Cell. Mol. Life Sci. 63 (2006) 1564–1573
1420-682X/06/131564-10
DOI 10.1007/s00018-006-6014-x

Cellular and Molecular Life Sciences

Research Article

Unique gene structure and paralogy define the 7D-cadherin family

M. W. Wendeler [a, +], R. Jung [a, ++], H. Himmelbauer [b, *] and R. Geßner [a, *]

[a] Biomedical Research Center, Virchow Hospital of Charité Medical School Berlin, Augustenburger Platz 1, 13353 Berlin (Germany), Fax: +49 30 450 569907, e-mail: gessner@charite.de
[b] Max-Planck-Institute of Molecular Genetics, Ihnestr. 73, 14195 Berlin (Germany), Fax: +49 30 8413 1380, e-mail: himmelbauer@molgen.mpg.de

Received 16 January 2006; received after revision 18 April 2006; accepted 11 May 2006
Online First 22 June 2006

Abstract. Cadherins are Ca^{2+}-dependent transmembrane glycoproteins crucial for cell-cell adhesion in vertebrates and invertebrates. Classification of this superfamily due to their phylogenetic relationship is currently restricted to three major subfamilies: classical, desmosomal and protocadherins. Here we report evidence for a common phylogenetic origin of the kidney-specific Ksp- (*Cdh16*) and the intestine-specific LI-cadherin (*Cdh17*). Both genes consist of 18 exons and the positions of their exon-intron boundaries as well as their intron phases are perfectly conserved. We found an extensive paralogy of more than 40 megabases in mammals as well as teleost fish species encompassing the Ksp- and LI-cadherin genes. A comparable paralogy was not detected for other cadherin gene loci. These findings suggest that the Ksp- and LI-cadherin genes originated by chromosomal duplication early during vertebrate evolution and support our assumption that both proteins are paralogues within a separate cadherin family that we have termed 7D-cadherins.

Keywords. Cadherin gene superfamily, molecular evolution, genomic paralogy, zebrafish, pufferfish.

Introduction

The cadherin superfamily consists of transmembrane glycoproteins that mediate Ca^{2+}-dependent cell-cell adhesion and thereby influence cell recognition and tissue morphogenesis [1, 2]. Cadherins play an important role in embryonic development and the maintenance of the tissue-specific differentiation state [3–5]. When expressed in fibroblastoid cells, the cadherin prototype E-cadherin leads to the formation of cell-cell junctions, and induces an epithelial phenotype [6–8]. Aberrant expression and mutations of cadherins are involved in invasiveness and metastasis of tumor cells [9, 10].

The ectodomain of cadherins is characterized by a variable number of homologous cadherin repeats (EC) [11]. Each repeat consists of about 110 amino acids and contains the conserved calcium binding motifs DXD, LDRE and DXNDN. A number of X-ray diffraction and NMR-based structural analyses of single and multiple cadherin repeats of classical E-, N- and C-cadherin revealed a highly conserved folding of the cadherin repeats in a β-barrel structure. The conserved calcium binding motifs were found to form Ca^{2+}-binding pockets at the interfaces between successive cadherin repeats [12–14]. Binding of Ca^{2+} ions stabilizes the cadherin ectodomain in a rod-like conformation with all repeats stacked upon each other [15]. This conformational change is a prerequisite for the cadherin function of mediating homotypic cell-cell adhesion [16].

* Corresponding author.
+ Present address: Division Pharmacology/Neurobiology, Biozentrum, University of Basel, Klingelbergstrasse 50/70, 4056 Basel (Switzerland).
++ Present address: Schering AG, Müllerstr. 178, 13342 Berlin (Germany).

Differences in the structure of the extracellular and the cytoplasmic domain within the cadherin superfamily led to the classification of cadherins into three major and several minor subfamilies [11, 17]. Classical type cadherins, like E-cadherin, are composed of five extracellular cadherin repeats, a single transmembrane domain and a cytoplasmic domain of about 160 amino acids [18]. The highly conserved cytodomain interacts with β-catenin, which in turn binds to α-catenin [19, 20]. Until recently, α-catenin via β-catenin was assumed to link the classical cadherins to the actin cytoskeleton and to affect thus cell morphology and migration [21]. This model has now been challenged and a more dynamic role of α-catenin as a molecular switch regulating actin dynamics has been proposed [22, 23]. Another catenin, p120, influences the adhesive function of cadherins by binding directly to the juxtamembrane region of classical cadherins [24]. Desmosomal cadherins, desmocollins and desmogleins, are the adhesive components of desmosomes [25]. Like classical cadherins they consist of five extracellular repeats and a single transmembrane region, but differ in their cytoplasmic domain that links them to the intermediate filament network [26]. The protocadherins comprise the largest subfamily of cadherins with more than 60 members [27]. Their number of extracellular cadherin repeats varies from four to seven and their cytoplasmic part is different from those of other cadherin subfamilies.

