Molecular cloning, genomic organization, genetic variations, and characterization of murine sterolin genes *Abcg5* and *Abcg8*

Kangmo Lu,[*] Mi-Hye Lee,[*] Hongwei Yu,[*] Yuehua Zhou,[*] Shelley A. Sandell,[*] Gerald Salen,[†] and Shailendra B. Patel[1,*]

Division of Endocrinology, Diabetes, and Medical Genetics,[*] Medical University of South Carolina, Charleston, SC 29403; and Division of Gastroenterology,[†] University of Medicine and Dentistry New Jersey, Newark, NJ 07018

Abstract **Mammalian physiological processes can distinguish between dietary cholesterol and non-cholesterol, retaining very little of the non-cholesterol in their bodies. We have recently identified two genes, *ABCG5* and *ABCG8*, encoding sterolin-1 and -2 respectively, mutations of which cause the human disease sitosterolemia. We report here the mouse cDNAs and genomic organization of *Abcg5* and *Abcg8*. Both genes are arranged in an unusual head-to-head configuration, and only 140 bases separate their two respective start-transcription sites. A single TATA motif was identified, with no canonical CCAT box present between the two genes. The genes are located on mouse chromosome 17 and this complex spans no more than 40 kb. Expression of both genes is confined to the liver and intestine. For both genes, two different sizes of transcripts were identified which differ in the lengths of their 3′ UTRs. Additionally, alternatively spliced forms for Abcg8 were identified, resulting from a CAG repeat at the intron 1 splice-acceptor site, causing a deletion of a glutamine. We screened 20 different mouse strains for polymorphic variants. Although a large number of polymorphic variants were identified, strains reported to show significant differences in cholesterol absorption rates did not show significant genomic variations in *Abcg5* or *Abcg8*.**—Lu, K., M-H. Lee, H. Yu, Y. Zhou, S. A. Sandell, G. Salen, and S. B. Patel. **Molecular cloning, genomic organization, genetic variations, and characterization of murine sterolin genes *Abcg5* and *Abcg8*. *J. Lipid Res.* 2002.** 43: **565–578.**

Supplementary key words dietary cholesterol • sitosterolemia • genetics • sterol transporter • ATP-binding cassette • inbred mouse strains

Sitosterolemia (also known as phytosterolemia, MIM 210250) is a rare autosomal recessively inherited metabolic disorder, that was first discovered in 1974 in two sisters (1). Sitosterolemia patients develop tendon and tuberous xanthomas, hemolytic episodes, arthralgias, and arthritis, and premature coronary and aortic atherosclerosis leading to cardiac fatalities (2–4). Affected individuals have very high levels of plasma plant sterols (sitosterol, campesterol, stigmasterol, avenosterol, and others) and 5∝-saturated stanols, particularly sitostanol, but their blood cholesterol levels may be normal or only moderately increased. Affected individuals show not only very high levels of dietary cholesterol absorption, they absorb all types of sterols and retain them in the body. Thus in contrast to normals, affected individuals absorb and retain plant sterols (2, 3), as well as shellfish sterols (5). Clinical studies further show that affected individuals have an inability to excrete sterols, both plant sterols as well as cholesterol, into bile (6, 7). Increased intestinal absorption, decreased hepatic excretion of sitosterol (the major plant sterol), and abnormally low cholesterol biosynthesis are reported to be features of sitosterolemia (5–9).

Previous studies on sitosterolemia have suggested that the defect in sitosterolemia may involve a putative sterol transporter expressed in the intestine and/or the liver (10, 11). Using linkage analyses, the *STSL* locus has been mapped to human chromosome 2p21 and no evidence of genetic heterogeneity was detected in a large cohort of sitosterolemia pedigrees (11, 12). By using positional cloning procedures (11–13), we isolated two novel genes from *STSL* locus, chromosome 2p21, between *D2S2294* and *D2S2298*. Both genes encode for proteins that contain an ATP binding signature sequence, characteristic of the ABC family of proteins, at the N-terminal and a transmembrane domain located at the carboxyl terminus consisting of 6 membrane-spanning helices. These findings have also been reported independently by Berge et al. (14, 15).

Abbreviations: EST, expressed sequence tag; IMAGE, integrated molecular analysis of genomes and their expression; LXR, liver X receptor; QTL, quantitative trait loci; SNP, single nucleotide change.

[1] To whom correspondence should be addressed at the Division of Endocrinology, Medical University of South Carolina, Endocrinology-Diabetes Medical Genetics, 114 Doughty Street, P.O. Box 250776, Charleston, SC 29403.
e-mail: patelsb@musc.edu

These genes, using the Human Genome Organization nomenclature, have been named *ABCG5*, encoding sterolin-1, and *ABCG8*, encoding sterolin-2, respectively. Mutations in *ABCG5* and *ABCG8* genes are responsible for causing sitosterolemia (14–16). Based upon the clinical defects in sitosterolemia, these proteins are predicted to play a pivotal role in the selective absorption of cholesterol from the intestinal lumen and in the selective excretion of noncholesterol sterols into bile (4, 16, 17). We report here (1) the murine genomic organization, cDNA, and tissue expression patterns of these genes, (2) an unusual, but naturally occurring alternative splicing of a mouse ABCG8 cDNA resulting in an in-frame deletion of a single amino acid, and (3) polymorphic variations in *Abcg5* and *Abcg8* in 20 inbred mouse strains.

MATERIALS AND METHODS

Database searches

Human ABCG5 and ABCG8 cDNA sequence information was used to search the databases for any murine homologous sequences using the Basic Local Alignment Search Tool (18, 19). Three murine expressed sequence tag (EST) integrated molecular analysis of genomes and their expression (IMAGE) clones (20) containing sequences homologous to human ABCG5 cDNA and three clones homologous to human ABCG8 cDNA were identified. These clones were obtained from Research Genetics (Huntsville, AL) and fully sequenced. IMAGE clones 1885393 and 1885458 contained Abcg8 coding sequences and were used for further analyses.

Northern-blot analysis

A mouse multiple-tissue poly(A)$^+$ RNA Northern blot (Origene, Bethesda, MD) was hybridized with a radio-labeled full-length mouse Abcg5 or Abcg8 cDNA as previously described (14). The hybridized filter was washed stringently with 0.1 × SSC-0.1% SDS at 68°C, exposed to a phosphorimager cassette, striped, and re-probed with a mouse β-actin probed for comparison of RNA loading.

cDNA library screening

A mouse cDNA lambda phage library (Stratagene, La Jolla CA) was screened with a partial length Abcg8 cDNA probe to identify full-length cDNA clones (21). Screening an estimated 3×10^6 for Abcg8, we identified more than 80 positive clones, 69 of which were plaque purified and characterized further. A library screen for Abcg5 cDNAs was not performed (see Results). All purified positive clones were excised in vivo into insert-containing pBSK phagemid vector by infection with ExAssist helper phage followed by transduction of filamentous phage particles into *E. coli* XLOLP according to the manufacturer's protocols. The phagemid DNA was prepared by plasmid miniprep Kit (Roche, Indianapolis, IN) and cycle-sequenced using BigDye Terminator Cycle Sequencing Ready Reaction Kit and an ABI 373 DNA genetic analyzer.

TABLE 1. Oligonucleotide primers used for amplification of *Abcg5* and *Abcg8*

Abcg5

Sense Primers			Antisense Primers		
Primer Name	Position in cDNA	Sequences 5′ to 3′	Primer Name	Position in cDNA	Sequences 5′ to 3′
mg5-F1	142-162	GGTGAGCTGCCCTTTCTGAGT	mg5-2R	363-343	CAAGGAGACATCTTTGAGGAT
mg5-F4	343-363	ATCCTCAAAGATGTCTCCTTG	mg5-3R	537-518	GACGTAGGAGAAGCAGTCTT
mg5-3F	469-489	GAAGGGGAGGTGTTTGTGAA	mg5-4R	608-589	AGGGCCAGCATCGCTGTGTA
mg5-4F	553-573	TTTCTGAGCAGCCTCACTGTG	mg5-5R	775-757	TGGGGTCCTGAAGGAGTTG
mg5-5F	643-662	GTAGAGGCAGTCATGACAGA	mg5-6R	915-896	TTGGAAGAGCTCAGAGCGAG
mg5-6F	776-796	AGGTCATGATGCTAGATGAGC	mg5-7R	1036-1016	CAAAGGGATTGGAATGTTCAG
mg5-7F	919-939	TTCGACAAAATTGCCATCCTG	mg5-8R	1250-1231	ACACCAAGCTTGCCGAACAT
mg5-8F	1050-1069	CTTGACATCAGTGGACACCC	mg5-9R	1462-1442	GATTCACAGCATTGAGCATGC
mg5-9F	1260-1282	GCGAGTAACAAGAAACTTAATGA	mg5-10R	1597-1577	ACACACTGCTGAAAATGACCG
mg5-10F	1466-1485	TTCCCATGCTGAGAGCCGTC	mg5-11R	1781-1762	CCAGATCCAATAAGCAGCCC
mg5-11F	1605-1624	GACTCTGGGCTTGTATCCTG	mg5-12R	1776-1756	CATTGACCACGAGAATCTCAC
mg5-12F	1756-1776	GTGAGATTCTCGTGGTCAATG	mg5-13R	2160-2140	TACTTCTCTGTGCTCCACAGT
mg5-13F	1968-1988	GAGAAAACCTGCCCAGGTGC	mg5-R3	1806-1486	TCCTGACTCTCCTGGTCGCT

