Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

J Phylogenetics Evol Biol. Author manuscript; available in PMC 2017 August 31.

Published in final edited form as:

J Phylogenetics Evol Biol. 2016 August ; 4(3): . doi:10.4172/2329-9002.1000167.

Evolution of Vertebrate Solute Carrier Family 9B Genes and Proteins (*SLC9B*): Evidence for a Marsupial Origin for Testis Specific *SLC9B1* from an Ancestral Vertebrate *SLC9B2* Gene

Roger S Holmes[1,*,2], **Kimberly D Spradling-Reeves**[2], and **Laura A Cox**[2]

[1]Eskitis Institute for Drug Discovery and School of Natural Sciences, Griffith University, Nathan, QLD, Australia

[2]Department of Genetics and Southwest National Primate Research Center, Texas Biomedical Research Institute, San Antonio, Texas, USA

Abstract

SLC9B genes and proteins are members of the sodium/lithium hydrogen antiporter family which function as solute exchangers within cellular membranes of mammalian tissues. *SLC9B2* and *SLC9B1* amino acid sequences and structures and *SLC9B*-like gene locations were examined using bioinformatic data from several vertebrate genome projects. Vertebrate *SLC9B2* sequences shared 56-98% identity as compared with ~50% identities with mammalian *SLC9B1* sequences. Sequence alignments, key amino acid residues and conserved predicted transmembrane structures were also studied. Mammalian *SLC9B2* and *SLC9B1* genes usually contained 11 or 12 coding exons with differential tissue expression patterns: *SLC9B2,* broad tissue distribution; and *SLC9B1*, being testis specific. Transcription factor binding sites and CpG islands within the human *SLC9B2* and *SLC9B1* gene promoters were identified. Phylogenetic analyses suggested that *SLC9B1* originated in an ancestral marsupial genome from a *SLC9B2* gene duplication event.

Keywords

Solute carrier; *SLC9B2*; *SLC9B1*; NHA2; NHA1; Sodium/hydrogen exchanger; Testis-specific; Transmembrane; Marsupial origin; Gene evolution

Introduction

SLC9/NHE gene family members (sodium lithium carrier 9/ sodium hydrogen exchanger) contribute to the maintenance of cellular and intracellular pH homeostasis and are predominantly responsible for the absorption of Na^+ ions in the kidney and the GI tract [1-3]. SLC9 comprises three subfamilies: SLC9A, with 9 members (SLC9A1-A9), of which SLC9A1-9A5 are localized in plasma membranes and SLC9A6-SLC9A9 in subcellular organelles; SLC9B, with two members, *SLC9B1* and *SLC9B2* [2], which are the subject of

*Corresponding author: Roger S Holmes, Eskitis Institute for Drug Discovery and School of Natural Sciences, Griffith University, Nathan QLD 4111, Australia, Tel: +61 7 37356008; r.holmes@griffth.edu.au.

Disclosure: No conflicts of interest are reported.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

this study; and SLC9C, with two members localized in sperm, SLC9C1 and SLC9C2 [2,4,5].

A mitochondrial inner membrane Na^+/H^+ exchanger (NHA2 or NHEDC2; designated as *SLC9B2*), was originally identified in yeast (*Saccharomyces cerevisiae*) and bacteria (*Escherichia coli*), with the human homologue showing ubiquitous expression, particularly in tissues with high mitochondrial content, including the kidney distal convoluted tubule [6,7]. *SLC9B2* has been identified as a contributor to sodium-lithium counter-transport activity (SLC), and may serve as a candidate gene for essential hypertension in human populations, particularly in urban black populations [7,8]. Moreover, genetic linkage studies of SLC activity in baboons have reported association with chromosome 5, the homologue to human chromosome 4 [9]. Human *SLC9B2* has also been shown to reverse the Na^+/H^+ exchanger-null phenotype when expressed in Na^+/H^+ exchanger deficient yeast [10], which supports a major role for *SLC9B2* as a sodium/hydrogen exchanger. A second gene (NHA1 or NHEDC1; designated as SLC9B1) has been located in tandem with *SLC9B2* on human chromosome 4 which is specifically expressed in testis in mammals [7,11]. This is in contrast to Drosophila, for which two SLC9B-like genes (designated as NHA1 and NHA2) are widely expressed in epithelial tissues and play crucial roles in organismal ion homeostasis and as Na^+/H^+ exchangers [12].

This paper reports the predicted gene structures and amino acid sequences for several vertebrate SLCB2 and mammalian SLC9B1 genes and proteins, the predicted structures for mammalian SLCB1 and SLCB2 proteins and the structural, phylogenetic and evolutionary relationships for these genes and enzymes. The results suggest that the mammalian *SLC9B1* gene arose from the duplication event of an ancestral mammalian *SLC9B2* gene with the appearance corresponding to the emergence of marsupial mammals during vertebrate evolution.

Methods

SLC9B gene and protein identification (*SLC9B1* and *SLC9B2*)

BLAST studies were undertaken using web tools from the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov) [13]. Protein BLAST analyses used vertebrate *SLC9B2* and mammalian SLC9B1 amino acid sequences previously described [2,7,10,11] (Table 1; Tables S1 and S2). Predicted SLC9B-like protein sequences were obtained in each case and subjected to protein and gene structure analyses.

