we noticed short branches between each species, supporting our hypothesis of a conserved role throughout animal evolution, due to selective pressure. However, the higher number of substitutions at the base of the ST6Gal II clade compared with the base of the ST6Gal I clade is indicative of changes difficult to interpret because no significant changes within the conserved motifs are recorded (see the R2/osteichthyans line in Table 2). In contrast, there is a long branch leading to ST6Gal I teleosts from osteichthyans ancestors, associated with an accumulation of substitution within the family motif b and sialylmotif S (1). This feature suggests an original function of this enzyme within bony fishes, although this hypothesis requires further study. In the tetrapod clade, we observed a greater number of substitutions in the branch leading to amniote ST6Gal I compared with the amniote ST6Gal II branch, associated with changes in the sialylmotif L. More interestingly, this accumulation is more clear in the branch leading to mammals, and this affects the sialylmotif L and family motif b. We can deduce that there is a progressive change in the function of ST6Gal I from tetrapods to mammals and that these changes are probably not of the same nature as those observed in teleosts.

As mentioned previously, there is a shift in the preferred specificity of acceptor substrate during the evolution of ST6Gal I in vertebrates (8). It is interesting to note that site-directed mutagenesis of the sialylmotif L and S conserved amino acids in the rat ST6Gal I and in the human ST3Gal I indicated that they are implicated in the donor (CMP-Neu5Ac) and acceptor binding, respectively (100–102). Thus, the present analysis of substitution changes suggests that the LacdiNAc to LacNAc shift would be correlative to the accumulation of substitutions during amniote to mammal evolution.

The stem region of vertebrate ST6Gal enzymes and their coding sequences also show important variations, and we observed more indel events in the branches leading to teleosts and to amniote ST6Gal I than in the ST6Gal II counterparts. The fish *st6gal1* genes exhibit two additional intron sequences that probably result from successive insertions of two spliceosomal introns in the first exon of the fish *st6gal1* gene, after teleost radiation. Moreover, we noticed an insertion of repeated genetic sequences in the amphibian *st6gal1* gene, leading to the formation of an acidic supercoiled region in the ST6Gal I stem region that may have variable impacts on the subcellular distribution in the *trans*-Golgi network and enzymatic activity of ST6Gal I. We molecularly cloned the amphibian *st6gal1* cDNA sequence and confirmed the presence of acidic repetitive sequences (REKDLE) in the *S. tropicalis* ST6Gal I protein sequence, which is also found in the bifunctional α2,3/α2,8-sialyltransferase of *Helicobacter acinonychis* Sheeba (YP_665016). The signals and mechanisms mediating Golgi localization have been studied extensively for various mammalian glycosyltransferases (reviewed in Refs. 103–106). The cytosolic tail/transmembrane domain/stem region of human ST6Gal I is probably implicated in subcellular traffic through functional homodimerization and/or interactions with other proteins, such as COP-I coated vesicle anterograde traffic of GT or COG (103, 107–111) or Golgi retention (112–114) and in the modulation of its enzymatic activity through substrate recognition (61). It has also been shown that the cytosolic tail/transmembrane domain/stem region impacts mammalian ST6Gal I secretion via BACE-1 aminopeptidase activity in Alzheimer disease (115–118). Altogether, our data on the molecular cloning of lower vertebrate *st6gal1* genes and their molecular evo-

Downloaded from http://www.jbc.org/ at INRA Institut National de la Recherche Agronomique on June 14, 2018

lution raise the question of the evolution of vertebrate ST6Gal I with regard to their subcellular localization, interaction with other glycosyltransferases, and activity *in vivo*.