Abcg8

Sense Primers			Antisense Primers		
Primer Name	Position in cDNA	Sequences 5′ to 3′	Primer Name	Position in cDNA	Sequences 5′ to 3′
mg8-F1	42-62	AAAGACAGAGAGAGCCCAACA	mg8-2R	264-244	AGGTGAGATCTCTGACCTCC
mg8-F3	188-208	CTCCTCGGAAAGTGACAACAG	mg8-R11	417-397	CCCTATGATGGCCAGCATCT
mg8-3F	285-305	CAGGTGCCTTGGTTTGAGCAG	mg8-R10	544-525	CACTTCCTCACCAGCTGAGG
mg8-F2	432-451	GGGAGAGCCTCACTACTCGA	mg8-4R	742-723	ACCCCACGTACATACGTGTT
mg8-4F	723-742	AACACGTATGTACGTGGGGT	mg8-R7	939-920	GCCTGAAGATGTCAGAGCGA
mg8-F7	920-939	TCGCTCTGACATCTTCAGGC	mg8-7R	1223-1203	GTGGGTGCTTGTGTTGAGTTC
mg8-F4	1075-1094	TGACCAGCATCGACAGACGC	mg8-8R	1274-1254	CTCAACAGCAGTCCCACAGTC
mg8-8F	1254-1274	GACTGTGGGACTGCTGTTGAG	mg8-9R	1505-1485	GACGACATCCAGGATGACATT
mg8-9F	1320-1340	ATTTCCAATGACTTCCGGGAC	mg8-10R	1565-1545	AGTGTACAGCCCGTCTTCCAG
mg8-10F	1545-1565	CTGGAAGACGGGCTGTACACT	mg8-R12	1807-1788	GAGTTGTAGAGGGCATTGCA
mg8-F12	1691-1709	ACACTTCCTGCTCGTGTGG	mg8-R13	1973-1953	GATGGAGAAGGTGAAGTTGCC
mg8-F13	1953-1973	GGCAACTTCACCTTCTCCATC	mg8-R2	2230-2211	GGTGGCTGCTGCCTGAGAGA

Reverse transcriptase and rapid amplification of cDNA ends-PCR

Total RNA was isolated from freshly isolated mouse liver using TRIzol reagent (Life Technologies, Gaithersburg, MD). Reverse transcription of 10 μg of total RNA was performed using oligonucleotide d(T) and the single-strand cDNA used as template to amplify mouse Abcg5 cDNA using primers mg5-F4 and mg5-R3 (**Table 1**). Rapid amplification of cDNA ends (RACE) was performed using a commercially available SMART RACE cDNA Amplification Kit (Clontech, Palo Alto, CA) according to the manufacturer's protocol, using total liver or intestinal RNA. Five prime and three prime RACE cDNA was synthesized using primers supplied with the kit. PCR products were directly sequenced or subcloned using TA cloning and isolated clones sequenced as described above.

BAC library screening

The sequences of Abcg5 and Abcg8 cDNA were used to design primer sets (Table 1) and used to screen a Mouse BAC library CitbCJ7 (Research Genetics) by PCR of DNA superpools and plates, using a modified protocol as previously described (13). The primers used for screening for Abcg5 were mg5-9F, 9R, mg5-13F, and mg5-13R. Primers mg8-1F, mg8-1R, mg8-13F, and mg8-13R were used for Abcg8 screening. Five μl of pooled DNA from the BAC library was used in a 10 μl PCR reaction containing 1.5 mM $MgCl_2$, 16 mM $(NH_4)_2SO_4$, 0.1 mM dNTPs, 67 mM Tris (pH 8.8), 1 μM primer, and 1 unit of Taq polymerase with the following protocol: 94°C for 30 s, 60°C for 30 s, 72°C for 1 min for 35 cycles. Positive mouse BAC clones were obtained from Research Genetics, Inc., plated on agar plates containing 12.5 μg/ml chloramphenicol, and confirmed by colony PCR using the above primer sets.

Direct BAC Sequencing and exon/intron boundary

BAC plasmid DNA was prepared and direct BAC sequencing was performed as previously described (13).

In addition to direct sequencing of the BAC, amplification of genomic fragments, using primers selected from the cDNA sequences (Table 1), was used to determine exon/intron boundary sequences.

Analysis of alternative splice forms

To quantitate the levels of the two alternatively spliced forms of Abcg8 RNAs, we utilized the loss of *Eco*O109 I restriction enzyme recognition site, as a result of the loss of the CAG triplet codon from the cDNA. Fragments containing the splice difference were amplified by RT-PCR from cDNA using mouse liver, jejunum, ileum, or colon total RNA using primers mg8-1F and mg8-R11, to yield an expected PCR product of 322–325 bp (Table 1). PCR reactions were spiked with ^{32}P end-labeled primer mg8-R11 at the ratio of 1:10 unlabeled primer for detection of PCR products. PCR products were digested with *Eco*O109 I and separated by 5% acrylamide gel electrophoresis under non-denaturing conditions. The gels were dried and exposed to a phosphorimager cassette (Molecular Dynamics, CA) and analyzed using ImageQuant *v*1.11 software. Although a 322–325 bp sized PCR products were expected, two additional bands migrating slightly slower were consistently identified following agarose or acrylamide gel electrophoresis. These DNA bands were purified and directly sequenced using the amplification primers and were found to represent heteroduplexes between the spliced two forms (see Results).

Chromosomal localization

To localize *Abcg5* and *Abcg8* on the mouse chromosomes, we performed electronic PCR. We initially identified the corresponding mouse chromosome syntenic to human chromosome 2p21 (http://www.ncbi.nlm.nih.gov/). We selected markers spanning the *STSL* locus whose murine counterparts were also known and had been mapped (http://www.informatics.jax.org/). To confirm the mapping, we identified mapped mouse EST sequences (http://www.genome.wi.mit.edu/mouse_rh/index.html) that shared sequence identity with Abcg5 and Abcg8 cDNA (see Results).

Animals

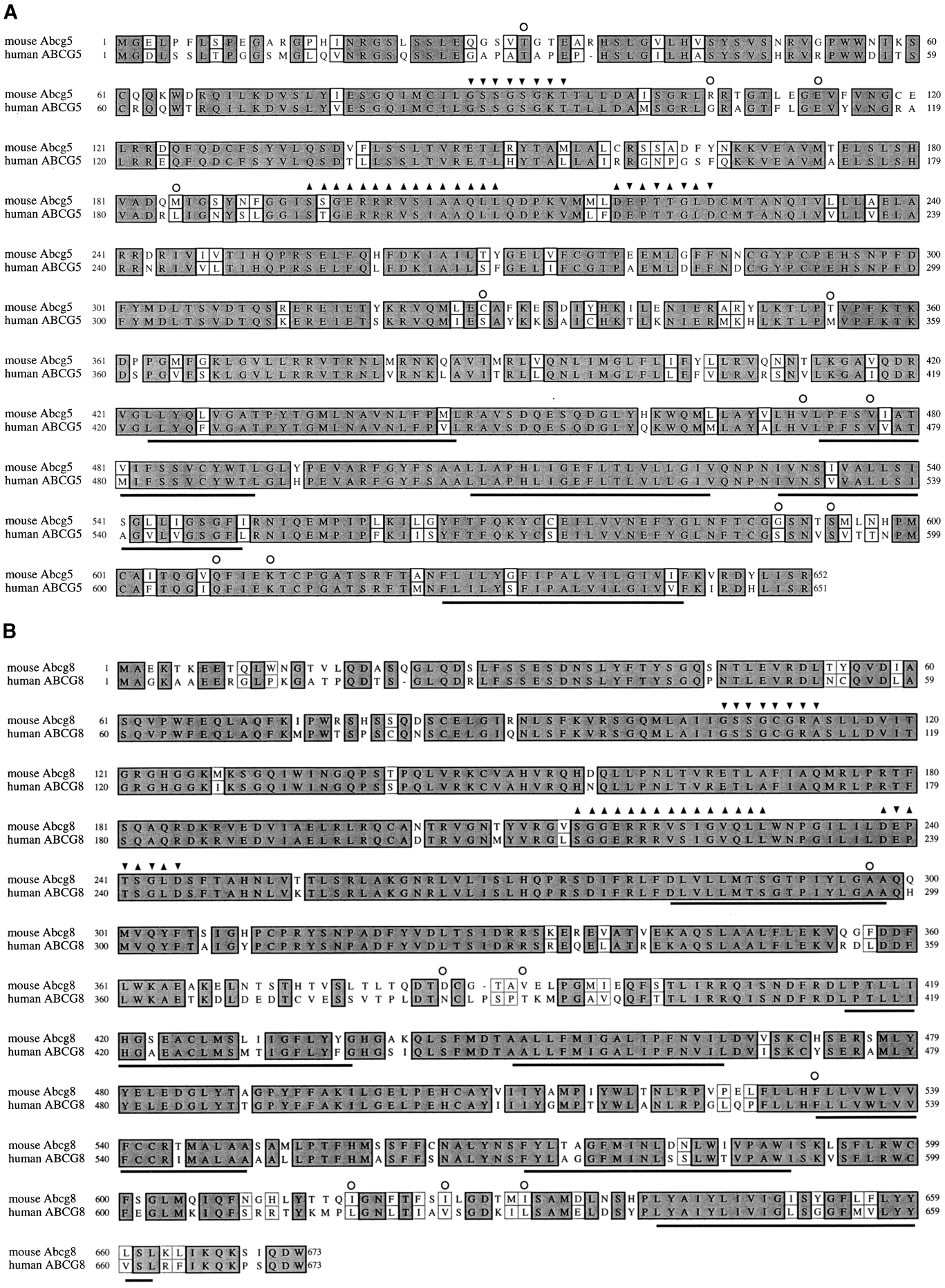
Various mouse inbred strain genomic DNAs were obtained from the DNA resource, Jackson Laboratory, Bar Harbor, ME (http://www.jax.org/resources/documents/dnares/index.html) and polymorphic variants detected by direct PCR amplification and sequencing. Mouse strains were chosen based up the reported plasma cholesterol levels (22) (http://aretha.jax.org/pub-cgi/phenome/mpdcgi?rtn=docs/home), as well from studies reported to show variations in dietary cholesterol absorption (23–26). For plasma sitosterol levels, two females of each selected strain were housed for 1 week, fed standard rodent chow, and fasted 2 h before sacrifice for blood and tissue sampling. Plasma sitosterol and campesterol levels were determined as previously described (27).