BLAT analyses were undertaken for each of the predicted SLC9B1 and SLC9B2 amino acid sequences using the UC Santa Cruz Genome Browser (http://genome.ucsc.edu) with the default settings to obtain the predicted locations for each of the vertebrate *SLC9B*-like genes, including exon boundary locations and gene sizes [14]. The structures for the major human *SLC9B2* and *SLC9B1* transcripts were obtained using the AceView website (http://www.ncbi.nlm.nih.gov/ieb/ research/acembly/) [15].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Alignments, structures and predicted properties of mammalian *SLC9B2* and *SLC9B1* proteins

Amino acid sequence alignments of vertebrate *SLC9B2* and *SLC9B1* proteins were undertaken using Clustal Omega [16]. Predicted secondary and tertiary structures for human and other mammalian *SLC9B2* and *SLC9B1* proteins were obtained using the SWISS-MODEL web-server [17] and the reported tertiary structures for bacterial (Thermus thermophilus) Na/H antiporter [18] (PDB:4bwzA) with modeling residue ranges of 114-512 for human *SLC9B2* and 94-502 for human *SLC9B1* (Figure S1). Predicted transmembrane structures for human *SLC9B2* and *SLC9B1* proteins were obtained using web tools (http://au.expasy.org/tools/pi_tool.html). Identification of conserved domains for vertebrate *SLC9B2* and *SLC9B1* proteins was made using NCBI web tools [19] (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

Comparative human *SLC9B2* and *SLC9B1* gene expression

RNA-seq gene expression profiles across 16 selected tissues from the Non-Human Primates References Transcriptome were used. (AceView http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/).

Phylogeny studies and sequence divergence

Phylogenetic analyses were undertaken using the http://phylogeny. fr platform [20]. Alignments of vertebrate *SLC9B2* and mammalian *SLC9B1* sequences were assembled using PMUSCLE [21] (Table 1 and Supplementary Table S1). Alignment ambiguous regions were excluded prior to phylogenetic analysis yielding alignments for comparisons of vertebrate *SLC9B2* and mammalian *SLC9B1* sequences. The phylogenetic tree was constructed using the maximum likelihood tree estimation program PHYML [22].

Results and Discussion

Alignments and possible roles of *SLC9B2* amino acid sequences

The deduced amino acid sequences for baboon (*Papio anubis*), chicken (*Gallus gallus*), zebra fish (*Danio rerio*) and coelacanth (*Latimeria chalumnae*) *SLC9B2* proteins are shown in Figure 1 together with the previously reported sequence for human and mouse *SLC9B2* proteins [2,7] (Table 1). Alignments of human with other vertebrate sequences examined were between 52-98% identical, suggesting that these are likely to be members of the same family of genes, whereas comparisons of sequence identities with vertebrate *SLC9B1*proteins exhibited lower levels of sequence identities, suggesting that these are members of a distinct but related gene family (Table 1; Tables S1 and Table S2).

The amino acid sequences for vertebrate *SLC9B2* proteins contained between 537 (for human *SLC9B2* isoform 1) and 564 (for zebra fish *SLC9B2*) residues (Figure 1). Previous studies have reported several key regions and residues for structurally related SLC9 Na^+/H^+ exchange proteins. These included an N-terminus region (residues 1-86), predicted to be external to the inner-mitochondrial membrane; hydrophobic transmembrane segments which may anchor the enzyme to the inner-mitochondrial membrane; and a C-terminal region (residues 513-537), predicted to be localized within the mitochondrial matrix. The N-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

terminal and C-terminal regions showed lower levels of amino acid sequence conservation which are in contrast to 13 identified hydrophobic regions (TM1-TM13), which are more highly conserved (Figure 1; Table 2). These regions were identified as transmembrane segments, also designated as 9B2:1/TM1-9B2:13/TM13 hydrophobic zones (Figure 2; Table 2). Several intra-hydrophobic sequences were also highly conserved, including the N-terminus region prior to 9B2:1/ TM1 (residues 77-84 designated pre9B2:1); ITM1 (residues 107-112); ITM2 (residues 136-138); ITM4, containing a basic amino acid cluster (residues 193-208); ITM5 (residues 230-232); and ITM6 containing the active site (residues 255-311), particularly the double aspartate sequence (278-279) [4,23]. In addition, the C-terminus region, immediately following the 9B2:13/TM13 transmembrane sequence (residues 514-519), was highly conserved as well (Table 2). Mutation analyses of a plasmid containing human *SLC9B2* expressed in a transfected salt sensitive strain of yeast cells demonstrated reduced ability to remove intracellular sodium when mutants in Val161 and Phe357 were used [23]. These residues were conserved among the mammalian *SLC9B2* sequences examined although conservative substitutions were observed in some lower vertebrate *SLC9B2* sequences (Figure 1).

The roles for these conserved amino acid sequences located outside of the hydrophobic and transmembrane regions for *SLC9B2* remain to be determined; however, it is proposed that the presence of multiple proline and glycine residues in these sequences may contribute to the sharp turns required for maintaining these transmembrane structures, particularly for the ITM1 and ITM6 sequences, which contained double or triple glycine residues (112-113 and 258-260, respectively), and the pre9B2:1 region, which contained a double proline sequence (79-80) (Figure 1; Table 2). A distinctive intra-hydrophobic region was observed for ITM4, which exhibited high basic amino acid content, namely K195, K198, K199 and K201. This positive ion cluster may assist in the transfer of salt ions across the inner-mitochondrial membrane or form a salt-ion with the double aspartate sequence (278-279DD), previously reported as the active site for *SLC9B2* [3,7].

Predicted structures for vertebrate *SLC9B2* and human *SLC9B1*

Predicted secondary structures for vertebrate *SLC9B2* and human *SLC9B1* sequences were examined, which were consistent with the presence of 13 hydrophobic regions, including at least 10 transmembrane structures, symmetrically placed on either side of the double aspartate (278-279DD) active site sequence (Figures 1-3). Predicted tertiary structures for human *SLC9B2* and *SLC9B1* were also examined which were based on the reported 3D structure for the sodium proton antiporter protein from Thermus thermophilus (also called NapA) (Figure S2) [18]. These transmembrane structures may be compared to pfam02080 domains [24], for which five transmembrane helices are found on either side of a 10 helical structure, predicted to form a pore involved in sodium proton exchange. The aspartate-alanine antiporter (aspT) from Tetragenococcus halophilus has also been shown to function with similar domain architecture [25].