The traditional paradigm is that duplication releases a selective constraint on one paralogous gene, offering the possibility of the appearance of new function(s). Here, we show on-going neofunctionalization of *st6gal1* genes in amniotes and maybe in teleosts. The consequence of neofunctionalization of *st6gal1* genes is a net increase in expression complexity following duplication. Those genes, such as *st6gal1*, implicated in immunity, host defense, reproduction, and olfaction are rapidly evolving, whereas those, such as *st6gal2*, implicated in intracellular signaling, neurogenesis, and neurophysiology are slowly evolving (119). Alternatively, another model of evolution of duplicated genes named subfunctionalization suggests duplication-degeneration-complementation (the DDC model), leading to pleiotropic expression (120). This appears to be more or less the case with *st6gal2* and *st6gal2-r* genes in zebrafish, issued from a specific teleost WGD R3, which has a complementary pattern of expression in adult tissues and embryos. Overlapping expression domains could produce fine graining of gene function (85).

A relationship has been established between the breadth of expression, expressed as the number of tissues in which ESTs are recorded, and evolutionary rates. Briefly, the wider the tissue expression, the weaker the evolutionary rate, a fact attributable to a greater selective pressure when a gene is expressed over a variety of tissues (121, 122). In the case of *st6gal* genes, we observe quite the opposite because *st6gal2* shows low variations of sequence despite the reduced number of tissues in which it is expressed. Thus, the conservation of the function of this gene would be driven by purifying selection. In contrast, the high evolutionary rates observed in *st6gal1* gene sequences instead result from changes of function and of specificity and an increase in expression breadth, through an increase in the number of TSSs. ST6Gal I progressively acquired functions in the immune system from a probable ancestral role in embryonic and adult CNS.

Acknowledgments—We acknowledge Prof. Mohammed Lemdani (University of Lille Nord de France, Lille 2, France) and Dr. François Foulquier (University of Lille Nord de France, Lille 1, France) for helpful discussions; Marianne Gérard, Amandine Verlande, and Leila Bekri for excellent technical assistance; and Dr. Rosella Mollicone (University of Paris Sud XI, France) and Dr. Benoit Laporte (University of Limoges, France) for constant interest in the work. We thank Prof. Jim Langeland (University of Wisconsin) for providing a P. marinus cDNA library and Prof. Jean-François Bodart (University of Lille Nord de France, Lille 1, France) for providing S. tropicalis tissues.