RESULTS

Identification of mouse Abcg5 and Abcg8 cDNA

By searching the mouse EST database with the human ABCG5 cDNA sequence, we identified three mouse ESTs (GenBank Accession Nos. AI505249, AI507053, and AA237916). Sequence analyses of these clones indicated that they contained partial Abcg5 cDNA sequences. To obtain full-length cDNA sequence, we used a combination of RT-PCR and 5′ RACE. The complete sequence analysis of mouse Abcg5 cDNA demonstrated that it encoded an open reading frame of 652 amino acids with a calculated molecular mass of 75 kDa (**Fig. 1**). The deduced amino acid sequence of mouse Abcg5 showed a high degree of conservation, with a 92.8% and 80.1% homology match with rat and human ABCG5 proteins, respectively. Mouse Abcg5 has an extra amino acid, R35, compared human ABCG5. A poly(A)$^+$ site was not identified in the 3′ UTR and 3′ RACE failed to extend the known 3′ end for this cDNA.

Using human ABCG8 cDNA sequence information, we identified three homologous mouse cDNA sequences in the databases (IMAGE clones 1925061,1885393, and 1885458). Clone 1925061 contained an internal deletion and was not analyzed further. The remaining two clones were identical and contained an insert of 2.6 kb. Sequence analyses showed that these clones contained the full-length mouse Abcg8 cDNA with a poly(A)$^+$ tail, located 13 bp down-stream of an AACAAA sequence motif. Two other poly(A) signals (AAUAAA) at 189 and 224 bp up-stream of the poly(A)$^+$ tail were also present. Translation of the longest open reading frame revealed a protein of 673 amino acids, with calculated molecular mass of 76 kDa (Fig. 1). The predicted amino acid sequence of mouse Abcg8 showed a high degree of conservation, with 90.6% and 81.5% homology to rat and human ABCG8, respectively (14, 16).



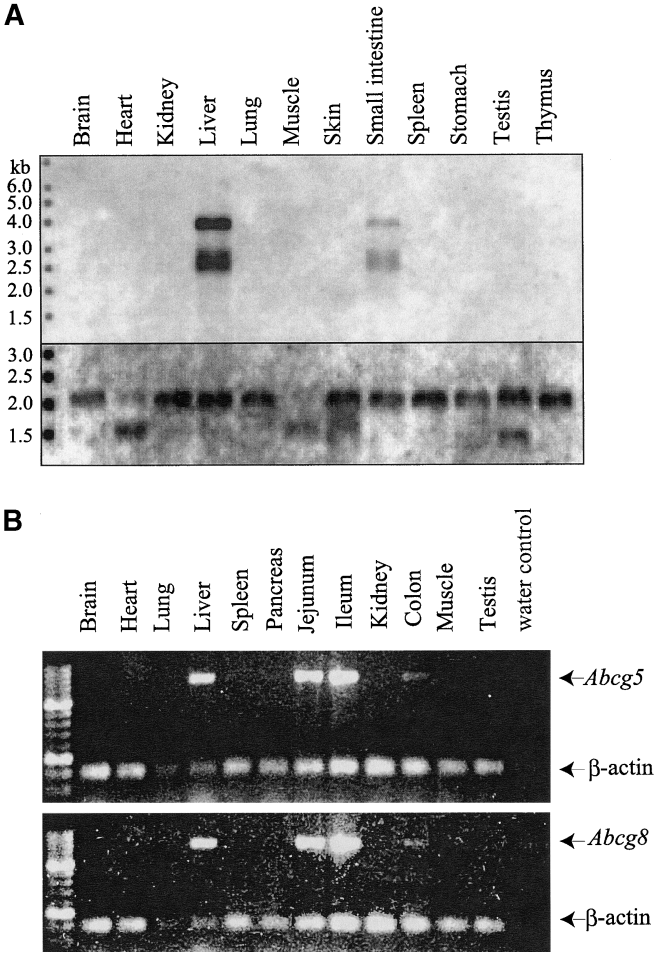


Fig. 2. Expression patterns of *Abcg5* and *Abcg8* genes. Northern blot analyses of poly(A)$^+$ RNA from a variety of tissues (see Materials and Methods) showed that expression of *Abcg8* was confined to the liver and small intestine (A). Data for mouse *Abcg5* mRNA expression have been published previously (14). RT-PCR analyses confirmed the expression patterns and further showed that expression was predominantly limited to the small intestine and liver. As a control, β-actin was co-amplified as an internal control and the last tracks indicate no template (water-only) controls.

Mouse Abcg5 and Abcg8 show a 30.8% homology to each other at the peptide level. Both proteins have highly conserved ATP-binding cassette signature motifs (Walker A, B, and C) (Fig. 1) located at the N-terminal half, and a predicted 6-transmembrane domain located at the C-terminal end (underlined, Fig. 1).

Tissue distribution of mouse Abcg5 and Abcg8 mRNA

Expression patterns of mouse *Abcg5* and *Abcg8* genes were examined by Northern blot and RT-PCR methods (**Fig. 2**). Expression of *Abcg5* (14) and *Abcg8* appear to be restricted to the liver and the small intestine, with a low level of expression in the colon. Size heterogeneity for both Abcg5 mRNA (2.3 kb and 3.3 kb) (14) and Abcg8 mRNA (2.6 kb and 3.7 kb) (Fig. 2A) was observed. In order to examine the nature of the size difference of the two transcripts for Abcg8, we screened a mouse liver phage cDNA library using Abcg8 as probe. Two major lengths of cDNAs, 2.56 and 3.67 kb respectively, were identified. Sequence analyses showed the size differences arose from 3′ UTR variation (data not shown. See Genbank submission AF324494). Although a library screening for Abcg5 cDNA was not performed, we predict that a similar variation on the 3′ UTR may account for the larger sized message.

RT-PCR analyses of RNA from different mouse tissues confirmed that the expression of these genes was confined to the liver, jejunum, and ileum, with weak expression in the colon (Fig. 2B).

Gene organization, exon-intron boundary, and promoter analyses of *Abcg5* and *Abcg8* genes

In order to obtain genomic information for both genes, we screened a mouse BAC library (CitbCJ7) using primer sets designed from the first and last exon sequences of *Abcg5* and *Abcg8*. Three BAC clones, 329B11, 329A22, and 376H16, were identified, all of which contained the full-length genomic fragments for both *Abcg5* and *Abcg8*. Exon-intron boundaries were determined by direct sequencing of the BAC DNA and/or long PCR amplified products using exon specific primers. The two genes are organized in a head-to-head tandem configuration and each gene contains 13 exons and 12 introns (**Fig. 3A** and **Table 2**). *Abcg5* spans about 23 kb of genomic DNA and *Abcg8* spans ~16 kb of genomic DNA. The exon sizes range from 98 to 206 bp and intron sizes from 87 bp and to about 5 kb. All exon-intron boundaries show canonical sequences with initial GT at the splice-donor and terminal AG at splice-acceptor sites.

By chromosome primer walking using reverse oligonucleotide primers located in exons 1 of mouse *Abcg5* and *Abcg8*, we discovered that both genes were in a head-to-head configuration, with a distance of 358 bp separating their respective initiation codons (Fig. 3B). Analysis of this region using TESS software (http://www.cbil.upenn.edu/tess/index.html) revealed a number of potential binding sites for transcription factors (Fig. 3B), but these were restricted to the mouse sequence only, the human sequence showing variances at almost all of these potential sites. A single TATA box was identified 232 bp up-stream of the mouse *Abcg5* ATG initiation codon, as well as two GATA motifs (Fig. 3B).

This putative promoter region sequence has a 63% homology to the corresponding human sequence (Fig. 3B), though the human sequence does not contain the TATA

Fig. 1. Polypeptide sequence and secondary features of mouse Abcg5 and Abcg8 cDNAs. The sequence and secondary features for *Abcg5* (A) (nucleotide sequence Genbank No. AF312713) and *Abcg8* (B) (nucleotide sequence Genbank No. AF324495) are as shown. The positions of the Walker A motif (inverted triangles), B motif (alternating triangles), and C motif (upright triangles) are as indicated. The predicted transmembrane domains are underlined. Shaded boxes indicate complete homology, with conserved sequence changes indicated by the open boxes. The open circles indicate the positions of coding polymorphic variants detected in different mouse strains (see also Table 3).

