Analysis of a large cluster of SLC22 transporter genes, including novel USTS, reveals species-specific amplification of subsets of family members

Wei Wu, Michael E. Baker, Satish A. Eraly, Kevin T. Bush, and Sanjay K. Nigam

Departments of Pediatrics, Medicine, and Cellular Molecular Medicine, School of Medicine, University of California, San Diego, La Jolla, California

Submitted 6 August 2008; accepted in final form 27 April 2009

Wu W, Baker ME, Eraly SA, Bush KT, Nigam SK. Analysis of a large cluster of SLC22 transporter genes, including novel USTS, reveals species-specific amplification of subsets of family members. Physiol Genomics 38: 116–124, 2009. First published May 5, 2009; doi:10.1152/physiolgenomics.90309.2008.—When the organic anion transporter Oat1 was first identified as NKT (Lopez-Nieto CE, You G, Bush KT, Barros EJ, Beier DR, Nigam SK. J Biol Chem 272: 6471–6478, 1997), it was argued that it, together with Oct1, may be part of a larger subfamily (now known as SLC22) involved in organic ion and xenobiotic transport. The least studied among SLC22 transporters are the so-called unknown substrate transporters (USTs). Here, five novel genes located in a cluster on mouse chromosome 19, immediately between Slc22a8/Oat3 and Slc22a6/Oat1 and Slc22a19/Oat5, were identified as homologs of human USTS. These genes display preferential expression in liver and kidney, and one gene, AB056422, has several splicing variants with differential tissue expression and embryonic expression. Along with Slc22a6/Oat1, Slc22a8/Oat2, and Slc22a19/Oat3, these USTS define the largest known cluster of mammalian Slc22 genes. Given the established functions of Oats, these genes may also be involved in organic anion transport. Usts have characteristic motifs and share a signature residue in the possible active site of transmembrane domain 7, a conserved, positively charged, amino acid, Arg356, possibly a site for interaction with organic anions. In certain species, Oat1 and Oat3 appeared to be highly conserved, whereas the Ust part of this cluster appeared to undergo repeated species-specific amplification, suggesting strong environmental selection pressure, and perhaps providing an explanation for copy number variation in the human locus. One Ust amplification in mouse appears to be recent. This cluster may be coordinately regulated and under selective pressure in a species-specific manner.

organic anion transporter; evolution; copy number variation

SLC22 FAMILY GENES ENCODE structurally and evolutionarily conserved transporters, with critical roles in elimination of chemicals in multicellular organisms (6). When organic anion transporter 1 (Oat1) was identified as NKT, it was proposed that it, together with Oct1 and NLT (now Oat2), constituted a new family of genes (14). Since then, the family has grown to at least 18 genes in human and 17 in mouse (12). SLC22 family genes can be further divided into subgroups according to their substrate specificity and include organic cation transporters (OCTs) (10), carnitine transporters (OCTNs), organic anion transporters (OATs) (14) and so-called unknown substrate transporters (USTs) (7, 21). These genes share a similar genomic structure, each with 9 or 10 exons encoding a ~550

amino acid polypeptide. Models of OCTs (17, 19, 29) and OATs (16) predict a protein with 12 transmembrane domains, with a large extracellular domain at its NH2 terminus between transmembrane helices 1 and 2, a large intracellular domain between transmembrane segments 6 and 7, and intracellular NH2 and COOH termini (5, 14, 16).

Genetic and physiological analyses have revealed critical roles for these transporters (4, 8, 9, 25, 27). Substrates for some SLC22 genes have been determined by in vitro expression in Xenopus laevis oocytes and cell/organ cultures (3, 13, 22, 26). For example, OAT1 (NKT/SLC22A6 and OAT3/SLC22A8 can effectively transport para-aminohippurate and estrone sulfate, respectively, as well as a wide variety of drugs, toxins, and metabolites. However, substrates for USTS, as their name implies, are not as well characterized. Nevertheless, the presence of these genes, closely related to OAT1 and OAT3, in human and mouse, suggests a role in transport of small organic solutes similar to that observed for the other SLC22 family members. For instance, SLC22A9 in human and the distantly related mouse gene Slc22a19, both of which group phylogenetically with the USTS, can mediate the absorption of organic anions (20, 28).