The phylogenetic relationship of cadherins within the superfamily can not only be revealed by their protein structure and their biological function, but also by their particular gene organization [28]. By comparing the protein sequences of the classical E-, P-, N- and R-cadherin from different species, Gallin [29] proposed the hypothesis that today's classical cadherins are derived from a single ancestral gene that has evolved to at least four classical cadherins via gene duplication. Interestingly, the majority of the classical cadherins are clustered on human chromosome 16 [30], whereas desmosomal cadherins are found in a cluster on chromosome 18 (18q12) close to the classical N-cadherin gene (18q11.2) [31, 32]. Recently, 52 protocadherin genes were found organized in three large clusters on human chromosome 5 [33].

Ksp- and LI-cadherin are members of the cadherin superfamily that share distinct structural features different from all other cadherins [34, 35]. Both proteins consist of seven extracellular cadherin repeats and a rather short cytoplasmic domain comprising only about 25 amino acids. Their cytoplasmic domains share no homology to that of classical cadherins and appear not to bind to β-catenin [36, 37]. Neither Ksp- nor LI-cadherin contain an N-terminal propeptide, which is a characteristic feature of immature classical cadherins [38]. Interestingly, Ksp- and LI-cadherin are each coexpressed with E-cadherin and evenly distributed along the lateral contact areas of highly resorptive polarized epithelia of the kidney or the intestine, respectively [35, 39, 40]. These similar structural and functional features of Ksp- and LI-cadherin suggest a common phylogenetic origin of both molecules. To further investigate our hypothesis of a phylogenetic relationship between Ksp- and LI-cadherin, we cloned the mouse Ksp-cadherin gene (*Cdh16*), analyzed its genomic structure and compared it to the mouse LI-cadherin gene (*Cdh17*). We proceeded to determine the chromosomal localization of both genes by interspecific backcross mapping in combination with a search in the mouse genome sequence to analyze their respective genomic environment. In addition, as draft genome assemblies have become available for several teleost fish species, we sought to clarify the evolution of Ksp- and LI-cadherin genes by exploiting this information.

Material and methods

Cloning of the mouse Ksp-cadherin gene. A 129/Ola mouse cosmid library [lib: 121 from the Resource Center of the German Human Genome Project (RZPD), Berlin], cloned in Lawrist 7, was screened to isolate the mouse Ksp-cadherin gene. As probes, two fragments (385 and 489 bp) of the recently cloned mouse Ksp-cadherin cDNA [37] were generated by PCR labeling (DIG PCR Labeling Kit; Roche, Mannheim, Germany) using specific forward (5′-GCACTTCTGCCACTGTTGTGATCC-3′ and 5′-TGGAGGAGACGTGCACTAC-3′) and reverse primers (5′-CCACCTTGCGCATACATTG-3′ and 5′-TGATCCTGCTAGGTCAACAGCCAGC-3′), respectively. Probes were detected with anti-digoxigenin Fab fragments conjugated to alkaline phosphatase, and visualized by incubation with the chemiluminescent substrate CSPD® (Roche, Germany) according to the manufacturer's protocol. Three independent cosmid clones were identified containing the complete Ksp-cadherin gene sequence and obtained from the RZPD for further analysis.

DNA sequence analysis. DNA sequences were determined by the dideoxy chain termination method [41] using fluorescent dye/Big-Dye terminators in 373A and 377 automated sequencers (Applied Biosystems, Weiterstadt, Germany). Sequence alignment was performed using Mac Molly® tetra (Version 3.7, Soft Gene GmbH, Berlin, Germany). Intron-exon boundaries were identified by sequence comparison with mouse Ksp-cadherin cDNA and conserved consensus splice sites.