A similar secondary structure was observed for human *SLC9B1*, although with a shorter N-terminus sequence (Figures 2 and 3) and a notable absence of 3 glycine zipper sequences (GlyxxxGlyxxxGly), observed for human *SLC9B2* (GZ1: 231-239; GZ2: 257-266; GZ3:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

313-321) (Table 2; Figure 3). Glycine zipper sequences have been previously shown to encourage close helix-helix packing within transmembrane structures, and to facilitate the formation of membrane pores associated with channel formation [26,27], which is particularly relevant to the role that *SLC9B2* plays as a mitochondrial inner membrane Na^+/H^+ exchanger. Vertebrate *SLC9B2* sequence alignments also showed that GZ1 and GZ3 glycine zipper sequences were conserved during vertebrate evolution, whereas GZ2 sequences were not fully conserved, at least among the vertebrate sequences examined (Figure 1). It is not known what impacts, if any, these differences in glycine zipper sequences have on *SLC9B2* and *SLC9B1* Na^+/H^+ exchanger functions.

Gene Locations, exonic structures and regulatory sequences for vertebrate *SLC9B2* and mammalian *SLC9B1* genes

Table 1 and Supplementary Tables S1 and S2 summarize the predicted locations for vertebrate *SLC9B2* and mammalian *SLC9B1* genes based upon BLAT interrogations of several vertebrate genomes using the reported sequences for human *SLC9B2* [7,28,29] and *SLC9B1* [7,11] and the predicted sequences for other vertebrate *SLC9B2* and mammalian *SLC9B1* proteins [11,14]. The vertebrate *SLC9B2* and mammalian *SLC9B1* genes examined were transcribed on either strand, depending on the species. Figure 1 summarizes the predicted exonic start sites for human, baboon, mouse, chicken, zebra fish and coelacanth *SLC9B2* genes, based on the ‘a’ isoform, with each having 11 or 12 coding exons, in identical or similar positions to those predicted for the human *SLC9B2* gene.

Figure 4 shows the predicted structures for the major human *SLC9B2* and *SLC9B1* transcripts. In each case, two major transcripts were observed, including the reference *SLC9B2* sequences (AK172823 and BC047447) which were 3601 and 2945 bps in length; and the reference *SLC9B1* sequences (BC136966 and AY461581) which were 1879 and 1755 bps in length. Te two major human *SLC9B2* transcript isoforms, designated as ‘a’ and ‘b’, encode identical proteins with 537 amino acids with 11 coding exons and 1 or 2 non-coding exons, respectively (Table 1; Figure 4). The two human major *SLC9B1* transcripts, also designated as ‘a’ and ‘b’, encoded proteins with 515 and 475 amino acids, with 11 and 10 coding exons respectively and 2 non-coding exons (Table 1; Figure 4). The first human *SLC9B1* intron is much larger in comparison with intron 1 for human *SLC9B2*, which explains the observation that primate *SLC9B1* genes were double the size for the primate *SLC9B2* genes (Table S1; Figure 4).

The human *SLC9B2* promoter region contained several potential transcription factor binding sites, including FOXD3, which promotes neural crest development; FOXJ2, which serves as a transcriptional activator; and HNF3B, which is involved in the development of liver, kidney and notochord (Table S3). It would appear that the *SLC9B2* gene promoter is well endowed with gene regulatory sequences which may contribute to the high levels of *SLC9B2* expression in mammalian neural, liver and kidney cells and to the maintenance of this expression during development. In addition, the human *SLC9B2* promoter region contained a large CpG island (CpG92), which may play an important role in *SLC9B2* gene silencing in association with methylation of the gene promoter as reported for other genes during carcinogenesis [30].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

SLC9B2 and *SLC9B1* tissue expression

Supplementary Figure S2 compares the tissue expression patterns for human and primate *SLC9B2* and *SLC9B1* genes respectively. Of particular interest are the distinct tissue expression profiles observed for these genes, with *SLC9B2* having a wider tissue expression pattern and showing much higher levels of expression in kidney, brain and liver; but with *SLC9B1* showing a testis-specific pattern in both human and rhesus macaque tissues. This is consistent with previous studies [2,3,11,14]. The higher levels and wider distribution patterns of expression for *SLC9B2* have been associated with the Na(+)/H(+) exchanger, sodium-lithium counter transport, cation/proton antiporter and salt homeostasis roles across tissue organelle lipid bilayers, including the inner mitochondrial membranes of kidney, which has led to this being designated as a candidate gene for essential hypertension [7]. No specific role for the unique testis-specific *SLC9B1* expression has been identified [11,31].

Phylogeny and divergence of vertebrate *SLC9B2* and mammalian *SLC9B1* genes and proteins

A phylogenetic tree (Figure 5) was calculated by the progressive alignment of 26 vertebrate and invertebrate *SLC9B2* amino acid sequences with 17 mammalian *SLC9B1* sequences which was ‘rooted’ with the coelacanth (*Latimeria chamumnae) SLC9B2* sequence (Table 1 and Table S1, S2). The phylogram showed clustering of the SLC9B-like sequences into groups consistent with their evolutionary relatedness as well as groups for the vertebrate *SLC9B2* and mammalian *SLC9B1* sequences. These groups were significantly different from each other (with bootstrap values >93). It is apparent from this study of vertebrate SLC9B-like genes and proteins that *SLC9B2* represents the vertebrate ancestral form while *SLC9B1* appeared early in marsupial and eutherian mammalian evolution for which a proposed common ancestor for these genes may have predated or coincided with the appearance of marsupial mammals during vertebrate evolution. In addition, the close localization of the *SLC9B2* and *SLC9B1* genes on human chromosome 4 as well as their similarities in sequence and exonic structure suggested that the mammalian *SLC9B1* gene evolved following a gene duplication event of the *SLC9B2* gene, similar to that reported for many gene families [32-34].

Conclusions

The results of the present study indicated that vertebrate *SLC9B2* and mammalian *SLC9B1* genes and encoded proteins represent a distinct gene and protein family of SLC9B-like proteins which share conserved sequences and active site residues with those reported for SLC9/NHE gene family members (sodium lithium carrier 9/sodium hydrogen exchanger), which contribute to the maintenance of cellular and intracellular pH homeostasis and are predominantly responsible for the absorption of Na^+ ions in the kidney and the GI tract [1-3].