REFERENCES

1. Harduin-Lepers, A. (2010) *Glycobiology Insights* **2,** 29–61
2. Harduin-Lepers, A., Mollicone, R., Delannoy, P., and Oriol, R. (2005) *Glycobiology* **15,** 805–817
3. Harduin-Lepers, A., Vallejo-Ruiz, V., Krzewinski-Recchi, M. A., Samyn-Petit, B., Julien, S., and Delannoy, P. (2001) *Biochimie* **83,** 727–737
4. Harduin-Lepers, A., Recchi, M. A., and Delannoy, P. (1995) *Glycobiology* **5,** 741–758
5. Krzewinski-Recchi, M. A., Julien, S., Juliant, S., Teintenier-Lelièvre, M., Samyn-Petit, B., Montiel, M. D., Mir, A. M., Cerutti, M., Harduin-Lepers, A., and Delannoy, P. (2003) *Eur. J. Biochem.* **270,** 950–961
6. Kurosawa, N., Kawasaki, M., Hamamoto, T., Nakaoka, T., Lee, Y. C., Arita, M., and Tsuji, S. (1994) *Eur. J. Biochem.* **219,** 375–381
7. Mercier, D., Wierinckx, A., Oulmouden, A., Gallet, P. F., Palcic, M. M., Harduin-Lepers, A., Delannoy, P., Petit, J. M., Levéziel, H., and Julien, R. (1999) *Glycobiology* **9,** 851–863
8. Rohfritsch, P. F., Joosten, J. A., Krzewinski-Recchi, M. A., Harduin-Lepers, A., Laporte, B., Juliant, S., Cerutti, M., Delannoy, P., Vliegenthart, J. F., and Kamerling, J. P. (2006) *Biochim. Biophys. Acta* **1760,** 685–692
9. Takashima, S., Tsuji, S., and Tsujimoto, M. (2002) *J. Biol. Chem.* **277,** 45719–45728
10. Takashima, S., Tsuji, S., and Tsujimoto, M. (2003) *J. Biochem.* **134,** 287–296
11. Kitagawa, H., and Paulson, J. C. (1994) *J. Biol. Chem.* **269,** 17872–17878
12. O'Hanlon, T. P., Lau, K. M., Wang, X. C., and Lau, J. T. (1989) *J. Biol. Chem.* **264,** 17389–17394
13. Wen, D. X., Svensson, E. C., and Paulson, J. C. (1992) *J. Biol. Chem.* **267,** 2512–2518
14. Dall'Olio, F., Chiricolo, M., Ceccarelli, C., Minni, F., Marrano, D., and Santini, D. (2000) *Int. J. Cancer* **88,** 58–65
15. Varki, A. (1997) *FASEB J.* **11,** 248–255
16. Hennet, T., Chui, D., Paulson, J. C., and Marth, J. D. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95,** 4504–4509
17. Lo, N. W., and Lau, J. T. (1999) *Glycobiology* **9,** 907–914
18. Collins, B. E., Smith, B. A., Bengtson, P., and Paulson, J. C. (2006) *Nat. Immunol.* **7,** 199–206
19. Crocker, P. R., Paulson, J. C., and Varki, A. (2007) *Nat. Rev. Immunol.* **7,** 255–266
20. Wang, P. H., Lee, W. L., Lee, Y. R., Juang, C. M., Chen, Y. J., Chao, H. T., Tsai, Y. C., and Yuan, C. C. (2003) *Gynecol. Oncol.* **89,** 395–401
21. Chiricolo, M., Malagolini, N., Bonfiglioli, S., and Dall'Olio, F. (2006) *Glycobiology* **16,** 146–154
22. Hedlund, M., Ng, E., Varki, A., and Varki, N. M. (2008) *Cancer Res.* **68,** 388–394
23. Lee, M., Park, J. J., and Lee, Y. S. (2010) *Oncol. Rep.* **23,** 757–761
24. Gagneux, P., Cheriyan, M., Hurtado-Ziola, N., van der Linden, E. C., Anderson, D., McClure, H., Varki, A., and Varki, N. M. (2003) *J. Biol. Chem.* **278,** 48245–48250
25. Gagneux, P., and Varki, A. (1999) *Glycobiology* **9,** 747–755
26. Dalziel, M., Lemaire, S., Ewing, J., Kobayashi, L., and Lau, J. T. (1999) *Glycobiology* **9,** 1003–1008
27. Jamieson, J. C., McCaffrey, G., and Harder, P. G. (1993) *Comp. Biochem. Physiol. B* **105,** 29–33
28. Appenheimer, M. M., Huang, R. Y., Chandrasekaran, E. V., Dalziel, M., Hu, Y. P., Soloway, P. D., Wuensch, S. A., Matta, K. L., and Lau, J. T. (2003) *Glycobiology* **13,** 591–600
29. Baum, L. G., and Paulson, J. C. (1990) *Acta Histochem. Suppl.* **40,** 35–38
30. Bishop, J. R., and Gagneux, P. (2007) *Glycobiology* **17,** 23R–34R
31. Koles, K., Irvine, K. D., and Panin, V. M. (2004) *J. Biol. Chem.* **279,** 4346–4357
32. Angata, T., and Varki, A. (2002) *Chem. Rev.* **102,** 439–469
33. Aoki, K., Perlman, M., Lim, J. M., Cantu, R., Wells, L., and Tiemeyer, M. (2007) *J. Biol. Chem.* **282,** 9127–9142
34. Koles, K., Lim, J. M., Aoki, K., Porterfield, M., Tiemeyer, M., Wells, L., and Panin, V. (2007) *Glycobiology* **17,** 1388–1403
35. Koles, K., Repnikova, E., Pavlova, G., Korochkin, L. I., and Panin, V. M. (2009) *Glycoconj. J.* **26,** 313–324
36. Repnikova, E., Koles, K., Nakamura, M., Pitts, J., Li, H., Ambavane, A., Zoran, M. J., and Panin, V. M. (2010) *J. Neurosci.* **30,** 6466–6476
37. Harduin-Lepers, A., Petit, D., Mollicone, R., Delannoy, P., Petit, J. M., and Oriol, R. (2008) *BMC Evol. Biol.* **8,** 258
38. Kanehisa, M., and Goto, S. (2000) *Nucleic Acids Res.* **28,** 27–30
39. Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., and Hirakawa, M. (2010) *Nucleic Acids Res.* **38,** D355–D360
40. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M. (2006) *Nucleic*