Interspecific backcross mapping. DNA from the inbred strain C57BL/6J (B6) was obtained from the Jackson Laboratory (Bar Harbor, ME, USA) and DNA from *Mus spretus* LS was obtained from the HGMP (Hinxton, UK). Genomic DNA from tail cuts of EUCIB (B6 × LS) × LS backcross animals was prepared following standard

procedures. Primer pairs mLIbcF1(5′-GTGGATATGGCGAAGAAGGGAAGTTCAGCGG-3′)/mLIbcR1 (5′-CTGAGTCGTTCTCCTTCCAC-3′) and mKbcF1(5′-GCACTTCTGCCACTGTTGTGATCC-3′)/mKbcF2 (5′-TGGAGGAGACGTGCACTAC-3′) were designed from genomic sequence of mouse cosmid clones containing Ksp- and LI-cadherin, respectively. Primers for the mouse *Rrad* gene Rad-F (5′-ACAAGAGTGACCTGGTGCGCT-3′) and Rad-R (5′-GAAACACAGAGGATGAACCA-3′) were designed from sequences AF084466 and AA450695, respectively. Primers for the mouse *Calb2* gene C2-F (5′-CGTGTAAAGGGGTGAAGGGA-3′) and C2-R (5′-GACAGAACATTTCCACTTTA-3′) were designed from sequence X73985. Amplification was carried out on 50 ng mouse genomic DNA with a primer concentration of 0.25 µM in 1 × PCR buffer (35 mM Tris-Base, 15 mM Tris-HCl, 50 mM KCl, 0.1% Tween 20, 1.5 mM $MgCl_2$ and 125 µM of each dNTP). B6 and *Mus spretus* LS alleles of Ksp-, LI-cadherin and Rrad were distinguished by digestion of PCR products with 5–10 U *Msp*I, *Hae*III or *Scr*FI in 1 × PCR buffer for 1 h at 37 °C. Primer pair C2-F/C2-R at the Calb2 locus detected a presence/absence polymorphism with no amplification of the *Mus spretus* allele detected at an annealing temperature of 65 °C.

Databases and data analysis. Protein sequences were analyzed with the MacMolly® Tetra software package (Version 3.7, Soft Gene GmbH). The module Align was used to calculate the amino acid identity of cadherin repeats and full-length proteins with the following parameters: pairwise local alignment, PAM 250; gap penalties: opening (–5), extending (0). Allele scores were written to text files and submitted to the MBx database for initial lod score calculation against EUCIB cross anchor markers. Fine mapping using MAPMAKER/Exp v 3.0 software [42] was carried out on downloads of complete EUCIB data sets (from http://www.hgmp.mrc.ac.uk/mbx-bin/mbx_panel) that represented mouse chromosomes 4 and 8. Data was compared with the annotated mouse genome sequence (http://www.ensembl.org). Protein sequences from annotations of known genes were downloaded from Ensembl using Ensmart (http://www.ensembl.org/EnsMart). The protein content of the following chromosomal regions was downloaded: mouse chromosome (MMU) 4, interval 1–40 Mb (198 proteins); MMU8, interval 70–140 Mb (877 proteins); MMU15, interval 1–78 Mb (644 proteins); MMU18, interval 1–65 Mb (629 proteins). Blastp homology searches for gene paralog identification was carried out using an e-value threshold of $1.0e^{-50}$. *Takifugu* and *Danio* whole genome peptide data sets were downloaded from Ensmart as well, corresponding to pufferfish data release v19.2a.2 and zebrafish v19.3.2. The *Tetraodon* whole-genome peptide data set (unpublished at the time of writing) was contributed by H. Roest-Crollius. Phylogenetic analyses were performed using ClustalX 1.83 and TreeView X (version 0.4.1).

Results

Protein structure of mouse Ksp- and LI-cadherin. Ksp- and LI-cadherin share many structural similarities, although their overall amino acid identity is less than 30%. Figure 1 shows a schematic drawing of both proteins with the average percentage of similarity on the amino acid level of corresponding cadherin repeats, transmembrane regions and cytoplasmic domains as well as the overall amino acid identity. The striking common features of both proteins are the seven extracellular cadherin repeats and the partial lack of the Ca^{2+}-binding motifs at the interface between cadherin repeats 2 and 3 (Zitt et al., submitted). Furthermore, Ksp- and LI-cadherin both exhibit two conserved cysteines in cadherin repeat 4, which are not present in classical cadherins and share a rather small cytoplasmic domain comprising only about 25 amino acids. To further investigate a potential common phylogenetic origin of Ksp- and LI-cadherin, we analyzed their gene structures.