SLC9B1 is encoded by a single gene (*SLC9B1*) among the marsupial and eutherian mammalian genomes studied and are highly expressed in human testis and usually contained 11 or 12 coding exons on the negative or positive strand, depending on the species. *SLC9B2* is also encoded by a single gene (*SLC9B2*) which is more widely expressed in human tissue,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

with highest levels being observed in kidney, liver and brain, being coexpressed with a proximal gene encoding 3-hydroxybutyrate dehydrogenase (BDH2). Several transcription factor binding sites were localized within the human *SLC9B2* gene promoter region, including FOXD3, FOXJ2 and HNF3B, which regulate gene expression, and may contribute significantly to the high level of gene expression in kidney, liver and neural cells. In addition, this region contains a large CpG island which may also contribute to *SLC9B2* gene regulation during development.

Predicted secondary and tertiary structures for vertebrate *SLC9B2* and mammalian *SLC9B1* proteins showed a strong similarity with similar Na/H antiporters originally identified in yeast (*Saccharomyces cerevisiae*) and bacterial (Termus thermophilus) Na/H antiporter proteins [17] (PDB:4bwzA). Several major structural domains were apparent for vertebrate *SLC9B2* and mammalian *SLC9B1* proteins, including a N-terminal tail of varying lengths predicted to be external to the inner-mitochondrial membrane; hydrophobic transmembrane segments which may anchor the enzyme to the inner-mitochondrial membrane; and a C-terminal region (residues 513-537), predicted to be localized within the mitochondrial matrix. The N-terminal and C-terminal regions showed lower levels of amino acid sequence conservation which are in contrast to 13 identified hydrophobic regions (9B2:1-9B2:13), which are more highly conserved.

A phylogenetic study used 17 mammalian *SLC9B1* and 26 *SLC9B2* protein sequences which indicated that the *SLC9B1* gene appeared early in marsupial mammalian evolution, prior to or coincident with the appearance of marsupial mammals, and derived from a gene duplication event of the ancestral vertebrate (and invertebrate)

SLC9B2 gene, which originated much earlier in evolution, since it has been reported in bird, reptile, amphibian, and fish genomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Grant support acknowledgements (to LAC): C06 RR013556, R01 HL118556 and P51 OD011133.

References

1. Orlowski J, Grinstein S. Diversity of the mammalian sodium/proton exchanger SLC9 gene family. Pflugers Arch. 2004; 447:549–565. [PubMed: 12845533]
2. Brett CL, Donowitz M, Rao R. Evolutionary origins of eukaryotic sodium/ proton exchangers. Am J Physiol Cell Physiol. 2005; 288:C223–239. [PubMed: 15643048]
3. Donowitz M, Ming Tse C, Fuster D. SLC9/NHE gene family, a plasma membrane and organellar family of Na^7/H^7 exchangers. Mol Aspects Med. 2013; 34:236–251. [PubMed: 23506868]
4. Wang D, King SM, Quill TA, Doolittle LK, Garbers DL. A new sperm-specific Na+/H+ exchanger required for sperm motility and fertility. Nat Cell Biol. 2003; 5:1117–1122. [PubMed: 14634667]
5. Liu T, Huang JC, Zuo WL, Lu CL, Chen M, et al. A novel testis-specific Na+/H+ exchanger is involved in sperm motility and fertility. Front Biosci. 2010; 2:566–581.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

6. Numata M, Petrecca K, Lake N, Orlowski J. The identification of a mitochondrial Na+/H+ exchanger. J Biol Chem. 1998; 2723:6951–6959.
7. Xiang M, Feng M, Muend S, Rao R. A human Na+/H+ antiporter sharing evolutionary origins with bacterial NhaA may be a candidate gene for essential hypertension. Proc Natl Acad Sci U S A. 2007; 104:18677–18681. [PubMed: 18000046]
8. Lindhorst J, Alexander N, Blignaut J, Rayner B. Differences in hypertension between blacks and whites: an overview. Cardiovasc J Afr. 2007; 18:241–247. [PubMed: 17940670]
9. Kammerer CM, Cox LA, Mahaney MC, Rogers J, Shade RE. Sodium-lithium countertransport activity is linked to chromosome 5 in baboons. Hypertension. 2001; 37:398–402. [PubMed: 11230307]
10. Fuster DG, Alexander RT. Traditional and emerging roles for the SLC9 Na+/H+ exchangers. Pfugers Arch. 2014; 466:61–76.
11. Ye G, Chen C, Han D, Xiong X, Kong Y, et al. Cloning of a novel human NHEDC1 (Na+/H+ exchanger like domain containing gene expressed specifically in testis. Mol Biol Rep. 2006; 33:175–180. [PubMed: 16850186]
12. Chintapalli VR, Kato A, Henderson L, Hirata T, Woods DJ, et al. Transport proteins NHA1 and NHA2 are essential for survival, but have distinct transport modalities. Proc Natl Acad Sci USA. 2015; 112:11720–11725. [PubMed: 26324901]
13. Altschul F, Vyas V, Cornfeld A, Goodin S, Ravikumar TS, et al. Basic local alignment search tool. J Mol Biol. 1990; 215:403–410. [PubMed: 2231712]
14. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, et al. The human genome browser at UCSC. Genome Res. 2002; 12:994–1006.
15. Thierry-Mieg D, Thierry-Mieg J. AceView: A comprehensive cDNA-supported gene and transcripts annotation. Genome Biol. 2006; 7:1–14.
16. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, et al. Fast, scalable generation of high quality protein multiple sequence alignments using Clustal Omega. Mol Systems Biol. 2011; 7:539.
17. Schwede T, Kopp J, Guex N, Pietsch MC. SWISS-MODEL: An automated protein homology-modelling server. Nucleic Acids Res. 2003; 31:3381–3385. [PubMed: 12824332]
18. Lee C, Kang HJ, von Ballmoos C, Newstead S, Uzdavinys P, et al. A two-domain elevator mechanism for sodium/proton antiport. Nature. 2013; 501:573–577. [PubMed: 23995679]
19. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, et al. CDD: a conserved domain database for the functional annotation of proteins. Nucleic Acids Res. 2011; 39:D225–D229. [PubMed: 21109532]
20. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, et al. Phylogeny. fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 2008; 36:W465–W469. [PubMed: 18424797]
21. Edgar RC. MUSCLE: a multiple sequence alignments method with a reduced time and space complexity. BMC Bioinformatics. 2004; 5:113. [PubMed: 15318951]
22. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003; 52:696–704. [PubMed: 14530136]
23. Huang X, Morse LR, Xu Y, Zahradka J, Schrova H, et al. Mutational analysis of NHAoc/NHA2 in Saccharomyces cerevisiae. Biochim Biophys Acta. 2010; 1800:1241–1247. [PubMed: 20713131]
24. Finn RD, Miller BL, Clements J, Bateman A. iPfam: a database of protein family and domain interactions found in the Protein Data Bank. Nucleic Acids Res. 2014; 42:D364–D373. [PubMed: 24297255]
25. Nanatani K, Ohonishi F, Yoneyama H, Nakajima T, Abe K. Membrane topology of the elctrogenic aspartate-alanine antiporter AspT of Tetragenoccus halophilus. Biochem Biophys Res Commun. 2005; 328:20–26. [PubMed: 15670744]
26. Kim S, Jeon TJ, Oberai A, Yanf D, Schmidt JJ, et al. Transmembrane glycine zippers: physiological and pathological roles in membrane proteins. Proc Natl Acad Sci USA. 2005; 102:14278–14283. [PubMed: 16179394]
27. Dong H, Sharma M, Zhou HX, Cross TA. Glycines: roleα in -helical membrane protein structures and a potential indicator of native conformation. Biochemistry. 2012; 51:4779–4789. [PubMed: 22650985]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