Downloaded from http://www.jbc.org/ at INRA Institut National de la Recherche Agronomique on June 14, 2018

Acids Res. **34,** D354–D357
41. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402
42. Patel, R. Y., and Balaji, P. V. (2006) *Glycobiology* **16,** 108–116
43. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680
44. Castresana, J. (2000) *Mol. Biol. Evol.* **17,** 540–552
45. Guindon, S., and Gascuel, O. (2003) *Syst. Biol.* **52,** 696–704
46. Kumar, S., Tamura, K., and Nei, M. (2004) *Brief Bioinform.* **5,** 150–163
47. Felsenstein, J. (2002) *Phylip (Phylogeny Inference Package)*, Version 3.6a3, University of Washington, Seattle, WA
48. Petit, D., Maftah, A., Julien, R., and Petit, J. M. (2006) *J. Mol. Evol.* **63,** 353–364
49. Martin, R., Gallet, P. F., Rocha, D., and Petit, D. (2009) *Recent Pat. DNA Gene Seq.* **3,** 63–71
50. Panopoulou, G., and Poustka, A. J. (2005) *Trends Genet.* **21,** 559–567
51. Vandepoele, K., De Vos, W., Taylor, J. S., Meyer, A., and Van de Peer, Y. (2004) *Proc. Natl. Acad. Sci. U.S.A.* **101,** 1638–1643
52. Hammer, Ø., Harper, D. A. T., and Ryan, P. D. (2001) *Paleontol. Electron.* **4,** 9
53. McLysaght, A., Hokamp, K., and Wolfe, K. H. (2002) *Nat. Genet.* **31,** 200–204
54. Ermonval, M., Petit, D., Le Duc, A., Kellermann, O., and Gallet, P. F. (2009) *Glycoconj. J.* **26,** 477–493
55. Chang, L. Y., Mir, A. M., Thisse, C., Guérardel, Y., Delannoy, P., Thisse, B., and Harduin-Lepers, A. (2009) *Glycoconj. J.* **26,** 263–275
56. Westerfield, M. (1995) *The Zebrafish Book: A Guide for the Laboratory Use of Zebrafish (Danio rerio)*, p. 385, University of Oregon Press, Eugene, OR
57. Thisse, B., Heyer, V., Lux, A., Alunni, V., Degrave, A., Seiliez, I., Kirchner, J., Parkhill, J. P., and Thisse, C. (2004) *Methods Cell Biol.* **77,** 505–519
58. Thisse, C., and Thisse, B. (1998) in *Zebrafish Science Monitor*, Vol. 5, pp. 8–9, University of Oregon Press, Eugene, OR
59. Thisse, C., and Thisse, B. (2008) *Nat. Protoc.* **3,** 59–69
60. Blair, J. E., and Hedges, S. B. (2005) *Mol. Biol. Evol.* **22,** 2275–2284
61. Legaigneur, P., Breton, C., El Battari, A., Guillemot, J. C., Auge, C., Malissard, M., Berger, E. G., and Ronin, C. (2001) *J. Biol. Chem.* **276,** 21608–21617
62. Ohno, S. (1970) *Evolution by Gene Duplication*, Springer-verlag, Heidelberg
63. Catchen, J. M., Conery, J. S., and Postlethwait, J. H. (2009) *Genome Res.* **19,** 1497–1505
64. Ehrlich, J., Sankoff, D., and Nadeau, J. H. (1997) *Genetics* **147,** 289–296
65. Amores, A., Force, A., Yan, Y. L., Joly, L., Amemiya, C., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y. L., Westerfield, M., Ekker, M., and Postlethwait, J. H. (1998) *Science* **282,** 1711–1714