Gene structure comparison of Ksp- and LI-cadherin. To isolate the mouse Ksp-cadherin gene, two PCR-labeled probes derived from the mouse Ksp-cadherin cDNA [37] were used for the screening of macroarrays from a mouse cosmid library. Three cosmid clones (MPMGc121-P1413Q3, -E13108Q2, and -D02293Q2) each containing the complete genomic sequence of the Ksp-cadherin gene were isolated. The Ksp-cadherin gene spans about 10 kb and contains 18 exons (Fig. 2). The position of the exon-intron boundaries were localized by comparing the genomic sequence of Ksp-cadherin with the corresponding cDNA sequences [37, 43]. When in doubt, the general rules for splice site consensus sequences were used to define their

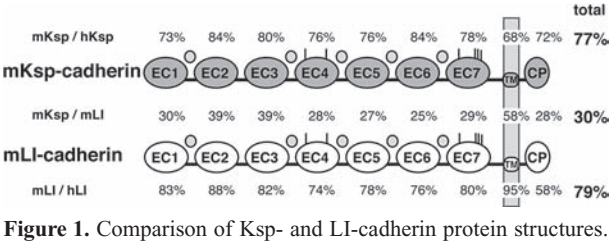

Figure 1. Comparison of Ksp- and LI-cadherin protein structures. EC1–EC7, extracellular cadherin homology repeats; TM, transmembrane domain; CP, cytoplasmic region. The small circles in between the cadherin repeats represent Ca^{2+} ions and the small vertical lines mark conserved cysteine residues. The amino acid identity of corresponding cadherin repeats and the overall identity of human and mouse Ksp- and LI-cadherin (hKsp, AAC34255; mKsp, AJ609635; hLI, NP_004054; mLI, AAD51125) were determined using the pairwise alignment with PAM 250.

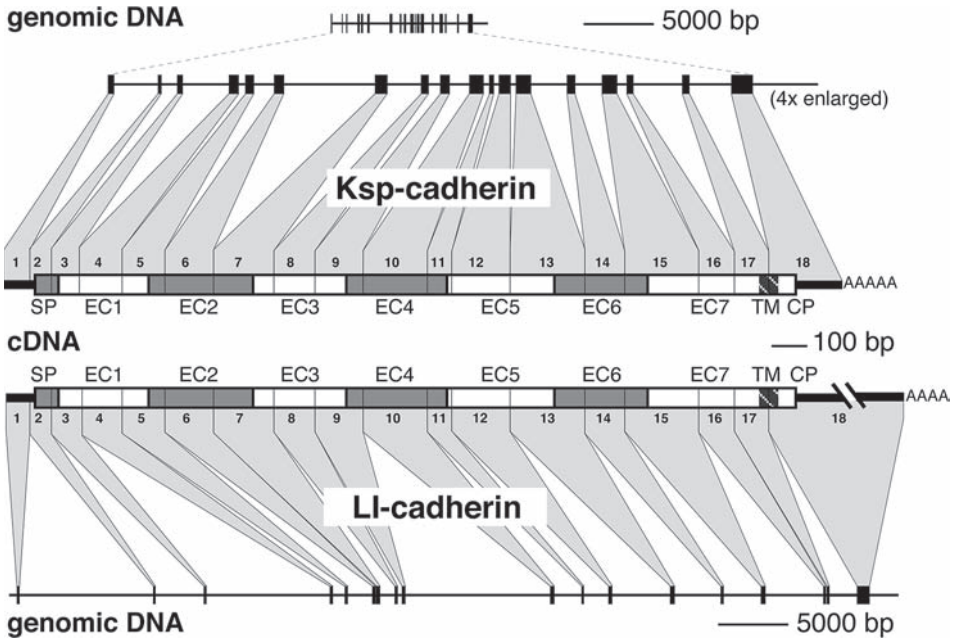

Figure 2. Cloning of the Ksp-cadherin gene and comparison with the mouse LI-cadherin gene. Three genomic clones were isolated by screening an arrayed mouse cosmid library with mouse Ksp-cadherin cDNA probes. The complete Ksp-cadherin gene spans about 10 kb of genomic DNA, and consists of 18 exons that are marked in the gene and cDNA structure. For comparison, the mouse LI-cadherin gene structure is drawn directly below. Note that the Ksp-cadherin gene is six times smaller than the LI-cadherin gene. SP, signal peptide; EC1–EC7, extracellular cadherin repeats 1–7; TM, transmembrane region; CP, cytoplasmic domain.

Table 1. Exon/intron distribution of mouse Ksp- and LI-cadherin [45] genes.