In this report, we describe a cluster of five novel USTS on mouse chromosome 19 adjacent to Slc22a19, Slc22a6/Oat1 and Slc22a8/Oat3 (also a part of the cluster); together, this eight-gene cluster appears to be largest in the SLC22 family. Their genomic location, sequence similarity and restricted expression patterns suggest they are homologs of human USTS. Sequence analyses identified 12 motifs that characterize mouse Usts. These UST genes possess a signature residue in the possible active site of transmembrane domain 7 that contains a conserved, positively charged, amino acid, Arg356, which could represent a region for interaction with organic anion substrates. The presence of six Ust paralogs in mouse, and three in humans, indicates unusual evolutionary events in the generation of USTS compared with other genes in the SLC22 family.

MATERIALS AND METHODS

Expression analysis and real-time RT-PCR. The tissue distribution of mouse Slc22 family members was determined by real-time RT-PCR. A panel of eight adult mouse tissues, four whole embryonic tissues, and one control cDNA (source RNAs were pooled from multiple individuals in each case) was purchased from Clontech (Mountain View, CA). Dilutions were made from these cDNAs such that 1 ng of template was used in each PCR reaction in a total volume of 10 μl. Primers used for amplification of mouse Ust genes were
designed using Primer3 (http://www-genome.wi.mit.edu). For
time RT-PCR, PCR products were in the 90 to 200 bp range and spanned
at least one intron-exon junction. PCR was carried out using an ABI 7500
Real-Time PCR system. Primer sequences were: For BC014805, Left
primer: 5’gagccagagctggtcg3’; Right primer: ccctcagatccaca. For
AB056442, Left primer: 5’gagccagagctggtcg3’; Right primer: attcagtgcctgggattcac. For D630002G06Rik,
Left primer: 5’ccattagttgcagcatgt3’; Right primer: aagaaggagggagcaaaagc. For Slc22a8, Left primer: 5’
gtcggttcaccaggtgc3’; Right primer: aaccaggccagagagagaca. PCR pa-
did not match the expected size and display the expected mobility.

Alignment and computational analysis. The list of mouse and
human SLC22 family members was obtained by searching the
HUGO database. Sequences of SLC22 family genes were obtained from
the UCSC Genome Browser (http://www.genome.ucsc.edu/
index.html, and the EMBL databases). Nucleotide and amino acid
sequences were aligned using ClustalX (http://www.ebi.ac.uk/
clustalw/). Phylogenograms were generated from the ClustalX output using
TreeView (http://www.ebi.ac.uk/clustalw/) and UCSD Workbench
(http://workbench.sdsc.edu/) with default parameters. The transmem-
brane regions of the novel Ust family proteins were predicted using
TMpred (http://www.ch.embnet.org/software/TMPRED_form.html)
and displayed with TOPO2 (http://www.sacs.ucsf.edu/TOPO-run/
wttopo.pl). Gene names in capitals denote human genes, while lower
case is used to indicate nonhuman genes.

RESULTS

Five clustered mouse genes on chromosome 19 are ho-
omologs of human Usts. A search of the latest mouse genome
sequence (mm9, UCSC Genome Browser, July 2007, Build
37.1) for orthologs of human Usts identified a cluster of 6
genes encoding full-length transporters on mouse chromosome
19q; from proximal to distal: Slc22a19, BC014805, AB056442,
EG43674, D630002G06Rik, and C730048C13Rik (Fig. 1). In
addition to the known mouse Ust Slc22a19, gene prediction
analyses suggested that each of these novel genes encodes a
product of ~550 amino acids, homologous to human UST6.
To deduce the relationship of the five novel mouse proteins
to other SLC22 proteins, a phylogenetic tree of the amino
acid sequences of these five proteins with various human
and mouse Slc22 proteins was constructed (Fig. 2). True
orthologous pairs should appear as pairs in a phylogram as
observed for the OATs and OCTs in Fig. 2. However, human
SLC22A9, SLC22A10, and UST6 and mouse Slc22a19,
BC014805, AB056442, EG43674, D630002G06Rik, and
C730048C13Rik instead clustered separately, indicating that the mouse
Usts are not orthologous to the human Usts (Fig. 2).