28. Yun CH, Tse CM, Nath SK, Levine SA, Brant SR. Mammalian Na+/H+ exchanger gene family: structure and function studies. Am J Physiol. 1995; 269:G1–G11. [PubMed: 7631785]
29. Malakooti J, Dahdal RY, Schmidt L, Layden TJ, Dudeja PK, et al. Molecular cloning, tissue distribution, and functional expression of the human Na(+)/H(+) exchanger NHE2. Am J Physiol. 1999; 277:G383–G390. [PubMed: 10444453]
30. Kang XC, Chen ML, Yang F, Gao BQ, Yang QH, et al. Promoter methylation and expression of SOCS-1 affect clinical outcome and epithelial-mesenchymal transition in colorectal cancer. Biomed Pharmacother. 2016; 80:23–29. [PubMed: 27133036]
31. Tusnady GE, Simon I. The HMMTOP transmembrane topology prediction server. Bioinformatics. 2001; 17:849–850. [PubMed: 11590105]
32. Holmes RS. Evolution of mammalian KELL blood group glycoproteins and genes (KEL): evidence for a marsupial origin from an ancestral M13 Type II endopeptidase gene. Phyl Evol Biol. 2013; 1:112.
33. Storz JF. Gene duplication and evolutionary innovations in hemoglobin-oxygen transport. Physiology. 2016; 31:223–232. [PubMed: 27053736]
34. Vervoort M, Muelemeester D, Behague J, Kerner P. Evolution of Prdm genes in animals: insights from comparative genomics. Mol Biol Evol. 2016; 33:679–696. [PubMed: 26560352]

Abbreviations

SLC	Sodium Lithium Countertransport Activity
NHE	Sodium Hydrogen Exchanger
BLAST	Basic Local Alignment Search Tool
BLAT	Blast-Like Alignment Tool
NCBI	National Center for Biotechnology Information
KO	Knock Out
GI	Gastro-Intestinal
AceView	NCBI Based Representation of Public mRNAs
SWISS-MODEL	Automated Protein Structure Homology-modeling Server