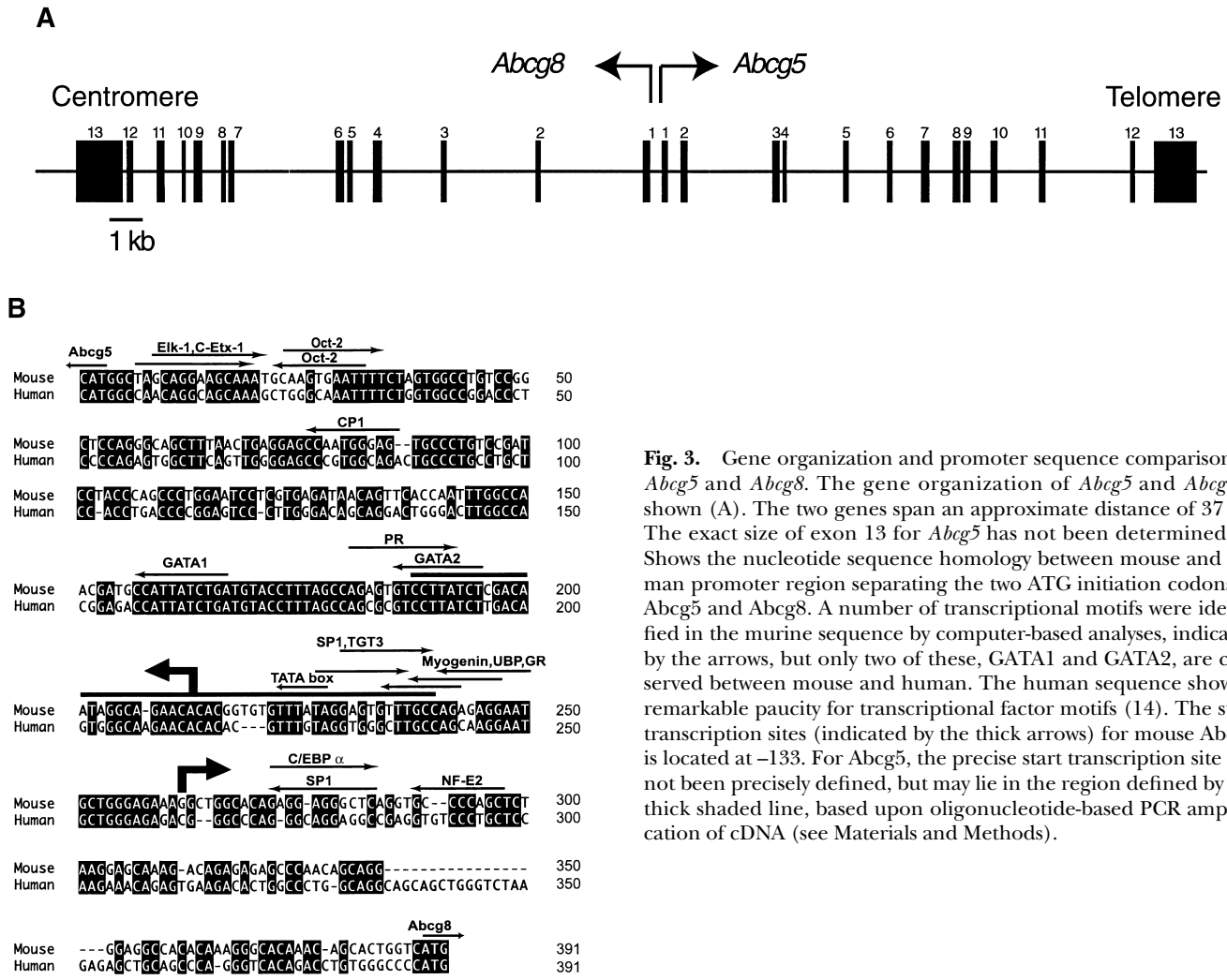


Fig. 3. Gene organization and promoter sequence comparison of *Abcg5* and *Abcg8*. The gene organization of *Abcg5* and *Abcg8* is shown (A). The two genes span an approximate distance of 37 kb. The exact size of exon 13 for *Abcg5* has not been determined. B: Shows the nucleotide sequence homology between mouse and human promoter region separating the two ATG initiation codons of Abcg5 and Abcg8. A number of transcriptional motifs were identified in the murine sequence by computer-based analyses, indicated by the arrows, but only two of these, GATA1 and GATA2, are conserved between mouse and human. The human sequence shows a remarkable paucity for transcriptional factor motifs (14). The start transcription sites (indicated by the thick arrows) for mouse Abcg8 is located at –133. For Abcg5, the precise start transcription site has not been precisely defined, but may lie in the region defined by the thick shaded line, based upon oligonucleotide-based PCR amplification of cDNA (see Materials and Methods).

motif. Preliminary results, using a minimal murine 358 bp fragment located between the two ATGs of *Abcg5* and *Abcg8* did not result in expression of a reporter gene when transfected into HEK 293 cells in either orientation (data not shown). Thus this sequence does not appear to function as a 'minimal' promoter element.

Chromosome localization

The human *STSL* locus, containing the *ABCG5* and *ABCG8* genes, maps to chromosome 2p21 between microsatellite markers *D2S2298* and *D2S2294* (11, 12). Human chromosome 2p21 is syntenic with mouse chromosome 17 (http://www.ncbi.nlm.nih.gov/Homology/). However, there are small regions within human 2p21-2p16 region that share synteny with mouse chromosome 11. To localize mouse *Abcg5* and *Abcg8*, we searched the Mouse Genome Informatics (http://www.informatics.jax.org/) and Whitehead Institute Center for Genome Research Mouse RH Map (http://www.genome.wi.mit.edu/mouse_rh/index.html) databases and performed electronic PCR using mouse *Abcg5* and *Abcg8* genomic sequences. Two murine sequences were identified (GenBank No. AI114946 and AW112016), that share complete sequence identity with exon 13 of mouse *Abcg8* and exon 10–13 of mouse *Abcg5* respectively. Both of these murine sequences have been mapped between chromosomal markers *D17Mit41* and *D17Mit189* (**Fig. 4**). Thus, based upon both synteny, as well as identification of *Abcg5* and *Abcg8* sequences, these genes map to between markers *D17Mit41* and *D17Mit75* at about 53 cM on the genetic map.

Alternative splicing for Abcg8 cDNA

Compared with the human ABCG8 polypeptide, the mouse Abcg8 polypeptide sequence has one more glutamine located at the beginning of exon 2 (16). By sequencing genomic BAC DNA and RT-PCR products using mouse liver RNA, we identified two alternatively spliced forms of mRNA (**Fig. 5A**). The 5′-end of intron 1 contains a motif (GTGAGT), which conforms to the consensus splice donor site (GURAGU). However, at the 3′-end of intron 1 there is a tandem CAG repeat (Fig. 5). This tandem CAG repeat at the splice acceptor site can result in a choice of splice acceptor sites; either the proximal or the distal AG can be utilized. Mouse liver cDNA library screening resulted in two sets of clones differing in the presence or absence of a CAG triplet codon encoding a glutamine

TABLE 2. Exon-intron boundaries and organization of Mouse *Abcg5* and *Abcg8*

Abcg5

GenBank Accession No.	Exon No.	Exon Size(bp)	Exon 3′ End Sequence	Intron Splice Donor	Intron No.	Intron Size(bp)	Intron Splice Acceptor	Exon 5′ End Sequence	Exon No.
AF351786	1	146	GTCCGACAGCGTCAG	gtaaggggacc	1	596	tttcctttaaag	CAACCGTGTCGGGCC	2
AF351787	2	122	TCTTAGGCAGCTCAG	gtaagtgcctgg	2	~5000	gtcgccccctag	GCTCAGGGAAGACCA	3
AF351788	3	136	TCCTACGTCCTGCAG	gtgggcgtgtcc	3	87	tctggcccctag	AGCGACGTTTTTCTG	4
AF351789	4	98	TTCTACAACAAGAAG	gtacttgaaagtt	4	~2000	gtgtctcttacag	GTAGAGGCAGTCATG	5
AF351790	5	133	TCCTTCAGGACCCCA	gtaagtgggaca	5	1228	ctttgccggcag	AGGTCATGATGCTAG	6
AF351791	6	140	TCTGAGCTCTTCCAA	gtaagggaatgc	6	900	ttgtccaagcag	CACTTCGACAAAATT	7
AF351792	7	129	CCTTTGATTTTTACA	gtaagtgtattct	7	688	gaaaactttag	TGGACTTGACATCAG	8
AF351793	8	214	TGGTGTCCTGCTGAG	gtaagagccttg	8	99	tttggttttcag	GCGAGTAACAAGAAA	9
AF351794	9	206	ATGCTGTGAATCTGT	gtaagtgcctgt	9	836	cttctatgccag	TTCCCATGCTGAGAG	10
AF351795	10	139	CAGTGTGTGTTATTG	gtaaggcggtgt	10	3000	atgttttctag	GACTCTGGGCTTGTA	11
AF351796	11	186	ATCTGGATTTATCAG	gtaagaagaaa	11	~4500	ttttcttaag	AAACATACAAGAGAT	12
AF351797	12	113	TGAACTTCACTTGTG	gtaagtgttctat	12	1495	ttttccttgcag	GTGGATCCAACACCT	13
AF351798	13	258							