66. Jaillon, O., Aury, J. M., Brunet, F., Petit, J. L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., Nicaud, S., Jaffe, D., Fisher, S., Lutfalla, G., Dossat, C., Segurens, B., Dasilva, C., Salanoubat, M., Levy, M., Boudet, N., Castellano, S., Anthouard, V., Jubin, C., Castelli, V., Katinka, M., Vacherie, B., Biémont, C., Skalli, Z., Cattolico, L., Poulain, J., De, Berardinis, V., Cruaud, C., Duprat, S., Brottier, P., Coutanceau, J. P., Gouzy, J., Parra, G., Lardier, G., Chapple, C., McKernan, K. J., McEwan, P., Bosak, S., Kellis, M., Volff, J. N., Guigó, R., Zody, M. C., Mesirov, J., Lindblad-Toh, K., Birren, B., Nusbaum, C., Kahn, D., Robinson-Rechavi, M., Laudet, V., Schachter, V., Quétier, F., Saurin, W., Scarpelli, C., Wincker, P., Lander, E. S., Weissenbach, J., and Roest Crollius, H. (2004) *Nature* **431,** 946–957
67. Taylor, J. S., Van de Peer, Y., and Meyer, A. (2001) *Curr. Biol.* **11,** R1005–1008
68. Altheide, T. K., Hayakawa, T., Mikkelsen, T. S., Diaz, S., Varki, N., and Varki, A. (2006) *J. Biol. Chem.* **281,** 25689–25702
69. Comelli, E. M., Head, S. R., Gilmartin, T., Whisenant, T., Haslam, S. M., North, S. J., Wong, N. K., Kudo, T., Narimatsu, H., Esko, J. D., Drickamer, K., Dell, A., and Paulson, J. C. (2006) *Glycobiology* **16,** 117–131
70. Ishii, A., Ikeda, T., Hitoshi, S., Fujimoto, I., Torii, T., Sakuma, K., Nakakita, S., Hase, S., and Ikenaka, K. (2007) *Glycobiology* **17,** 261–276
71. Nairn, A. V., York, W. S., Harris, K., Hall, E. M., Pierce, J. M., and Moremen, K. W. (2008) *J. Biol. Chem.* **283,** 17298–17313
72. Varki, A. (2007) *Nature* **446,** 1023–1029
73. Schauer, R. (2009) *Curr. Opin. Struct. Biol.* **19,** 507–514
74. Nichols, S. A., Dirks, W., Pearse, J. S., and King, N. (2006) *Proc. Natl. Acad. Sci. U.S.A.* **103,** 12451–12456
75. Müller, W. E., Arendes, J., Kurelec, B., Zahn, R. K., and Müller, I. (1977) *J. Biol. Chem.* **252,** 3836–3842
76. Danchin, E. G., Gouret, P., and Pontarotti, P. (2006) *BMC Evol. Biol.* **6,** 5
77. Marchal, I., Jarvis, D. L., Cacan, R., and Verbert, A. (2001) *Biol. Chem.* **382,** 151–159
78. Tomsa, J. M., and Langeland, J. A. (1999) *Dev. Biol.* **207,** 26–37
79. Ravi, V., and Venkatesh, B. (2008) *Curr. Opin. Genet. Dev.* **18,** 544–550
80. Ding, G., Sun, Y., Li, H., Wang, Z., Fan, H., Wang, C., Yang, D., and Li, Y. (2008) *Nucleic Acids Res.* **36,** D255–D262
81. Wang, Z., Ding, G., Yu, Z., Liu, L., and Li, Y. (2009) *Algorithms Mol. Biol.* **4,** 2
82. Wang, Y., and Gu, X. (2000) *J. Mol. Evol.* **51,** 88–96
83. Robinson-Rechavi, M., Boussau, B., and Laudet, V. (2004) *Mol. Biol. Evol.* **21,** 580–586
84. Meyer, A., and Van de Peer, Y. (2005) *BioEssays* **27,** 937–945
85. Aparicio, S. (2000) *Trends Genet.* **16,** 54–56
86. Christoffels, A., Koh, E. G., Chia, J. M., Brenner, S., Aparicio, S., and Venkatesh, B. (2004) *Mol. Biol. Evol.* **21,** 1146–1151