No.	Mouse Ksp-cadherin			Mouse LI-cadherin		
	Exon-exon sequence	Exon/bp	Intron/bp	Exon-exon sequence	Exon/bp	Intron/bp
1	AAG/GTC	147*	683	AAG/GAG	112	9452
2	CAG/GCT	*13* + 45	243	TTG/ACC	*22* + 48	3464
3	AAG/CTA	90	721	CAG/TTT	99	8726
4	CAG/GTC	156	69	CAG/CTT	135	924
5	CCTG/GGGTC	139	311	CCAG/GAAAG	139	1792
6	AGTG/GAAGC	159	1423	GAAG/GATCC	159	95
7	CAG/GTG	197	523	CAG/GTG	200	1094
8	GAG/TAC	123	178	TCA/CAT	132	364
9	CCAG/GAACT	151	304	TTGG/GTAAC	151	10218
10	AGTG/GCCTC	228	84	GTAG/ATTTC	216	1994
11	CAG/ATT	77	74	AAT/TAT	77	1791
12	AAG/AAC	189	77	AAG/CCT	192	4126
13	CTCAG/GTTC	242	558	GTGAG/TTAT	245	3310
14	ACAG/ATAAG	134	425	GTAG/GTGGG	131	2675
15	AACG/ATTCC	243	145	AATG/GTACA	240	4140
16	AAAG/TGATT	108	761	CCAG/TTACT	117	129
17	ATAG/GCTTC	117	652	ATTG/GTATA	114	2015
18		92 + *248*			89 + *799*	
Coding region			2490			2484
Gene			10155			59726

exact position. The observed splice donor and acceptor sites match, in general, the typical consensus sequences [44]. The mouse LI-cadherin gene [45] also consists of 18 exons but is, with nearly 59 kb, six times as large (Fig. 2). The intron positions with respect to the repeat structure are highly conserved between both molecules.

The exon/intron sizes and the splice sites of the Ksp- and LI-cadherin genes are summarized in Table 1. The smallest exon of Ksp-cadherin gene is 58 bp in size, the biggest, located at the 3′-end, has a size of 340 bp. The introns range in size between 69 and 1423 bp. The identified exons of LI-cadherin (70–888 bp) are of similar size

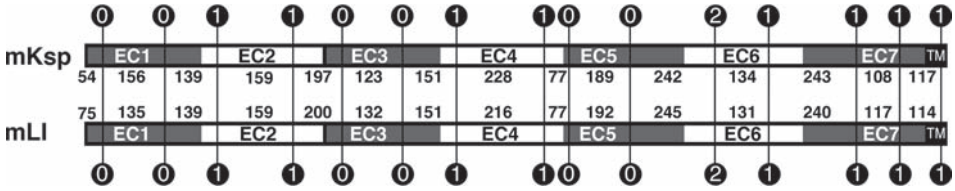

Figure 3. Gene structure comparison of mouse Ksp- and LI-cadherin. The cDNAs encoding the extracellular (EC1–EC7) and transmembrane parts (TM) of mouse Ksp- and LI-cadherin (mKsp and mLI) are aligned schematically and the relative positions of the splice sites as well as their phases (white numbers on black bullets) are shown. Both the exon-intron structure of Ksp- and LI-cadherin as well as their phases are completely conserved. The exon lengths are indicated in base pairs.

as those of Ksp-cadherin but the introns are generally much larger varying between 95 and 10 218 bp. In four cases, the exon lengths are exactly the same; six exons differ only by three nucleotides corresponding to one amino acid, and another three exons by nine nucleotides. The biggest difference is found in exon 4 with 21 nucleotides. The length of the first exon of the Ksp-cadherin gene has already been determined using RNase protection and 5′-RACE (rapid amplification of cDNA ends) approaches [46].

In Figure 3, the intron patterns of the Ksp- and LI-cadherin ectodomains and transmembrane regions are superimposed. Despite the variable intron length of both genes, their positions are highly conserved. Within the coding regions, the introns of Ksp- and LI-cadherin are at the same positions (intron phases) of coding triplets, respectively, and all differences in exon length are multiples of three (Table 1).