Abcg8

GenBank Accession No.	Exon No.	Exon Size(bp)	Exon 3′ End Sequence	Intron Splice Donor	Intron No.	Intron Size(bp)	Intron Splice Acceptor	Exon 5′ End Sequence	Exon No.
AF351799	1	257	CTTCAGGATGCTTCG	gtgagtgagctctgc	1	~2500	ctacatgtctcccag	CAGGGCCTCCAGGAC	2
AF351800	2	105	GATCTCACCTACCAG	gtaggggcacgtgc	2	1818	acctctccccacag	GTGGACATCGCCTCT	3
AF351801	3	157	TCATAGGGAGCTCAG	gtaccaacagaggct	3	3102	tctctggatttgcag	GCTGCGGGAGAGCCT	4
AF351802	4	239	CAGCGTGACAAACGG	gtaacagttggctgg	4	442	tccgcctgtcctcag	GTGGAAGACGTAATC	5
AF351803	5	133	TCCTGTGGAACCCAG	gtgaggcctgggaa	5	94	gtgcaatatccccag	GAATCCTCATTCTGG	6
AF351804	6	270	CTGCGGACTTCTACG	gtgagtggtaaaggc	6	1995	atcttctgcttacag	TGGACTTGACCAGCA	7
AF351805	7	163	CACCCACACAGTCAG	gtacgggaagcccg	7	87	actgctcccaacag	CCTGACCCTCACACA	8
AF351806	8	81	TTCCACCCTGATCCG	gtaaatctccctccc	8	600	gaggctttctttcag	TCGTCAGATTTCCAA	9
AF351807	9	201	ATGTCGTCTCCAAAT	gtgagtgtcacctgc	9	522	ctcccccatctttag	GTCACTCGGAGAGGT	10
AF351808	10	77	TATTTCTTTGCCAAG	gtcagggctgggag	10	545	agctgtgctttgcag	ATCCTAGGAGAATTG	11
AF351809	11	268	ACAACCTGTGGATAG	gtgaggcctgctgcc	11	887	tcttgctgtcttcag	TGCCTGCATGGATCT	12
AF351810	12	128	ATCCTCGGAGACACG	gtacgtagcgaagg	12	84	aatgtctgtccgcag	ATGATCAGTGCCATG	13
AF351811	13	605							

residue at this position (nucleotides 165–167, Genbank No. AF324495). Since the presence or absence of this CAG alters the nucleotide recognition site in the cDNA for the restriction endonuclease *Eco*O109I, we used this to quantitate the prevalence and tissue expression patterns of the two different isoforms. RT-PCR was performed using ^{32}P-labeled primer and total RNA from mouse liver jejunum, ileum, or colon (Fig. 5B). The PCR products were digested by *Eco*O109I and separated by non-denaturing acrylamide gel electrophoresis. Undigested PCR products showed the presence of three bands, two of which migrated slower than the expected size. Upon digestion, the expected size bands were diminished in amounts, though the slower migrating bands were unchanged (digestion results in a 50 bp fragment size that is not shown, Fig. 5B). The two slower migrating bands were shown to be heterodimers of the two forms of spliced products produced during the PCR reaction, and were verified by direct sequencing. These heterodimers would be expected to be resistant to endonuclease digestion. Quantitative analysis of bands (Fig. 5B) showed that these two spliced forms were present in almost equal abundance in mouse liver, jejunum, or colon; 41% of the cDNAs contained the CAG and 59% had the CAG deleted. This is comparable to the equal numbers of such clones identified in the liver cDNA library.

Genetic variations in inbred mouse strains

Inbred mouse strains have been used to identify genes whose genetic variations may be important determinants of atherosclerosis, gall stone formation, or biliary cholesterol secretion (28–32). Some of these inbred mouse strains have been screened for differences in dietary cholesterol absorption (23–26). To identify whether genetic variations in *Abcg5* and *Abcg8* may be responsible for some of these phenotypes, we screened 20 strains. These strains were selected based upon documentation of either cholesterol absorption rates, or having very high levels of plasma cholesterol levels. The latter phenotype was chosen because some sitosterolemia patients presented with elevated levels of plasma cholesterol and were initially diagnosed as pseudohomozygous familial hypercholesterolemia. **Table 3** shows a compilation of both coding and non-coding alterations detected for *Abcg5* and *Abcg8*. There were 21 polymorphisms (14 in Abcg5 and 7 in Abcg8) that altered amino acid coding, 77 intronic nucleotide changes, 7 single nucleotide changes (SNPs) in the 5′UTR, 21 SNPs in the 3′UTR, and 50 single nucleotide changes in exonic regions that did not alter amino acid coding. All of these changes were present as homozygous changes, compatible with the inbreeding of these lines. The greatest variations were noted in DBA/1J and DBA/2J (both identical to each other), with CAST/Ei, SPRET/

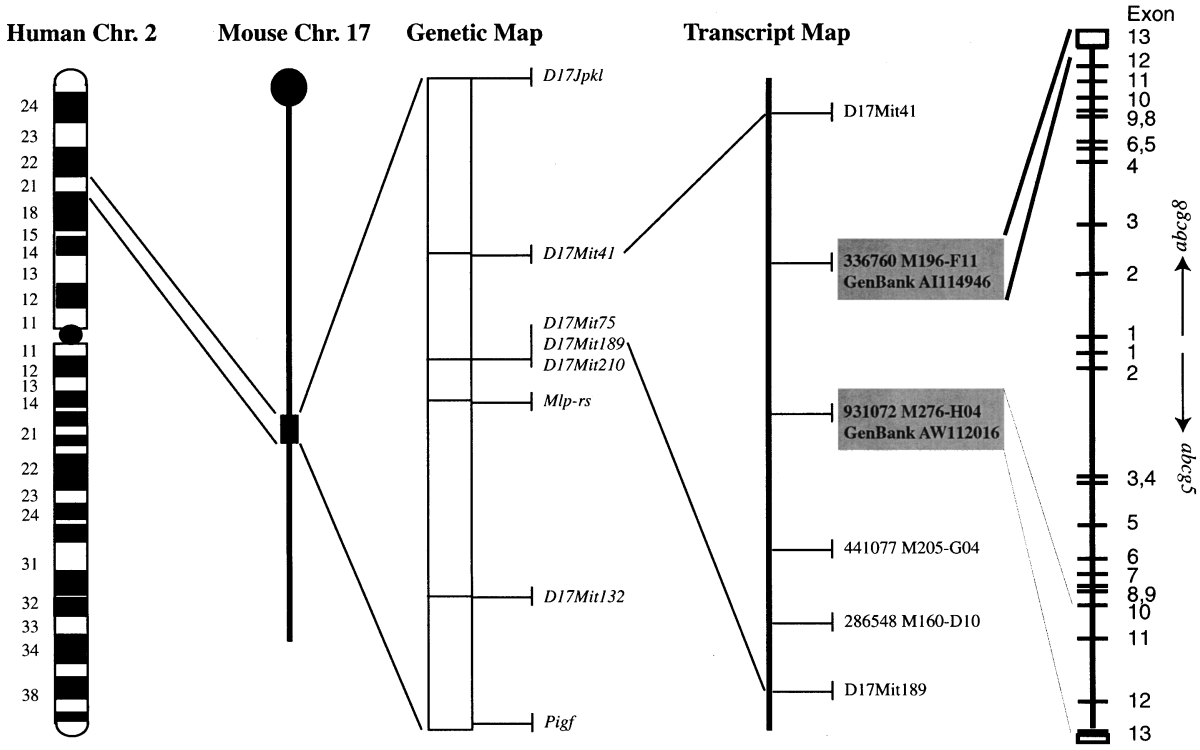


Fig. 4. Chromosomal localization of *Abcg5* and *Abcg8*. Based upon both synteny, as well as identification of expressed sequence tag (EST)s identical to *Abcg5* or *Abcg8*, these genes can be firmly placed on mouse chromosome 17, at 53 cM between microsatellite markers *D17mit41* and *D17mit189*.

Ei, MOLF/Ei, RBF/DnJ, CBA/J, CE/J, and FVB/NJ all containing some genetic variations (Table 3). Remarkably, 129/SvEv, C57BL/6J, SM/J, BALB/cJ, C3H/HeJ, AKR/J, and C57L/J were all essentially identical at these loci.

We selected strains that showed genetic variations (C57BL/6J, CAST/Ei, MOLF/Ei, RBF/DnJ, DBA/1J, DBA/2J, FVB/NJ, and CE/J) and measured their plasma sitosterol levels after a week on standard rodent chow (61 mg/kg cholesterol, 31 mg/kg campesterol, and 105 mg/kg sitosterol). Despite the large genetic variations identified between some of the strains, none of these mice showed any detectable levels of plasma sitosterol (lower limit of detection ~0.1 mg/dl). A small amount of campesterol was detectable in most samples (<1.5 mg/dl), but these did not show any significant patterns. Given the small sample size (n = 2 per group), we do not draw any conclusions about the campesterol values.

DISCUSSION

We report here the identification and characterization of the murine genes *Abcg5* and *Abcg8*, which are the homologs of human *ABCG5* and *ABCG8* genes, mutations of which are now known to cause the autosomal recessive disorder sitosterolemia (14–16). The protein products of these genes, sterolin-1 and sterolin-2 respectively, are predicted to play a crucial role in the exclusion of dietary non-cholesterol sterols from the body, as well as sterol excretion by the liver into bile. The identification of these genes as defective in sitosterolemia has given credence to the hypothesis that dietary cholesterol absorption is regulated by a molecular mechanism(s) specific for cholesterol (1). Under normal circumstances, although our normal diets consist of almost equal amounts of plant sterols and cholesterol, on average about 55% of dietary cholesterol and less than 1% of plant sterols are retained by the body (33). This selectivity of intestinal absorption appears to extend to other sterols such as shellfish sterols and this selectivity is also exhibited by the normal liver for its preferential excretion of these sterols into bile (non-cholesterol sterols > cholesterol) (5). These processes are disrupted in sitosterolemia (2, 3).