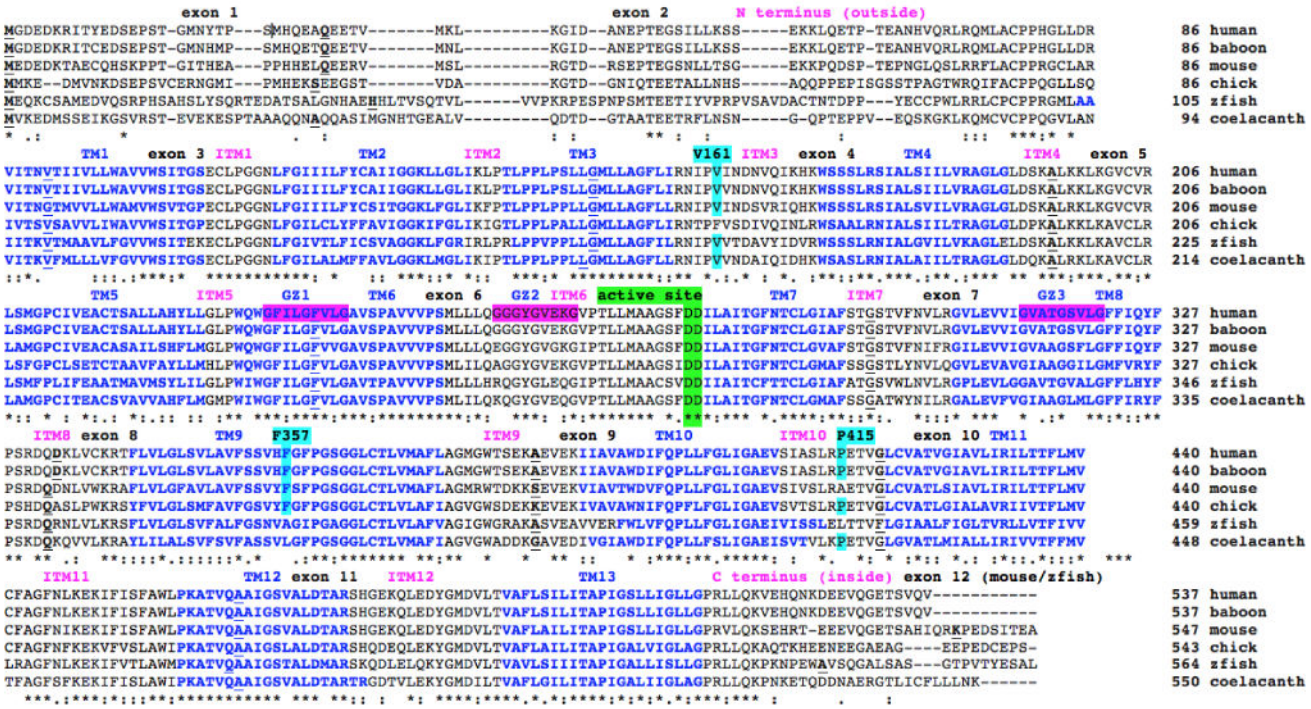

Figure 1.

Amino acid sequence alignments for human and other vertebrate SLC9B2 sequences. See Table 1 for sources of SLC9B-like sequences; *shows identical residues for SLC9B2 subunits; similar alternate residues; dissimilar alternate residues; predicted transmembrane residues are shown in blue and numbered in sequence TM1, TM2 etc.; predicted interhelical segments are numbered in sequence (ITM1, ITM2 etc.); active site residues are in green; predicted glycine zipper sequences are shaded pink; bold underlined font shows residues corresponding to known or predicted exon start sites; exon numbers refer to human SLC9B2 gene exons; key functional residues are shown in turquoise [23]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

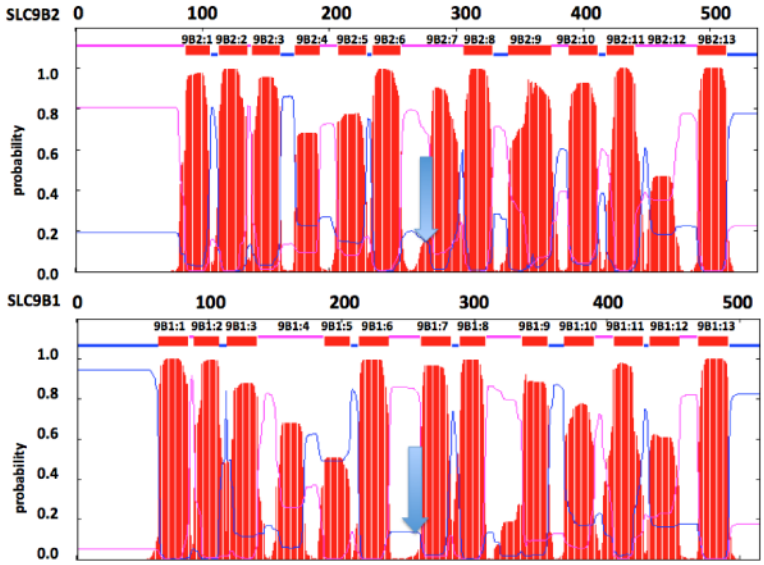

Figure 2.

Comparisons of the predicted transmembrane structures for human SLC9B2 and SLC9B1. The transmembrane structures for the human SLC9B2 and SLC9B1 subunits are based on results [31]. Amino acid residues and predicted transmembrane structures are numbered in order from N-terminus as 9B2:1, 9B2:2 etc. for SLC9B2 and 9B1:1, 9B1:2 etc. for human SLC9B1. The arrow indicates position of the active site. Predicted transmembrane regions are shown as red bars. Pink and blue colors indicate external and internal inter-membrane regions

Figure 3.

Amino acid sequence alignments for human SLC9B2 and SLC9B1 sequences.

See Table 1 for sources of human SLC9B2 and SLC9B1 sequences; *shows identical residues for SLC9B subunits; similar alternate residues; dissimilar alternate residues; predicted transmembrane residues are shown in blue and numbered in sequence TM1, TM2 etc.; predicted interhelical segments are numbered in sequence (ITM1, ITM2 etc.); active site residues are in green; predicted glycine zipper sequences are shown as GZ1, GZ2 and GZ3 for human SLC9B2; bold underlined font shows residues corresponding to known or predicted exon start sites; exon numbers refer to human SLC9B2 gene exons; proline residues are shaded in pink; glycine residues in turqouise

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

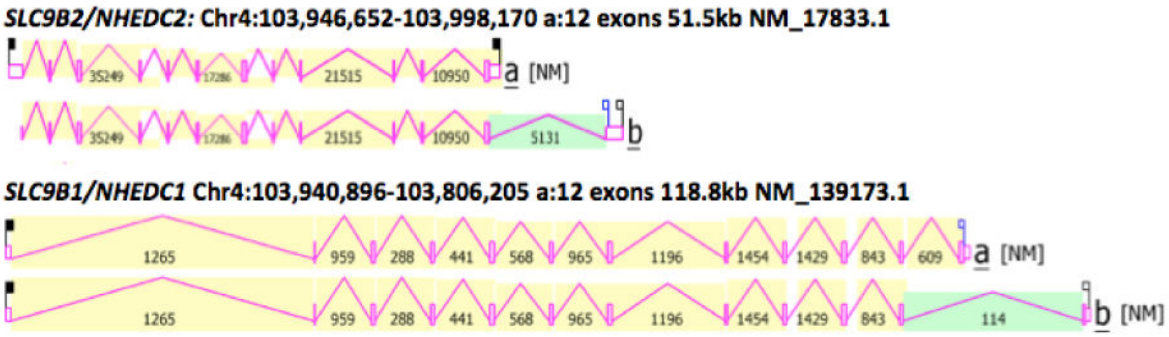

Figure 4.

Gene structures, major transcript isoforms and chromosomal locations for human *SLC9B2* and human *SLC9B1* genes.