Chromosomal localization of mouse Ksp- and LI-cadherin genes. The chromosomal positions of the mouse Ksp- and LI-cadherin genes were determined by interspecific backcross mapping using the European Collaborative Interspecific Backcross (EUCIB). The large *Mus spretus*/B6 backcross of 982 progeny has a genetic resolution of 0.3 cM at the 95% confidence level (EUCIB, 1994; [47]). Ksp-cadherin was mapped on 115 mice of the EUCIB BSS backcross and linkage to anchor marker *D8Mit35* (lod score 17.0) was found in agreement with Thomson and colleagues [43], who used a different backcross panel. The Ksp-cadherin gene colocalized with the microsatellite marker *D8Mit185* and is positioned between *D8Mit313* (1.7 ± 1.0 cM) and *D8Mit138* (1.7 ± 1.0 cM) on mouse chromosome (MMU) 8 (Fig. 4). To genetically map the mouse LI-cadherin gene, 119 mice of the EUCIB mouse BSS backcross were genotyped and linkage to the EUCIB anchor marker *Mos* on proximal mouse chromosome 4 was detected (lod score 20.9). Fine mapping of the LI-cadherin gene resulted in a location between the markers *D4Mit101* (0.84 ± 1.0 cM) and *D4Mit292* (0.84 ± 1.0 cM).

Paralogy relationship of the regions encompassing Ksp- and LI-cadherin genes. We observed paralogy between the genomic regions that contain the loci encoding Ksp- and LI-cadherin on chromosomes 8 and 4, respectively. Initial evidence came from genetic mapping of the ras GTPase gene superfamily member *Rrad* and the calbindin gene family member *Calb2* to mouse chromosome 8 using the EUCIB cross. No recombinations were found between *Rrad, Calb2* and Ksp-cadherin (*Cdh16)* in 60 backcrossed mice typed for all three genes (data not shown). Their respective paralogs *Gem* and *Calb1* as well as LI-cadherin (*Cdh17*) are located on MMU 4 (Fig. 4 and supplementary material Table S1 on web page www.molgen.mpg.de/~rodent/projects.html). For a comprehensive identification of pairs of gene paralogs shared between MMU 4 and 8, we utilized the mouse genome annotation, as available in the Ensembl database. We downloaded the protein sequences of the known genes contained within proximal MMU4 (1–40 Mb) and distal MMU8 (70–140 Mb) and compared them using pairwise blastp analysis, applying a stringent e-value cutoff of $1.0e^{-50}$. Excluding retroviral elements and retroposons, we identified a total of 15 pairs of gene paralogs on MMU4 and MMU8. These pairs of gene paralogs are contained within a subsegment of the chromosomal regions analyzed, *i.e.* within interval 7.6–24.4 Mb on MMU4 and 79.9–122.8 Mb on MMU8 (Fig. 4). In addition to the gene pairs described above (Ksp and LI-cadherin genes *Cdh16*/*Cdh17*; calbindin genes *Calb1*/*Calb2* and GTPases *Gem*/*Rrad*), newly identified paralogs included pyruvate dehydrogenase phosphatase genes *Pdp1*/*Pdp2*, cyclinD-related proteins *Cbfa2t1h*/*Cbfa2t3h*, ATPase subunits *Atp6v0d2*/*Atp6v0d1* and other paralogs with currently ill-defined functions, *e.g.* proteins encoding a chromodomain or a KELCH domain, respectively. The observed paralogy of the chromosomal regions encompassing the mouse Ksp- and LI-cadherin genes is also conserved in rat and human (supplementary data Fig. S1 on web page www.molgen.mpg.de/~rodent/projects.html).

Chromosomal organization of other cadherin genes. Classical cadherin genes are found on several mouse chromosomes. For instance, the genes *Cdh6*, *Cdh10* and *Cdh12* map to MMU15, and *Cdh2* locates to MMU18. We compared the proteins encoded in chromosomal segments of MMU15 and MMU18 to MMU4, in the same