Murine *Abcg5* and *Abcg8* are located on chromosome 17 at ~53 cM, a region syntenic to human chromosome 2p21, where *ABCG5* and *ABCG8* have been mapped previously. Both the polypeptides as well as the genes show a high degree of conservation between human and mouse (14). The intron-exon organization is well preserved between the two species (data not shown) and at the protein level there is greater than 80% identity, with more than 90% conservation, suggesting that these genes play an important and highly conserved function in both species. *Abcg5* and *Abcg8* are also related to each other and appear to have arisen by a mechanism of gene duplication and inversion. Both proteins belong to the ABCG family, containing highly conserved ABC motifs, and overall organization; both proteins contain an ATP binding domain

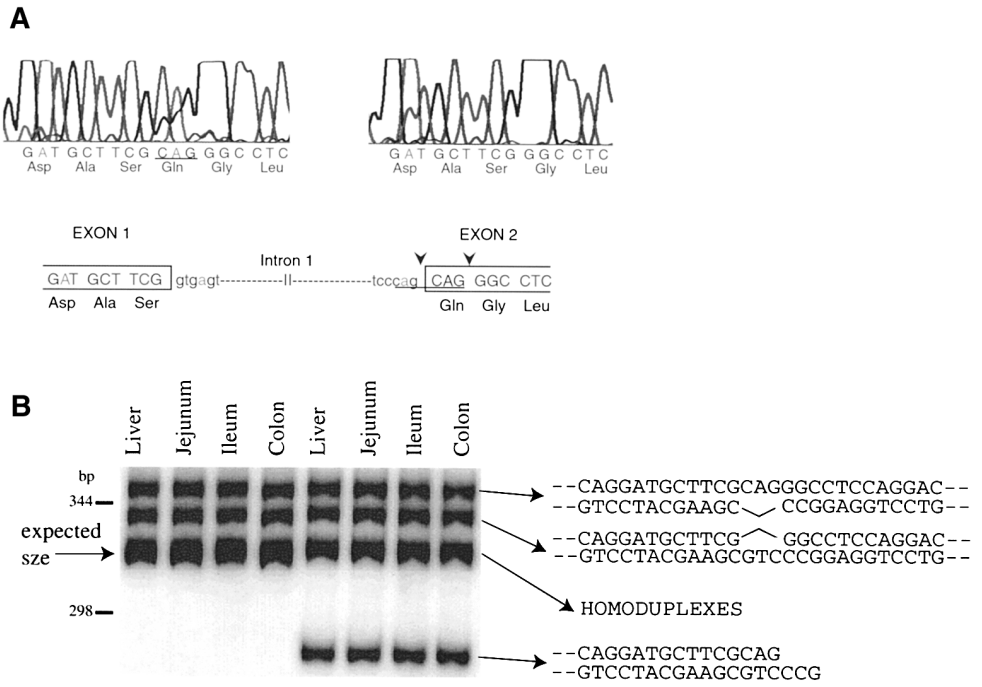


Fig. 5. Alternative splicing for Abcg8 results from a CAG repeat at splice-acceptor boundary region in intron 1. Two sets of cDNAs were identified from the library screening that differed in the presence or absence of a CAG triplet codon at the boundary of exon 1 and exon 2 in the cDNA (A, electropherograms). Examination of the exon-intron boundary suggested that two possible choices for splice-acceptor sites existed (B, underlined). To confirm that alternatively spliced forms of Abcg8 were present in RNA isolated from normal mouse liver, jejunum, ileum, or colon, we exploited the loss of *Eco*O19I recognition sequence, if the CAG was deleted. Labeled PCR products (see Materials and Methods) were digested and separated by acrylamide gel electrophoresis (C). Although a PCR product size of 325 or 322 bp was expected, two other slower bands were identified, resulting from heteroduplexes formation (confirmed by direct sequencing of these bands). About 40% of the mRNA utilized the first CAG for splice-site identification and 60% utilized the next CAG (underlined). No significant differences between the various tissues in RNA alternative splice forms was detected.

located in a larger N-terminal domain, predicted to be cytoplasmic, with a 6-transmembrane domain located at the carboxyl termini. The genes are arranged in a head-to-head configuration in the genome and are unusual as they are separated from each other by no more than 140 bases (4, 14–16). *Abcg5* and *Abcg8* are expressed in the intestine and liver only and we have previously presented genetic evidence that their two protein products, sterolin-1 and -2, may function as obligate heterodimers (4, 16). Given their proximity to each other, the lack of an obvious promoter (except for a TATA box identified in the mouse sequence only), a head-to-head configuration, and the very short distance separating them, the transcriptional control of these genes is likely to be unusual. Berge et al. identified Abcg5 and Abcg8 cDNAs as transcripts that were induced after rexinoid exposure, suggesting that liver X receptor (LXR)-retinoid X receptor may be involved in their regulation (15). Repa et al. showed that LXR deficiency affected cholesterol absorption (34). Thus LXR is a strong candidate as a regulatory transcriptional factor. If these proteins act as obligate heterodimers for normal function, they would be predicted to be transcriptionally active in the same cell and transcribed in a co-ordinate manner. Given the very short distance separating the two transcriptional start-sites, some interference of transcription would be predicted, unless transcription of each gene was temporally separated. Preliminary data suggests that the minimal region, encompassing about 358 bases between the two respective ATGs, is not sufficient to drive expression of a reporter gene when transfected in human HEK 293 cells (data not shown).

Both the human *ABCG8* and the mouse *Abcg8* exhibit alternative splicing caused by a CAGCAG repeat at a splice-acceptor site. However, for the human, this repeat occurs in intron 7 (16), whereas such a repeat is present in intron 1 for the mouse. Additionally, almost 50% of the mouse transcripts show alternative splicing, whereas only about 10% of the human ABCG8 transcripts are alternatively spliced (16). Alternative splicing results in the deletion of a single amino acid, the functional consequence of which is not known at present. Amino-acid residues that are affected by the splicing variants are not highly conserved between human, rat, and mouse, thus alternative splicing may not have any functional significance (4, 16).

TABLE 3. Polymorphisms identified *Abcg5* and *Abcg8* in 17 mouse strains

Gene	Position	Nucleotide change[a]	Effect[b]	DBA/1J	Cast/Ei	MOLF/Ei	RBF/DnJ	SPRET/Ei	CBA/J	CE/J	FVB/NJ	RIIIS/J	C58/J	129/SvFv	ARK/J	C57L/J	C3H/HeJ	C57BL/6J[c]
Abcg5	5′-UTR	−39C>T	−			✓												
	exon 1	**230C>T**	**T31M**							✓								
		237A>G	−	✓				✓										
	intron 2	IVS2−8C>T	−			✓												
		IVS2−49A>G	−								✓							
		IVS2−50A>G	−				✓			✓								
		IVS2−101T>C	−		✓		✓											
	exon 3	**454C>T**	**R106C**	✓														
		477G>C	**E113D**	✓														
		489T>C	−		✓													
	exon 4	624G>A	−		✓	✓												
	intron 4	IVS4−35C>T	−		✓	✓												
		IVS4−48G>A	−	✓														
	exon 5	684G>A	−	✓														
		691A>G	**M185V**	✓														
	intron 5	IVS5+19G>A	−	✓														
		IVS5+35A>G	−	✓														
		IVS5−24T>C	−	✓														
	intron 6	IVS6+19G>A	−	✓														
		IVS6+35A>C	−	✓														
		IVS6−29C>A	−	✓														
		IVS6−37T>G	−					✓										
		IVS6−44C>T	−					✓										
	exon 7	987C>A	−	✓				✓										
		1041T>C	−	✓				✓			✓							
	intron 7	IVS+12C>T	−	✓				✓										
		IVS7+34G>A	−	✓														
		IVS7+44C>T	−					✓										
		IVS7+66 67delT	−					✓										
		IVS7−45 56delTG	−	✓														
		IVS7−50G>A	−					✓										
		IVS7−88C>A	−					✓										
	exon 8	1095G>A	−	✓														
		1121G>C	**C328S**	✓				✓										
		1199C>T	**T354M**								✓	✓						
		1164C>T	−					✓										
		1209C>T	−	✓				✓										
		1224T>G	−		✓	✓												
		1236C>T	−					✓										
	intron 8	IVS816A>C	−					✓										
		IVS8+19C>T	−	✓														
		IVS8+23C>T	−					✓										
		IVS8+86C>G	−					✓										
		IVS8−16C>G	−					✓										
	exon 9	1359C>A	−	✓														
		1374G>A	−					✓										
		1383C>T	−	✓														
	intron 9	IVS9−13G>A	−	✓														
	exon 10	**1553T>C**	**V472A**	✓	✓	✓												
		1567G>A	**V477I**	✓														
		1581C>G	−	✓	✓	✓												
	intron 10	IVS10−77A>G	−	✓														
	exon 11	1701G>A	−	✓	✓	✓												
		1764G>A	−	✓	✓	✓												
	intron 11	IVS11+43A>G	−	✓				✓										
		IVS11+45G>A	−					✓										
		IVS11−8 15delTTTTCTCT	−					✓										
		IVS11−73 74delCT	−	✓	✓	✓												
		IVS11−149T>C	−	✓														
		IVS11−165C>T	−	✓														
		IVS11−179G>A	−	✓														
	intron 12	IVS12−24G>A	−	✓				✓										
		IVS12−34 40delGAAGTCC	−	✓														

(Continued)