87. Inohaya, K., Yasumasu, S., Ishimaru, M., Ohyama, A., Iuchi, I., and Yamagami, K. (1995) *Dev. Biol.* **171,** 374–385
88. Laporte, B., Gonzalez-Hilarion, S., Maftah, A., and Petit, J. M. (2009) *Glycobiology* **19,** 1082–1093
89. Boehm, T., and Bleul, C. C. (2007) *Nat. Immunol.* **8,** 131–135
90. Dall'Olio, F. (2000) *Glycoconj. J.* **17,** 669–676
91. Galliot, B., Quiquand, M., Ghila, L., de Rosa, R., Miljkovic-Licina, M., and Chera, S. (2009) *Dev. Biol.* **332,** 2–24
92. Mannherz, O., Mertens, D., Hahn, M., and Lichter, P. (2006) *Genomics* **87,** 665–672
93. Lehoux, S., Groux-Degroote, S., Cazet, A., Dhaenens, C. M., Maurage, C. A., Caillet-Boudin, M. L., Delannoy, P., and Krzewinski-Recchi, M. A. (2010) *Glycoconj. J.* **27,** 99–114
94. Wang, X., Vertino, A., Eddy, R. L., Byers, M. G., Jani-Sait, S. N., Shows, T. B., and Lau, J. T. (1993) *J. Biol. Chem.* **268,** 4355–4361
95. Cañestro, C., Catchen, J. M., Rodríguez-Marí, A., Yokoi, H., and Postlethwait, J. H. (2009) *PLoS Genet.* **5,** e1000496
96. Varki, A. (2010) *Proc. Natl. Acad. Sci. U.S.A.* **107,** Suppl. 2, 8939–8946
97. Varki, A. (2009) *Glycoconj. J.* **26,** 231–245
98. Varki, A. (2006) *Cell* **126,** 841–845
99. Kawar, Z. S., Haslam, S. M., Morris, H. R., Dell, A., and Cummings, R. D. (2005) *J. Biol. Chem.* **280,** 12810–12819
100. Datta, A. K. (2009) *Curr. Drug Targets* **10,** 483–498
101. Datta, A. K., and Paulson, J. C. (1995) *J. Biol. Chem.* **270,** 1497–1500
102. Datta, A. K., Sinha, A., and Paulson, J. C. (1998) *J. Biol. Chem.* **273,** 9608–9614
103. Tu, L., and Banfield, D. K. (2010) *Cell Mol. Life Sci.* **67,** 29–41
104. Tu, L., Tai, W. C., Chen, L., and Banfield, D. K. (2008) *Science* **321,** 404–407
105. Patel, R. Y., and Balaji, P. V. (2007) *Protein Pept. Lett.* **14,** 601–609
106. Breton, C., Mucha, J., and Jeanneau, C. (2001) *Biochimie* **83,** 713–718
107. Chen, T. L., Chen, C., Bergeron, N. Q., Close, B. E., Bohrer, T. J., Vertel, B. M., and Colley, K. J. (2003) *Glycobiology* **13,** 109–117
108. Ma, J., Qian, R., Rausa, F. M., 3rd, and Colley, K. J. (1997) *J. Biol. Chem.* **272,** 672–679
109. Ma, J., Simonovic, M., Qian, R., and Colley, K. J. (1999) *J. Biol. Chem.* **274,** 8046–8052
110. Hassinen, A., Rivinoja, A., Kauppila, A., and Kellokumpu, S. (2010) *J. Biol. Chem.* **285,** 17771–17777
111. Rivinoja, A., Hassinen, A., Kokkonen, N., Kauppila, A., and Kellokumpu, S. (2009) *J. Cell Physiol.* **220,** 144–154
112. Munro, S. (1995) *EMBO J.* **14,** 4695–4704
113. Munro, S. (1991) *EMBO J.* **10,** 3577–3588
114. Fenteany, F. H., and Colley, K. J. (2005) *J. Biol. Chem.* **280,** 5423–5429
115. Kitazume, S., Nakagawa, K., Oka, R., Tachida, Y., Ogawa, K., Luo, Y.,