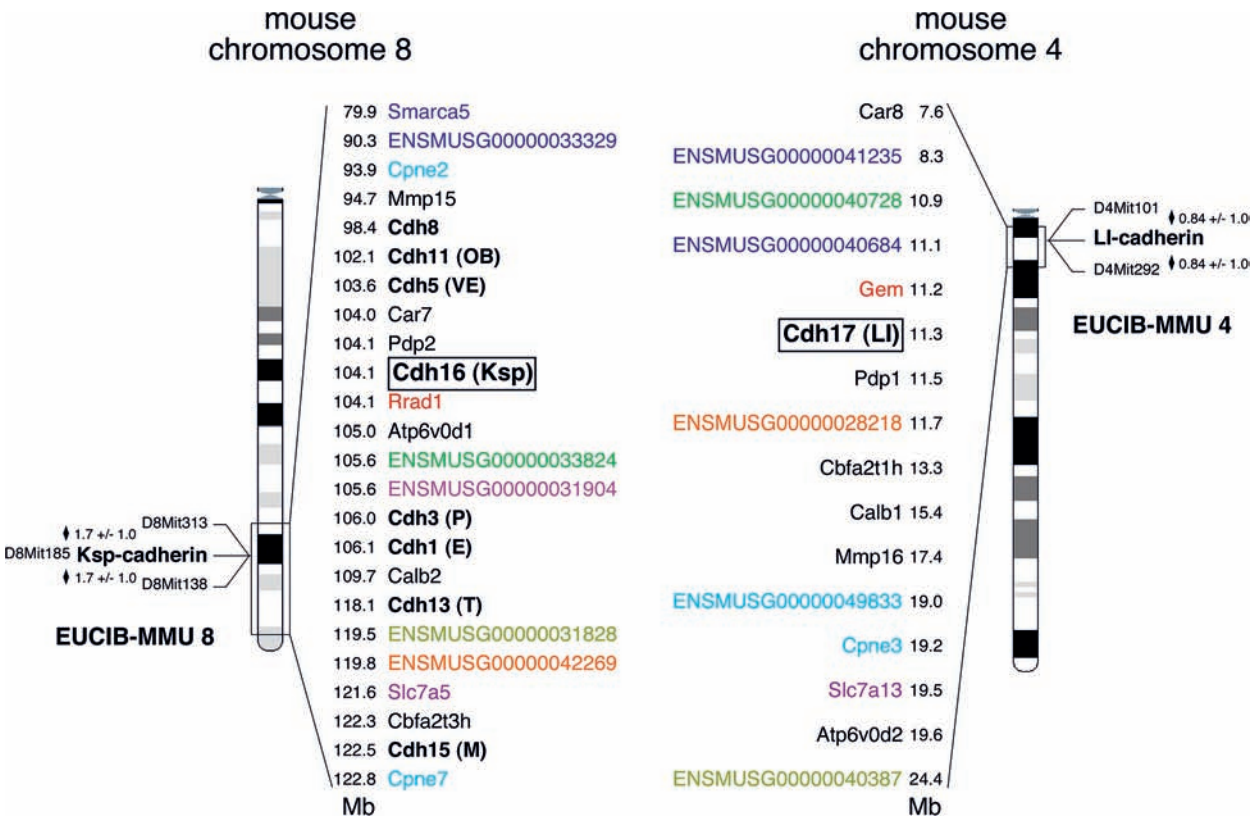

Figure 4. Chromosomal localization of the mouse Ksp- and LI-cadherin genes. The gene loci of Ksp- and LI-cadherin (*Cdh16* and *Cdh17*) on mouse chromosome 8 and 4, respectively, revealed by interspecific backcross analysis (EUCIB) and mouse genome database search are compared. Ksp- and LI-cadherin were located near the indicated microsatellite markers (recombination distance in cM) by interspecific backcross analysis on the respective mouse chromosomes. A search of the annotated mouse genome in the Ensembl database revealed that Ksp-cadherin is located within a large cluster of classical cadherin genes (*Cdh8*, *Cdh11*, *Cdh5*, *Cdh3*, *Cdh1*, *Cdh13*, and *Cdh15*), whereas LI-cadherin is located isolated on chromosome 4. However, both chromosomal regions harboring the Ksp- and LI-cadherin genes exhibit an extensive paralogy involving at least 15 gene pairs (see also supplementary data). Matching pairs of gene paralogs that cannot be related to each other due to their gene names are highlighted with the same color. The positions of the genes are indicated in megabases (Mb) as annotated by the Ensembl mouse genome database (www.ensembl.org).

way as described above for MMU8. Intriguingly, no paralogy relationship was detected, with the exception of a pair of ion transporter proteins on MMU4/MMU18 and, as expected, matching cadherin proteins. The remainders of gene hits above threshold were elements occurring in multiple copies throughout the genome (genes or pseudogenes encoding ribosomal proteins, hnRNP genes). In one case, a multi-exon gene on MMU15 that encoded poly(A)-binding protein matched a processed pseudogene on MMU4. These findings indicate that the observed extensive paralogy between the Ksp- and LI-cadherin gene loci is a particular feature of those two genes and is not common for other cadherins.