OA1T and OA3T are also a part of this cluster, but human
OA1T appears to be a true ortholog of mouse Oat1, as does
human OA3T of mouse Oat3.

Similarly, comparing the amino acid sequences of mouse
genes in the Ust-Oat cluster region with those in human,
**Expression analysis of five novel mouse Usts.** To profile the gene expression patterns of the novel mouse Ust genes, RT-PCR analyses were carried out in a panel of mouse tissues (Fig. 3). Although different patterns of expression were found for each gene, all had highest or second highest expression in the kidney. Substantial liver expression was also found for four of the five genes, and one (D630002G06Rik) had higher expression in testis (as well as kidney and liver). Low but significant levels were found in other tissues such as brain, heart, and lung. Also noteworthy was substantial embryonic expression for D630002G06Rik. Expression of Slc22a8/Oat3 was also analyzed with the same set of samples; as expected, dominant expression was found in the kidney, with lower expression in the brain and heart, consistent with data from previous reports (2, 24, 28).

**Analysis of alternative splicing variants of AB056442.** To further characterize the expression of the novel mouse Ust AB056442, PCR primers were designed to span the entire coding region of the gene, and PCR products were generated from mouse liver cDNA. Four PCR fragments were found of sizes, 1.3, 1.6, 1.8, and 2 kb, with the 1.8 kb fragment being the most abundant species (Fig. 4). These PCR fragments were sequenced, confirming that the 2 kb transcript encoded a product of 551 amino acids and 12 transmembrane domains. RT-PCR with primers flanking the differentially spliced sites revealed tissue-specific expression of splice variants with heart and testis expressing the 1.8 kb transcript and lung expressing the 1.6 kb splicing variant, while kidney and liver, where expression was most abundant, expressed multiple transcripts (Fig. 4B). A changing pattern of splicing variants during development was also observed with the 1.8 kb transcript levels being highest.

**Analysis of sequence conservation in USTs.** We aligned the amino acid sequences of the six mouse Usts and the three human USTs (SLC22A9, SLC22A10, and UST6). Twenty-nine percent of the amino acids were identical in all nine Usts (Fig. 5). Charged amino acids were preferentially conserved; of the 81 charged residues 53% were conserved (55% of the acidic residues and 49% of the basic residues). Charged residues have been postulated to be important in modulating the trans-membrane movement of small charged substrates (16). We used the three-dimensional model of Perry et al. (16) to investigate the location of the conserved charged residues on AB056422, which we use as a reference molecule for the Ust subfamily. Out of the 44 total conserved charged amino acids, 43 were in extracellular or intracellular domains. Many of these were adjacent to transmembrane segments, in regions potentially critical for substrate recognition and movement (Fig. 6). Of 10 charged residues in the predicted 12 transmembrane segments, only one was conserved in all 9 Usts, Arg356,

---

**Table 1. Sequence identity comparisons of selected SLC22A family genes between mouse and human**

<table>
<thead>
<tr>
<th>Identity (Amino Acid)</th>
<th>Slc22a19</th>
<th>Bco14805</th>
<th>AB056442</th>
<th>EG43674</th>
<th>D630002G06Rik</th>
<th>C730048C13Rik</th>
<th>Slc22a8</th>
<th>Slc22a6</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLC22A9</td>
<td>55</td>
<td>52</td>
<td>57</td>
<td>53</td>
<td>53</td>
<td>53</td>
<td>37</td>
<td>35</td>
</tr>
<tr>
<td>SLC22A10</td>
<td>54</td>
<td>52</td>
<td>56</td>
<td>54</td>
<td>54</td>
<td>54</td>
<td>39</td>
<td>36</td>
</tr>
<tr>
<td>UST6</td>
<td>56</td>
<td>54</td>
<td>59</td>
<td>55</td>
<td>55</td>
<td>55</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>SLC22A8</td>
<td>36</td>
<td>32</td>
<td>35</td>
<td>33</td>
<td>34</td>
<td>33</td>
<td>78</td>
<td>49</td>
</tr>
<tr>
<td>SLC22A6</td>
<td>37</td>
<td>37</td>
<td>38</td>
<td>37</td>
<td>36</td>
<td>38</td>
<td>48</td>
<td>83</td>
</tr>
</tbody>
</table>