TABLE 3. Polymorphisms identified *Abcg5* and *Abcg8* in 17 mouse strains

Gene	Position	Nucleotide change[a]	Effect[b]	DBA/1J	Cast/Ei	MOLF/Ei	RBF/DnJ	SPRET/Ei	CBA/J	CE/J	FVB/NJ	RIIIS/J	C58/J	129/SvEv	ARK/J	C57L/J	C3H/HeJ	C57BL/6J[c]
	exon 13	1905T>C	?Splicing					✓										
		1907G>A	**G590E**	✓	✓	✓												
		1909T>C	**S591P**					✓										
		1918T>A	**S594T**	✓	✓	✓												
		1950C>T	–					✓										
		1960C>G	**Q608E**	✓														
		1960C>A	**Q608K**		✓	✓												
		1972A>G	**K612E**					✓										
		1977C>T	–					✓										
		1988T>C	–					✓										
		2055A>G	–		✓	✓		✓										
Abcg8	5′-UTR	−67A>T	–				✓											
		−135G>A	–					✓										
		−148G>A	–					✓										
		−168C>T	–					✓										
		−212C>T	–					✓										
		−225C>A	–					✓										
	intron 1	IVS1+33C>G	–	✓														
	exon 2	277T>C	–	✓														
	intron 2	IVS2+25G>T	–	✓														
		IVS2+59A>G	–					✓										
		IVS2−44T>A	–								✓							
	exon3	284T>C	–					✓										
	intron 3	IVS3+80C>T	–	✓														
		IVS3+96G>C	–	✓														
	exon 4	449C>T	–		✓	✓												
		509T>C	–	✓				✓										
		548T>G	–					✓										
		551G>A	–					✓										
		620G>A	–	✓					✓									
	intron 4	IVS4+39A>T	–	✓														
		IVS4−22T>G	–	✓					✓									
		IVS4−24C>T	–	✓					✓									
	exon 5	683C>T	–		✓													
	intron 5	IVS5+28A>G	–						✓									
		IVS5+62G>A	–	✓														
		IVS5+80G>A	–	✓														
		IVS5−16G>A	–	✓														
		IVS5−34G>A	–	✓														
	exon 6	854T>C	–	✓	✓	✓			✓									
		991C>A	**A297E**		✓	✓												
		1058G>A	–	✓					✓									
	intron 6	IVS6+5T>C	–						✓									
		IVS6+6T>C	–	✓														
	exon7	1163A>G	–					✓										
	intron7	IVS7+25G>A	–					✓										
		IVS7−35G>A	–					✓										
		IVS7−40C>T	–					✓										
	exon8	**1254G>A**	**D385N**					✓										
		1270T>C	**V390A**					✓										
	intron 8	IVS8−18A>T	–					✓										
		IVS8−52T>C	–	✓														
	exon 9	1394T>C	–	✓		✓		✓										
		1400C>T	–					✓										
		1445A>G	–	✓	✓	✓		✓										
		1472G>A	–					✓										
	intron9	IVS9+46A>G	–					✓										
		IVS9+64T>A	–					✓										
		IVS9+79G>C	–					✓										
		IVS9+81 81delC	–					✓										
		IVS9+93T>C	–					✓										
	intron 10	IVS10+24G>A	–						✓									
		IVS10−24A>G	–	✓														
		IVS10−39A>G	–					✓										
		IVS10−43T>C	–					✓										
		IVS10−62G>A	–					✓										
		IVS10−79A>G	–	✓														

(Continued)

TABLE 3. Polymorphisms identified *Abcg5* and *Abcg8* in 17 mouse strains

Gene	Position	Nucleotide change[a]	Effect[b]	DBA/1J	Cast/Ei	MOLF/Ei	RBF/DnJ	SPRET/Ei	CBA/J	CE/J	FVB/NJ	RIIIS/J	C58/J	129/SvFv	ARK/J	C57L/J	C3H/HeJ	C57BL/6J[c]
	exon 11	**1697C>A**	**F532L**	✓			✓											
		1703C>T	−	✓			✓											
		1820T>C	−	✓			✓	✓		✓			✓					
		1835A>C	−					✓										
	intron 11	IVS11+12T>C	−					✓										
		IVS11+16 18delCCA	−					✓										
		IVS11+24C>T	−					✓										
		IVS11+27C>T	−					✓										
		IVS11−27A>G	−		✓													
	exon 12	1871C>A	−	✓														
		1883G>T	−		✓													
		1892G>C	−					✓										
		1904G>C	−					✓										
		1946A>G	−					✓										
		1950A>G	**I617V**					✓										
		1971A>G	**I624V**		✓													
	intron 12	IVS12−17C>T	−					✓										
		IVS12−28T>C	−	✓	✓	✓												
		IVS12−34C>T	−	✓														
	exon 13	**1989A>G**	**I630V**		✓			✓										
		1997C>T	−					✓										
		2045C>T	−		✓													
		2057C>T	−					✓										
		2075C>T	−		✓													
		2081A>G	−					✓										
	3′-UTR	*25G>A	−					✓										
		*29G>A	−					✓										
		*37 37delT	−	✓														
		*41G>A	−					✓										
		*65C>G	−					✓										
		*86A>G	−					✓										
		*108 110delACCinsTAG	−					✓										
		*193A>G	−					✓										
		*199G>A	−					✓										
		*208G>A	−					✓										
		*297G>A	−					✓										
		*361A>G	−					✓										
		*362C>T	−	✓														
		*412C>T	−											✓	✓	✓	✓	
		*413G>A	−					✓										
		*417T>C	−	✓				✓										
		*451C>T	−					✓										
		*461T>C	−	✓				✓										
		*510T>C	−	✓				✓										
		*513T>C	−					✓										
		*559T>A	−					✓										

[a] Numbering based on Genbank No. AF312713 (*Abcg5*) and Genbank No. AF324495 (*Abcg8*)
[b] Numbering based upon the first ATG as +1
[c] Strains C57/L, C57BL/6ByJ, BALB/c, and PERA/Ei were identical to C57BL/6J.
All sequences are wild-type, unless indicated by the check mark. Sequences in bold indicate known coding changes.

Although relatively rare, three other genes that harbor such repeats at the splice-acceptor boundaries and have been shown to lead to alternatively spliced mRNAs have been previously reported (35–37).

Finally, the murine genes are located at ~53 cM on chromosome 17, and although a large number of quantitative trait loci (QTL) for obesity, atherosclerosis, cholesterol metabolism, or lithogenic bile in mouse have been reported (28–32, 38), none of these QTLs map to this region. Thus whether genetic variation at these loci is an important determinant of any of these phenotypes remains to be determined. Direct comparison between different mouse strains has also been reported (23, 25, 26, 39). We screened 20 mouse strains and identified a very large number of polymorphic variants in both *Abcg5* and *Abcg8*. However, in a sub-set of these mice fed a standard rodent chow diet, plasma sitosterol levels were essentially undetectable (our lower limit of detection is 0.1 mg /dl). Standard rodent chow is rich in plant sterols, containing more than 1.5-fold more plant sterol relative to cholesterol. Based upon an average daily food consumption, a 25 g mouse may consume 5 g of rodent chow, representing a daily intake of ~20 mg/kg body weight of sitosterol. In comparison, an average daily diet in humans contains ~400 mg of sitosterol per day, equivalent to ~6 mg/kg body weight. Thus the rodent chow results in ~3-fold more plant sterol daily intake in the mouse relative to the human. One explanation for a lack of detection of an ef-

fect of polymorphisms in *Abcg5* or *Abcg8* on plasma sitosterol levels is that none of the variations identified have any consequences, despite some of these polymorphisms altering amino acids. These data are limited in that cholesterol, campesterol, and sitosterol absorption rates were not directly measured and none of these strains were placed on a challenge diet (such as the atherogenic or lithogenic diets). However, the information presented should allow a more direct approach in examining whether any of the genetic changes identified lead to physiological consequences. Another important observation is that some of the inbred mouse strains were highly variant, such as DBA, CAST/Ei, and MOLF/Ei, while others were almost completely identical at these loci, such as AKR/J, FVB/NJ, 129/SvEv, and C57BL/6J. The most variant strain is the DBA, a mouse that has been reported to have hyporesponsiveness to high fat and cholesterol diets (23). In this context, it is important to note that many investigators maintain their own long term mouse colonies and these animals may show genetic variations not seen in our animals obtained from the Jackson Laboratories. Nevertheless, accurate measurements of dietary sterol absorption rates, as well as biliary excretion could now be undertaken to test the hypothesis that Abcg5 and Abcg8 are involved in dietary sterol balance.

The authors are grateful to the General Clinical Research Center, Medical University of South Carolina (MUSC) and the Biomolecular Resource Center, MUSC for assistance with sequencing, and the members of the Patel lab for helpful discussions. This work was funded by a Scientist Development Award from the American Heart Association grant 9730087N (S.B.P.), the National Institutes of Health grants HL60616 (S.B.P.), MO1 RR01070-25 (MUSC GCRC), the Summer Undergraduate Program, MUSC (S.A.S.), and by an intramural award from the University Research Committee, MUSC (S.B.P.).

Manuscript received 17 October 2001 and in revised form 10 January 2002.