Derived from the AceView [15]; shown with capped 5′- and 3′- ends for the two isoform mRNA sequences, in each case; NM refers to the NCBI reference sequence; exons are in pink; numbers within the intron sequences refer to bps

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

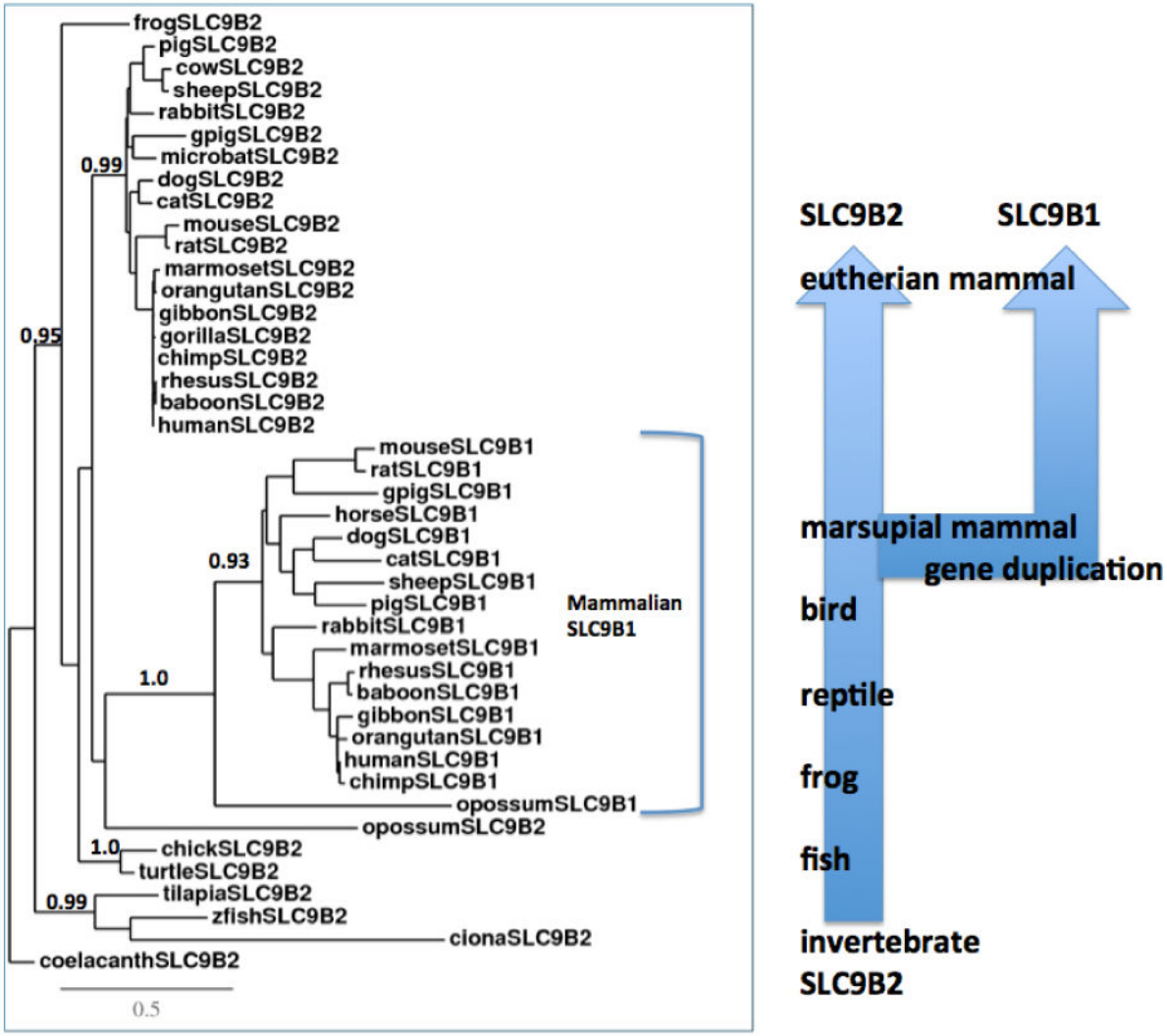

Figure 5.

Phylogenetic tree of vertebrate SLC9B2 and mammalian SLC9B1 amino acid sequences. The tree is labeled with the SLC9B-like name and the name of the animal and is 'rooted' with the coelacanth (*Latimeria chalumnae)* SLC9B2 sequence, which was used to 'root' the tree. Note the 2 major groups corresponding to the *SLC9B2* and *SLC9B1* gene families. A genetic distance scale is shown. The number of times a clade (sequences common to a node or branch) occurred in the bootstrap replicates are shown. Only replicate values of 0.9 or more, which are highly significant, are shown with 100 bootstrap replicates performed in each case. A proposed duplication event is shown arising from an ancestral invertebrate *SLC9B2*-like gene, generating the mammalian *SLC9B1* gene

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Vertebrate *SLC9B2* and mammalian *SLC9B1* genes and proteins.