Values are shown as percentages.
Fig. 3. Expression profiling of a cluster of 5 novel mouse SLC22 family genes. The relative expression of 5 mouse SLC22 family genes was determined in 8 adult tissues and at multiple stages of embryonic development (E, embryonic day). The expression pattern of the well-characterized mouse gene Slc22a8 was determined as a control.

Fig. 4. A: diagram depicting the genomic organization of four splice variants of the novel SLC22 family member, AB056442. The 4 splice variants were cloned and sequenced, and the genomic structure of each variant was determined. B: expression analysis of AB056442 by RT-PCR. PCR primers designed encompassing the spliced sites revealed multiple bands corresponding to the splice variants of the gene. The 1 kb band represents the full-length transcript (see A), while the 0.8 kb band represents the 1.8 kb transcript and the 0.4 kb band represents the 1.3 kb transcript. The 1.8 kb fragment is the most abundantly expressed transcript variant. e7–e17, Age in embryonic days.
located on transmembrane segment 7. However, Arg356 was not found in any of the other Slc22 family genes in mouse or human (Fig. 5 and Supplemental Fig. S11). According to the aforementioned model (16), transmembrane segment 7 is a critical segment forming part of the active site of the "pore" for substrate recognition and translocation. A positive charge at Arg356 might be expected to repel cationic substrates and attract anionic substrates. The conservation of Arg356 in this critical region suggests that Usts may have a preference for compounds that contain a negatively charged group. Although the endogenous substrate remains unknown for the novel Usts, two of the Usts, SLC22A9/OAT7 and Slc22a19/Slc22a9/Oat5, are known to transport at least one organic anion (20, 28). In addition, phylogenetic analysis (Fig. 2) placed human USTs (SLC22A9 and SLC22A10) in the proximity of the known organic anion transporter SLC22A12, also known as URAT1, which mediates the reabsorption of uric acid in the proximal tubule. Thus, it may be that the novel mouse Usts can also transport organic anionic drugs, metabolites and/or toxins.

Motif and structure analyses of mouse Usts. To further characterize the sequences of these novel mouse Usts, we used MEME to identify 12 motifs from a training set of three mouse Usts and then used MAST to map these motifs onto mouse and human USTs, as well as the other SLC22 genes (Supplemental Fig. S2). The MAST scores in Supplemental Fig. S2 are in agreement with the phylogenetic analysis in Fig. 2. For example, the MAST scores for SLC22A10 (1.9 H1002 10/11003 10) and SLC22A11 (1.8 H1002 73/H1002 10) place them closest to the USTs, in agreement with the phylogenetic analysis (Fig. 2).

Evolutionary history of the Usts. In an attempt to trace the evolutionary pathways that gave rise to the Ust clusters in mouse and human, we initially aligned sequences of well known human and mouse Slc22 family genes including the five novel mouse genes (Fig. 2). Given that the Usts and Oat1 and Oat3 are located next to each other (Fig. 1), the Ust and Oat1/3 ancestor genes appear to have arisen from a tandem duplication event. The duplication process then appears to have been

\[[1] The online version of this article contains supplemental material.\]
Table 2. MAST score table