REFERENCES

1. Bhattacharyya, A. K., and W. E. Connor. 1974. Beta-sitosterolemia and xanthomatosis. A newly described lipid storage disease in two sisters. *J. Clin. Invest.* **53:** 1033–1043.
2. Bjorkhem, I., and K. M. Boberg. 1995. Inborn errors in bile acid biosynthesis and storage of sterols other than cholesterol. *In* The Metabolic Basis of Inherited Disease. C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle, editors. McGraw-Hill Inc., New York. 2073–2102.
3. Salen, G., S. Shefer, L. Nguyen, G. C. Ness, G. S. Tint, and V. Shore. 1992. Sitosterolemia. *J. Lipid Res.* **33:** 945–955.
4. Lee, M. H., K. Lu, and S. B. Patel. 2001. Genetic basis of sitosterolemia. *Curr. Opin. Lipid.* **12:** 141–149.
5. Gregg, R. E., W. E. Connor, D. S. Lin, and H. Brewer, Jr. 1986. Abnormal metabolism of shellfish sterols in a patient with sitosterolemia and xanthomatosis. *J. Clin. Invest.* **77:** 1864–1872.
6. Salen, G., V. Shore, G. S. Tint, T. Forte, S. Shefer, I. Horak, E. Horak, B. Dayal, L. Nguyen, A. K. Batta. 1989. Increased sitosterol absorption, decreased removal, and expanded body pools compensate for reduced cholesterol synthesis in sitosterolemia with xanthomatosis. *J. Lipid Res.* **30:** 1319–1330.
7. Bhattacharyya, A. K., W. E. Connor, D. S. Lin, M. M. McMurry, and R. S. Shulman. 1991. Sluggish sitosterol turnover and hepatic failure to excrete sitosterol into bile cause expansion of body pool of sitosterol in patients with sitosterolemia and xanthomatosis. *Arterioscler. Thromb.* **11:** 1287–1294.
8. Salen, G., E. Ahrens, Jr., and S. M. Grundy. 1970. Metabolism of beta-sitosterol in man. *J. Clin. Invest.* **49:** 952–967.
9. Nguyen, L. B., G. Salen, S. Shefer, G. S. Tint, V. Shore, and G. C. Ness. 1990. Decreased cholesterol biosynthesis in sitosterolemia with xanthomatosis: diminished mononuclear leukocyte 3-hydroxy-3-methylglutaryl coenzyme A reductase activity and enzyme protein associated with increased low-density lipoprotein receptor function. *Metab: Clin. & Exp.* **39:** 436–443.
10. Patel, S. B., A. Honda, and G. Salen. 1998. Sitosterolemia: exclusion of genes involved in reduced cholesterol biosynthesis. *J. Lipid Res.* **39:** 1055–1061.
11. Patel, S. B., G. Salen, H. Hidaka, P. O. Kwiterovich, A. F. Stalenhoef, T. A. Miettinen, S. M. Grundy, M. H. Lee, J. S. Rubenstein, M. H. Polymeropoulos, and M. J. Brownstein. 1998. Mapping a gene involved in regulating dietary cholesterol absorption. The sitosterolemia locus is found at chromosome 2p21. *J. Clin. Invest.* **102:** 1041–1044.
12. Lee, M-H., D. Gordon, J. Ott, K. Lu, L. Ose, T. Miettinen, H. Gylling, A. F. Stalenhoef, A. Pandya, H. Hidaka, J. Brewer, B. H. Kojima, N. Sakuma, R. Pegoraro, G. Salen, and S. B. Patel. 2001. Fine mapping of a gene responsible for regulating dietary cholesterol absorption; founder effects underlie cases of phytosterolemia in multiple communities. *Eur. Hum. Genet.* **9:** 375–384.
13. Lu, K., M-H. Lee, J. D. Carpten, M. Sekhon, and S. B. Patel. 2001. High-Resolution Physical and Transcript Map of Human Chromosome 2p21 Containing the Sitosterolemia Locus. *Eur. Hum. Genet.* **9:** 364–374.
14. Lee, M-H., K. Lu, S. Hazard, H. Yu, S. Shulenin, H. Hidaka, H. Kojima, R. Allikmets, N. Sakuma, R. Pegoraro, A. K. Srivastava, G. Salen, M. Dean, and S. B. Patel. 2001. Identification of a gene, ABCG5, important in the regulation of dietary cholesterol absorption. *Nat. Gen.* **27:** 79–83.
15. Berge, K. E., H. Tian, G. A. Graf, L. Yu, N. V. Grishin, J. Schultz, P. Kwiterovich, B. Shan, R. Barnes, and H. H. Hobbs. 2000. Accumulation of dietary cholesterol in sitosterolemia caused by mutations in adjacent ABC transporters. *Science.* **290:** 1771–1775.
16. Lu, K., M.-H. Lee, S. Hazard, A. Brooks-Wilson, H. Hidaka, H. Kojima, L. Ose, A. F. H. Stanlenhoef, T. Mietinnen, I. Bjorkhem, E. Brukert, A. Pandya., H. B. Brewer, Jr., G. Salen, M. Dean, A. Srivastava, and S. B. Patel. 2001. Two genes that map to the *STSL* locus cause sitosterolemia: Genomic structure and spectrum of mutations involving sterolin-1 and sterolin-2, encoded by *ABCG5* and *ABCG8* respectively. *Am. J. Hum. Genet.* **69:** 278–290.
17. Lu, K., M. Lee, and S. B. Patel. 2001. Dietary cholesterol absorption; more than just bile. *Trends Endo. Metab.* **12:** 314–320.
18. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Molec. Biol.* **215:** 403–410.
19. Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389–3402.
20. Lennon, G., C. Auffray, M. Polymeropoulos, and M. B. Soares. 1996. The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics.* **33:** 151–152.
21. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual. Laboratory Press, Cold Spring Harbor.
22. Paigen, K., and J. T. Eppig. 2000. A mouse phenome project. *Mamm. Genome.* **11:** 715–717.
23. Kirk, E. A., G. L. Moe, M. T. Caldwell, J. A. Lernmark, D. L. Wilson, and R. C. LeBoeuf. 1995. Hyper- and hypo-responsiveness to dietary fat and cholesterol among inbred mice: searching for level and variability genes. *J. Lipid Res.* **36:** 1522–1532.
24. Howles, P. N., C. P. Carter, and D. Y. Hui. 1996. Dietary free and esterified cholesterol absorption in cholesterol esterase (bile salt-stimulated lipase) gene-targeted mice. *J. Biol. Chem.* **271:** 7196–7202.
25. Carter, C. P., P. N. Howles, and D. Y. Hui. 1997. Genetic variation in cholesterol absorption efficiency among inbred strains of mice. *J. Nutr.* **127:** 1344–1348.
26. Jolley, C. D., J. M. Dietschy, and S. D. Turley. 1999. Genetic differences in cholesterol absorption in 129/Sv and C57BL/6 mice: effect on cholesterol responsiveness. *Am. J. Physiol.* **276:** G1117–G1124.
27. Salen, G., P. Kwiterovich, Jr., S. Shefer, G. S. Tint, I. Horak, V. Shore, B. Dayal, and E. Horak. 1985. Increased plasma cholestanol

and 5 alpha-saturated plant sterol derivatives in subjects with sitosterolemia and xanthomatosis. *J. Lipid Res.* **26:** 203–209.
28. Nishina, P. M., J. Wang, W. Toyofuku, F. A. Kuypers, B. Y. Ishida, and B. Paigen. 1993. Atherosclerosis and plasma and liver lipids in nine inbred strains of mice. *Lipids.* **28:** 599–605.
29. Purcell-Huynh, D. A., A. Weinreb, L. W. Castellani, M. Mehrabian, M. H. Doolittle, and A. J. Lusis. 1995. Genetic factors in lipoprotein metabolism. Analysis of a genetic cross between inbred mouse strains NZB/BINJ and SM/J using a complete linkage map approach. *J. Clin. Invest.* **96:** 1845–1858.
30. Mehrabian, M., L. W. Castellani, P. Z. Wen, J. Wong, T. Rithaporn, S. Y. Hama, G. P. Hough, D. Johnson, J. J. Albers, G. A. Mottino, J. S. Frank, M. Navab, A. M. Fogelman, and A. J. Lusis. 2000. Genetic control of HDL levels and composition in an interspecific mouse cross (CAST/Ei x C57BL/6J). *J. Lipid Res.* **41:** 1936–1946.
31. Paigen, B., N. J. Schork, K. L. Svenson, Y. C. Cheah, J. L. Mu, F. Lammert, D. Q. Wang, G. Bouchard, and M. C. Carey. 2000. Quantitative trait loci mapping for cholesterol gallstones in AKR/J and C57L/J strains of mice. *Physiol. Genomics.* **4:** 59–65.
32. Perusse, L., Y. C. Chagnon, S. J. Weisnagel, T. Rankinen, E. Snyder, J. Sands, and C. Bouchard. 2001. The human obesity gene map: the 2000 update. *Obes. Res.* **9:** 135–169.
33. Gould, R. G., R. J. Jones, G. V. LeRoy, R. W. Wissler, and C. B. Taylor. 1969. Absorbability of beta-sitosterol in humans. *Metabolism.* **18:** 652–662.
34. Repa, J. J., S. D. Turley, J. A. Lobaccaro, J. Medina, L. Li, K. Lustig, B. Shan, R. A. Heyman, J. M. Dietschy, and D. J. Mangelsdorf. 2000. Regulation of absorption and ABC1-mediated efflux of cholesterol by RXR heterodimers. *Science.* **289:** 1524–1529.
35. Yee, D., G. S. Lebovic, R. R. Marcus, and N. Rosen. 1989. Identification of an alternate type I insulin-like growth factor receptor beta subunit mRNA transcript. *J. Biol. Chem.* **264:** 21439–21441.
36. Moingeon, P., C. C. Stebbins, L. D'Adamio, J. Lucich, and E. L. Reinherz. 1990. Human natural killer cells and mature T lymphocytes express identical CD3 zeta subunits as defined by cDNA cloning and sequence analysis. *Eur. J. Immunol.* **20:** 1741–1745.
37. Rogina, B., and W. B. Upholt. 1995. The chicken homeobox gene Hoxd-11 encodes two alternatively spliced RNA species. *Biochem. Mol. Biol. Int.* **35:** 825–831.
38. Paigen, B. 1995. Genetics of responsiveness to high-fat and high-cholesterol diets in the mouse. *Am. J. Clin. Nutr.* **62:** 458S–462S.
39. Keelan, M., D. Y. Hui, G. Wild, M. T. Clandinin, and A. B. Thomson. 2000. Variability of the intestinal uptake of lipids is genetically determined in mice. *Lipids.* **35:** 833–837.