Gene	Species	RefSeq ID	GenBank ID	UNIPROT Amino		Chromosome	Coding Exons	Gene Size	Subunit	p[1]	% Identity	% Identity
SLC9B2	-	'NCB! ID	-	ID	Acids	Location	(Strand)	bps	MW		Human SLC9B2	Human SLC9B1
Human	Homo sapiens	NM_178833.5	BC009732	Q86UD5	537	4:103,947,530-103,988,707	11 (-ye)	41,178	57,564	6.3	100	53
Rhesus	Macaca mulatta	'XP_001110438.1	CH471198	F6QDG5	537	597,274,361-97,312,198	11 (-ye)	37,838	57,468	6	98	52
Mouse	Mus musculus	IXP_519445	BC090977	Q5BKR2	547	3:135,317,104-135,336,575	12 (+ve)	19,472	58,934	6.8	81	55
Rat	Rattus norvegicus	NM_001109385.2	BC167815	B2RYK8	547	2:259,014,927.259,035,569	12 (+ve)	20,643	58,738	6.5	82	53
Pig	Sus scrofa	NP_003129331.2	GACC01000293	F1S111	537	8:126,983,541-127,018,193'	11 (+ve)	34,653•	57,298	5.7	84	54
Opossum	Monodelphis domestica	NP_007495925.1	na	F7E5XS	656	5:46,800,244-46,865,285	12 (+ye)	65,042	71,611	5	52	42
Chicken	Gallus gallus	'XP 004936029.1	na	na	568	4:60,564,448-60,583,177	12 (-ye)	18,730	60,576	5.1	66	53
Zebra fish	Dania rerio	'XP_699710.2	na	na	564	1:44,078,058-44,094,585	12 (-ye)	16,528	6M88	8.9	56	47
Coelacanth	Latimeria chalumnae	NP_006009735.1	na	na	550	1H128666:67,204-118,949	11 (+ye)	51,746	58,525	8	65	52
Sea squirt	Ciona intestinalis	'XP 002128513.1	na	na	535	14p:1,088,299-1,094,369	13 (-ye)	6,071	56,556	5.3	43	41
SLC9B2												
Human	Homo sapiens	NM_139173	6CO22079	441114	515	4:103,822,277-103,912,868	11 (-ye)	90,592	56,054	8.3	53	100
Rhesus	Macaca mulatta	NP_001110263.1	na	F6URU9	511	5:97,140,216-97,227,958	11 (-ye)	87,743	55,504	9	51	92
Mouse	Mus musculus	NM_028946.3	AK029525	G5E812	565	3:135,355,009-135,397,805	12 (+ve)	42,797	61,986	6.8	50	64
Rat	Rottus norvegicus	NM_001173773.1	na	F1LPN2	519	2:259,055,735-259,096,954	12 (+ve)	41,220	56,581	6.4	51	64
Pig	Sus scrofa	na	CU862009	F1S110	514	8:127,051,477-127,103,437	11 (+ve)	51,961	55,503	8.5	53	67
Opossum	Monodelphis domestica	IXP_007496035 1	na	F6V2D6	495	5:46,890,82646,981,997	11 (+ve)	91,172	53,953	9.3	47	51

RefSeq: the reference amino acid sequence; [1]predicted Ensembl amino acid sequence; na-not available; GenBank IDs are derived NCBI http://www.ncbi.nlm.nih.gov/ genbank/; Ensembl ID was derived from Ensembl genome database http://www.ensembl.org; UNIPROT refers to UniprotKB/Swiss-Prot IDs for individual proteins (see http://kr.expasy.org); bps refers to base pairs of nucleotide sequences; pI refers to theoretical isoelectric points; the number of coding exons are listed

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2

Conserved key residues for human SLC9B2.

SLC982				
Function	**Name**	**Location**	**Residues**	**Sequence**
N-terminus	N-terminus	Outside membrane	1…86	See Figure 1
Book end to TM1	PreTMl	Outside membrane	77…84	CPPHRGLLDR
TM helix 1	TM1/992:1	Transmembrane	87.-107	VITNVTIIVLLWAWWSITGS
Interhelix spacer	ITM1	Inside membrane	108…114	ECLPGGN
TM helix 2	TM2/9B2:2	Transmembrane	115.-135	LEGIIILFYCAUGGKLLGLI
Interhelix spacer	ITM2	Outside membrane	136…138	KLP
TM helix 3	TM3/9B2:3	Transmembrane	139.156	TLPPLPSUGMLLAGFLI
Interhelix spacer	ITM3	Inside membrane	157_.171	RNIPVINDNVQIKHK/R
TM helix 4	TM4/982:4	Transmembrane	172…191	WSSSUISIALSIILVRAGLG
Interhelix spacer	ITM4	Outside membrane	192-.206	LDSKALKKLKGVCVR
TM helix 5	TM5/992:5	Transmembrane	207-227	LSMGPCIVEACTSALLAHYLLGL
Interhelix spacer	ITMS	Inside membrane	228-230	GLP
TM helix 6	TM6/9B2:6	Transmembrane	231-252	WQWGFILGFVLGAVSPAWVPS
Active site zone	ITM6	Outside membrane	253-279	PALLIQGGGYGVEKGVPRIMAAGSFDD
TM helix 7	TM7/9B2:7	Transmembrane	280-294	ILAITGFNTCLGIAF
Interhelix spacer	ITM7	Inside membrane	295-305	STGSTVSTGSTVVFN VAR
TM helix 8	TM8/9B2:8	Transmembrane	306-327	GVLEVVIGVATGSVLGFFICILF
Interhelix spacer	ITM8	Outside membrane	328-327	PSRQQDKLVCKRT
TM helix 9	TM9:9B2:9	Transmembrane	328-340	FLVLGLSVLAVFSSVHFGFPGSGGLCTLVMAFL
Interhelix spacer	ITM9	Inside membrane	341-373	AGMGWTSEKAEVEK
TM helix 10	TM10/982:10	Transmembrane	374-408	HAVAWDIFQPLLFGLIGAEV
Interhelix spacer	ITM10	Outside membrane	409-419	SIASLRPETVG
TM helix 11	TM11/9B2:11	Transmembrane	420-440	LCVATVGIAVLIRILTTFLMV
Interhelix spacer	ITM11	Inside membrane	441-458	CFAGFNLKEKIFISFAWI
TM helix 12	TM12/9132:12	Transmembrane	459-476	PKATVQAAIGSVALDTAR
Interhelix spacer	ITM12	Outside membrane	477-492	SHGEKQLEDVGMDVLT
TM helix 13	TM13/992:13	Transmembrane	493-513	VAFLSILITAPIGSLLIGLLG
C-terminus	postTM13	Inside membrane	514-537	PRLIQKVEHQNKDEEVQGETSVQV

Predicted key residues are identified based upon their positioning within transmembrane and/or inter-transmembrane structures; conserved residues are underlined; predicted Glycine Zipper sequences were identified designated as GZ1 (GFILGFVLG), GZ2 (GGGYGVEKG) and GZ3 (GVATGSVLG); a highly basic amino acid residue region was observed within ITM4, shown in blue; single and double Gly (G) and Pro (P) sequences were identified which may support the formation of transmembrane helices; the Asp-Asp (DD) active site sequence is shown in red.