<table>
<thead>
<tr>
<th>Gene Motif</th>
<th>XI</th>
<th>IV</th>
<th>II</th>
<th>I</th>
<th>VII</th>
<th>VI</th>
<th>VIII</th>
<th>V</th>
<th>IX</th>
<th>X</th>
<th>III</th>
<th>XII</th>
</tr>
</thead>
<tbody>
<tr>
<td>D630002G06Rik</td>
<td>-25</td>
<td>-27</td>
<td>-27</td>
<td>-26</td>
<td>-26</td>
<td>-25</td>
<td>-26</td>
<td>-24</td>
<td>-25</td>
<td>-26</td>
<td>-17</td>
<td>-17</td>
</tr>
<tr>
<td>AB056442</td>
<td>-25</td>
<td>-27</td>
<td>-26</td>
<td>-26</td>
<td>-25</td>
<td>-26</td>
<td>-26</td>
<td>-24</td>
<td>-26</td>
<td>-26</td>
<td>-17</td>
<td>-16</td>
</tr>
<tr>
<td>Slc22a19</td>
<td>-17</td>
<td>-25</td>
<td>-24</td>
<td>-21</td>
<td>-20</td>
<td>-20</td>
<td>-21</td>
<td>-16</td>
<td>-14</td>
<td>-16</td>
<td>-19</td>
<td>-18</td>
</tr>
<tr>
<td>SLC22A9</td>
<td>-15</td>
<td>-22</td>
<td>-21</td>
<td>-18</td>
<td>-20</td>
<td>-16</td>
<td>-16</td>
<td>-17</td>
<td>-8</td>
<td>-14</td>
<td>-12</td>
<td>-7</td>
</tr>
<tr>
<td>SLC22A10</td>
<td>-17</td>
<td>-20</td>
<td>-22</td>
<td>-17</td>
<td>-10</td>
<td>-17</td>
<td>-11</td>
<td>-12</td>
<td>-10</td>
<td>-12</td>
<td>-12</td>
<td>-12</td>
</tr>
<tr>
<td>SLC22A11</td>
<td>-11</td>
<td>-18</td>
<td>-18</td>
<td>-18</td>
<td>-10</td>
<td>-12</td>
<td>-14</td>
<td>-12</td>
<td>-7</td>
<td>-5</td>
<td>-9</td>
<td>-9</td>
</tr>
<tr>
<td>SLC22A12</td>
<td>-17</td>
<td>-18</td>
<td>-12</td>
<td>-18</td>
<td>-10</td>
<td>-13</td>
<td>-10</td>
<td>-6</td>
<td>-5</td>
<td>-6</td>
<td>-11</td>
<td>-11</td>
</tr>
<tr>
<td>Slc22a12</td>
<td>-15</td>
<td>-19</td>
<td>-8</td>
<td>-13</td>
<td>-12</td>
<td>-9</td>
<td>-10</td>
<td>-8</td>
<td>-6</td>
<td>-7</td>
<td>-11</td>
<td>-11</td>
</tr>
<tr>
<td>SLC22A6</td>
<td>-14</td>
<td>-9</td>
<td>-8</td>
<td>-9</td>
<td>-12</td>
<td>-15</td>
<td>-12</td>
<td>-5</td>
<td>-6</td>
<td>-6</td>
<td>-11</td>
<td>-11</td>
</tr>
<tr>
<td>SLC22a6</td>
<td>-15</td>
<td>-9</td>
<td>-7</td>
<td>-9</td>
<td>-10</td>
<td>-13</td>
<td>-14</td>
<td>-5</td>
<td>-6</td>
<td>-6</td>
<td>-9</td>
<td>-9</td>
</tr>
<tr>
<td>SLC22A7</td>
<td>-11</td>
<td>-6</td>
<td>-6</td>
<td>-8</td>
<td>-13</td>
<td>-5</td>
<td>-7</td>
<td>-5</td>
<td>-6</td>
<td>-9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slc22a7</td>
<td>-10</td>
<td>-5</td>
<td>-6</td>
<td>-9</td>
<td>-14</td>
<td>-6</td>
<td>-5</td>
<td>-6</td>
<td>-10</td>
<td>-10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Motif</th>
<th>Width</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>20</td>
<td>FSNMTEPDTEPCVDGWYYDR</td>
</tr>
<tr>
<td>II</td>
<td>20</td>
<td>PLDSNLRDLKCRRFAPQWH</td>
</tr>
<tr>
<td>III</td>
<td>20</td>
<td>PETKNQPLPSHDVNDW</td>
</tr>
<tr>
<td>IV</td>
<td>20</td>
<td>ENFTAAPNRCHWVFILDND</td>
</tr>
<tr>
<td>V</td>
<td>20</td>
<td>NPKQGKLKERVAYHMNGMK</td>
</tr>
<tr>
<td>VI</td>
<td>19</td>
<td>QFAITECTVAFAPSFIYC</td>
</tr>
<tr>
<td>VII</td>
<td>20</td>
<td>TIVTEDWVCESQALNSVAK</td>
</tr>
<tr>
<td>VIII</td>
<td>20</td>
<td>AGLAFLFLRWHHLQLAMSVP</td>
</tr>
<tr>
<td>IX</td>
<td>20</td>
<td>MDKELEAATKTPSLRLFHT</td>
</tr>
<tr>
<td>X</td>
<td>19</td>
<td>PLFMMLATYYANLPWFYFG</td>
</tr>
<tr>
<td>XI</td>
<td>20</td>
<td>MAFQELLNQVQSHRFQILQ</td>
</tr>
<tr>
<td>XII</td>
<td>15</td>
<td>SROQKEEPFPHKVTR</td>
</tr>
</tbody>
</table>