Downloaded from http://www.jbc.org/ at INRA Institut National de la Recherche Agronomique on June 14, 2018

Citron, M., Shitara, H., Taya, C., Yonekawa, H., Paulson, J. C., Miyoshi, E., Taniguchi, N., and Hashimoto, Y. (2005) *J. Biol. Chem.* **280,** 8589–8595
116. Kitazume, S., Oka, R., Ogawa, K., Futakawa, S., Hagiwara, Y., Takikawa, H., Kato, M., Kasahara, A., Miyoshi, E., Taniguchi, N., and Hashimoto, Y. (2009) *Glycobiology* **19,** 479–487
117. Kitazume, S., Saido, T. C., and Hashimoto, Y. (2002) *Seikagaku* **74,** 1180–1183
118. Kitazume, S., Tachida, Y., Oka, R., Kotani, N., Ogawa, K., Suzuki, M., Dohmae, N., Takio, K., Saido, T. C., and Hashimoto, Y. (2003) *J. Biol. Chem.* **278,** 14865–14871
119. Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., Yamada, T., Nagayasu, Y., Doi, K., Kasai, Y., Jindo, T., Kobayashi, D., Shimada, A., Toyoda, A., Kuroki, Y., Fujiyama, A., Sasaki, T., Shimizu, A., Asakawa, S., Shimizu, N., Hashimoto, S., Yang, J., Lee, Y., Matsushima, K., Sugano, S., Sakaizumi, M., Narita, T., Ohishi, K., Haga, S., Ohta, F., Nomoto, H., Nogata, K., Morishita, T., Endo, T., Shin-I, T., Takeda, H., Morishita, S., and Kohara, Y. (2007) *Nature* **447,** 714–719
120. Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., and Postlethwait, J. (1999) *Genetics* **151,** 1531–1545
121. Duret, L., and Mouchiroud, D. (2000) *Mol. Biol. Evol.* **17,** 68–74
122. Subramanian, S., and Kumar, S. (2004) *Genetics* **168,** 373–381
123. Tsuji, S., Datta, A. K., and Paulson, J. C. (1996) *Glycobiology* **6,** v–vii

Downloaded from http://www.jbc.org/ at INRA Institut National de la Recherche Agronomique on June 14, 2018

Molecular Phylogeny and Functional Genomics of β-Galactoside α 2,6-Sialyltransferases That Explain Ubiquitous Expression of *st6gal1* Gene in Amniotes

Daniel Petit, Anne-Marie Mir, Jean-Michel Petit, Christine Thisse, Philippe Delannoy, Rafael Oriol, Bernard Thisse and Anne Harduin-Lepers

J. Biol. Chem. 2010, 285:38399-38414.
doi: 10.1074/jbc.M110.163931 originally published online September 20, 2010

Access the most updated version of this article at doi: 10.1074/jbc.M110.163931

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2010/09/20/M110.163931.DC1

This article cites 120 references, 31 of which can be accessed free at
http://www.jbc.org/content/285/49/38399.full.html#ref-list-1

Downloaded from http://www.jbc.org/ at INRA Institut National de la Recherche Agronomique on June 14, 2018