P value in e, relative to respective motif.
repeated as the Oat1/3 ancestral gene gave rise to Oat1 and Oat3, and the Ust ancestral gene gave rise to SLC22A9, SLC22A10, and Ust6 in human, and Slc22a19, BC014805, AB056442, EG43674, D630002G06Rik, and C730048C13Rik in mouse (Figs. 2 and 7).

The human Ust cluster has three full-length genes, while the mouse Ust cluster has six (the five novel genes plus Slc22a19). To delineate the evolutionary pathways leading to the generation of these additional genes in mouse, we compared the amino acid sequences of the Ust/Oat cluster of genes (Table 3). When the six mouse Ust genes were compared, four of them, AB056442, EG43674, D630002G06Rik, and C730048C13Rik, shared high amino acid identity with each other with a minimum of 81%; strikingly, AB056442 and D630002G06Rik were 95% identical, and EG43674 and C730048C13Rik were 97% identical. These results suggested that these four genes were generated from recent duplication events that occurred after divergence of the mouse and human lineages. Taken together, the data suggest that genes in this Ust/Oat cluster were generated through successive rounds of segmental duplication, with AB056442, EG43674, D630002G06Rik, and C730048C13Rik generated in the most recent rounds. Slc22a19 and BC014805 were generated and survived from the previous round of duplication, and the Ust and Oat1/3 ancestral genes were generated in the earliest round of duplication (Fig. 7).

Given their lack of orthologs in the human and highly conserved sequences, the duplication events that gave rise to AB056442, EG43674, D630002G06Rik, and C730048C13Rik genes were likely mouse-specific (Fig. 7 and Supplemental Fig. S3). Alternatively, the orthologs for the mouse Usts might not have survived purifying selection in human, and vice versa.

To determine if the species-specific duplications were unique to mouse, we examined the Ust/Oat clusters in several vertebrate species. Such clusters were found in Homo sapiens (Build 36.3), Mus musculus (Build 37.1, mouse), Rattus norvegicus (RGSC v3.4, rat), Ornithorhynchus anatinus (Build 1.1, duck-billed platypus), Monodelphis domestica (MonDom5, opossum), Bos taurus (Btau_4.0, cow), Canis lupus familiaris (Build 2.1, dog), Equus caballus (EquCab2.0, horse), and Sus scrofa (Susrof5, pig). By homology (BLAST) search, Oat orthologs were also found in Danio rerio (Zv7, zebrafish). Phylogenetic analysis of gene sequences derived from these clusters revealed species-specific duplications in several genomes (Supplemental Fig. S3). Orthologs for Oat1 and Oat3 were found in most of the genomes examined, which may suggest conserved function in these species. On the other hand, species-specific amplifications of Ust genes were common. In rats, similar to the observed mouse-specific amplifications of the Usts, Ust5r and Ust4r appeared to be rat-specific. In horse, gene amplifications of LOC100064462 and LOC100064545, and of LOC100064402 and LOC10006420 were horse-specific. Similarly, in cows, amplification of Slc22a10 and LOC532368 was a species-specific event. In humans, SLC22A9 and SLC22A25 appear to be species-specific.

Notably, while both opossum and platypus have orthologs for OAT1 (LOC100013196 in opossum and LOC100088844 in platypus) and an OAT3 ortholog can also be found in platypus (LOC100093352), several additional genes in platypus (LOC100087269, LOC100078871, and LOC100093355) and opossum (LOC100030315, LOC100033978, LOC100030335, LOC100030283, LOC100013455, and LOC100033502) were found that were more similar to Oat1 and Oat3 than to the Usts. These results suggest that a strategy of preferential amplification of Oat1/3-like genes, rather than of Usts, was employed during the evolution of opossum and platypus.

**DISCUSSION**

Based on comparative analyses of the syntenic regions of the mouse and human UST-OAT clusters, we identified a group of five genes belonging to the Slc22 family of solute transporters on mouse chromosome 19. Along with the previously identified Slc22a19, these genes form a group of six paralogs, which likely arose by species-specific serial duplication of an ancestral Ust.

**Table 3. Sequence identity comparisons between novel mouse Slc22 family members**

<table>
<thead>
<tr>
<th>Identity</th>
<th>Amino Acid</th>
<th>Slc22a19</th>
<th>BC014805</th>
<th>AB056442</th>
<th>EG43674</th>
<th>D630002G06Rik</th>
<th>C730048C13Rik</th>
<th>Slc22a8</th>
<th>Slc22a6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slc22a19</td>
<td></td>
<td>100</td>
<td>64</td>
<td>76</td>
<td>62</td>
<td>63</td>
<td>63</td>
<td>34</td>
<td>36</td>
</tr>
<tr>
<td>BC014805</td>
<td></td>
<td>100</td>
<td>71</td>
<td>60</td>
<td>60</td>
<td>61</td>
<td>33</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>AB056442</td>
<td></td>
<td>100</td>
<td>81</td>
<td>95</td>
<td>82</td>
<td>33</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EG43674</td>
<td></td>
<td>100</td>
<td>84</td>
<td>97</td>
<td>34</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D630002G06Rik</td>
<td></td>
<td>100</td>
<td>100</td>
<td>86</td>
<td>35</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C730048C13Rik</td>
<td></td>
<td>100</td>
<td>100</td>
<td>34</td>
<td>37</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slc22a8</td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slc22a6</td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sequence identity comparison values shown are percentages.
Similar species-specific Ust clusters were identified in other mammalian species. In certain species (e.g., opossum and platypus), it was the Oat1/Oat3 cluster that was amplified. These species-specific clusters seem to be relatively unique in the SLC22 family and may reflect the adaptation of each species to its particular environment.

Identifying a cluster of mouse UST genes without clear orthologs among human USTs helps to clarify a confusing area of nomenclature for the SLC22 family. For example, mouse Slc22a19 was “slashed” with Slc22a9 to indicate it as the ortholog of human SLC22A9, and both can transport organic anions.

**Phylogeny of the Oats and the Usts.** Based on the amino acid sequence similarities of the Usts (>50% for the most divergent of these genes), the Oats (<50% between Oat1 and Oat3), and of mouse and human orthologs (~80% for both mouse Oat1 vs. human OAT1 and for mouse Oat3 vs. human OAT3) (see Table 1) the ancestral genes of the former were generated at a time between the divergence of Oat1 and Oat3 and the divergence of the mouse and human lineages. Alternatively, if the mouse and human Usts shared their most recent common ancestor at the time of the divergence of the mouse and human lineages, reduced similarity compared with that observed for other orthologs might suggest accelerated genetic change.

The apparent recent duplication events involving the Ust-Oat cluster of genes suggest active genomic adaptation among these genes. Indeed, such species-specific amplification can be observed in the duplication of SLC22A9 and SLC22A25 pair. Recent genomic studies using whole genome hybridization methods have also identified copy number variants (CNV) in the region that includes the SLC22A10 locus (cnp923) (18), presumably due to segmental duplication/deletion events. Furthermore, finding multiple copies of amplified Slc22a10-like genes (LOC100064402 and LOC100064420) in horse raises the possibility that the gene duplication events resulting in the current UST cluster on human chromosome 11 and the CNV observed in this region may share the same etiology.

**Different expression patterns of mouse and human Ust genes.** Gene expression analyses in the better characterized species of mouse and human suggested that most SLC22 family members have predominant expression in the major excretory organs, kidney and liver. The five novel Slc22 family members in mouse display expression in both kidney and liver. This is somewhat surprising, considering the mouse Ust cluster is syntenic to the human UST cluster, and the genes in the latter are expressed in liver only (6, 21). One can speculate that, if there is a locus of control that dictates the shared expression of the genes in the cluster, then kidney expression of the Usts might have been gained in mouse along with formation of new Ust paralogs by gene duplication, or, alternatively, kidney expression might have been lost in the human USTS. Another possible explanation of this difference in gene expression between mouse and human Usts is that each cluster acquired its expression pattern independently as it arose from repeated tandem duplications. Nevertheless, the shared expression of these genes may have both genetic and physiological implications. OAT1 and OAT3 are an example of the clustering of structurally and functionally similar genes: both are expressed in the proximal tubule and both are broad-spectrum drug/metabolite/toxin transporters that have similar mechanisms. It is possible that the renal USTS, which appear to be in proximal tubule cells (28), may complement OATs in the same cell, mediating the transepithelial movement of organic anionic drugs, metabolite, and toxins. Because the physiological role of SLC22 family members may primarily be that of endogenous metabolite handling, the embryonic expression and splicing patterns may be an important area for future work. It has previously been shown that certain Oats and Octs have embryonic expression as both within and outside the kidney (15, 23).

**Preferential conservation of charged amino acids.** Compared with noncharged residues, charged residues were twice as likely to be conserved among the nine mouse and human Usts, consistent with a function of the Usts as transporters of charged small organic molecules. Most of the conserved charged residues (43 of 44) were located in intracellular and extracellular domains, rather than in transmembrane domains. The lone transmembrane charged residue conserved in all nine Usts is Arg356 located in helix 7. Along with helices 5, 8, 10 and 11, helix 7 is a part of the putative active site based on modeling of OATs (16). The positive charge of Arg356 in the active site may be critical for substrate recognition. The other area that may influence substrate preference is the first extracellular loop where several charged residues, both positive and negative, were conserved (Fig. 6). The first extracellular loop is also the region harboring several predicted N-glycosylation sites, and four disulfide bond-forming Cystine residues that were also conserved (Fig. 6). Overall, this loop is the most conserved region of the Usts with 59 of 113 residues conserved in all nine Usts.

Although USTS form a large subclass within SLC22 transporters, unlike other family members, their function is poorly understood. We have identified a cluster of six mouse paralogs in the Ust subfamily of the SLC22 family that are not orthologous to the human USTS. On the other hand, the Ust genes are highly similar to each other with many conserved residues, including a uniquely conserved positively charged amino acid in transmembrane segment 7 that may correspond to a functional site. Evolutionary analyses indicate that the clustering and species-specific amplification of the Usts and Oats may underlie species’ adaptations to their environment. Some of the Ust genes underwent recent duplication, which may represent active selection and may correlate with copy number variations observed in the human SLC22A10 locus.

**ACKNOWLEDGMENTS**

We thank Megan Bettilyon for assistance.

**GRANTS**

This work was supported by National Institute of Health Grants AI-057695 and DK-079784 (to S. K. Nigam).

**REFERENCES**