Ksp- and LI-cadherin genes in lower vertebrates. Recently, the draft genome sequences for three teleost fish species have become available, for the zebrafish (*Danio rerio*) and for two different pufferfish species, *Takifugu rubripes* [48] and *Tetraodon fluviatilis* [49]. For either species, sequence coverage of the genome is well above 95%. *Takifugu* and *Tetraodon* are closely related, at an evolutionary distance comparable to mouse and rat. The split between zebrafish and pufferfish occurred much earlier, 120–150 million years ago, at the beginning of the radiation of the modern teleost fish species. In whole-genome datasets comprising all proteins predicted from the genome sequence of either fish species, we identified the fish proteins that are homologous to the mouse gene paralogs mentioned above. These proteins, together with their orthologs in man and rodents (rat or mouse), were subjected to a phylogenetic analysis. Representative examples are shown in Figure 5a for Ksp- and LI-cadherin (*Cdh16* and *Cdh17*), classical cadherin and desmosomal cadherin genes and in Figure 5b for pyruvate dehydrogenase phosphatase genes *Pdp1* and *Pdp2*. In both examples, each of the teleost genomes contains orthologs to either *Cdh16*/*Cdh17* (Ksp-/LI-cadherin) or *Pdp1*/*Pdp2*, respectively. For *Pdp1* two orthologs were found in the zebrafish genome (Fig. 5b). Similar results were obtained for Ras GTPase superfamily genes *Rrad*/*Gem* (supplementary data Fig. S2 on web page www.molgen.mpg.de/~rodent/projects.html). For the calbindin gene

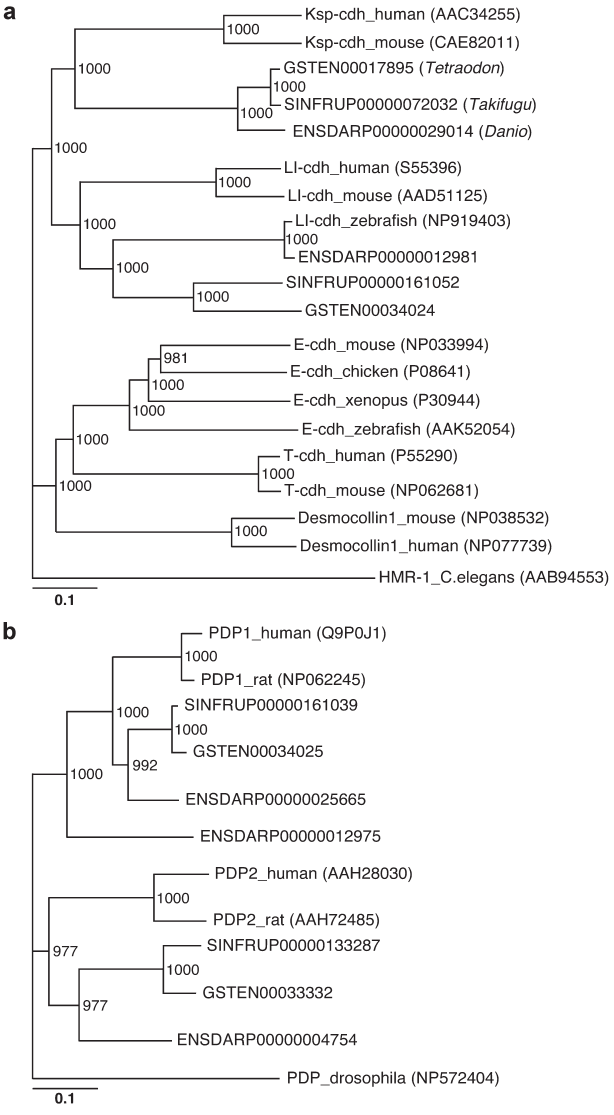

Figure 5. Phylogenetic trees (N-J) derived from the alignment of the identified fish proteins homologous to mouse Ksp- and LI-cadherin (*a*) or pyruvate dehydrogenase phosphatases PDP1 and PDP2 (*b*). Each of the identified teleost genomes of *Danio rerio*, *Takifugu rubripes* and *Tetraodon fluviatilis* contains orthologs of the Ksp-/LI-cadherin as well as of PDP1/PDP2. Homologous protein sequences from other species as indicated have been included to allow better grouping (accession nos. in brackets). For zebrafish LI-cadherin, both the recently identified sequence as deposited at the NCBI protein database (NP919403) and the LI-cadherin match of the whole genome peptide data set as was downloaded from Ensmart (ENSDARP00000012981) are shown. Bootstrap values are indicated at the branching points.

family, *Calb2* orthologs were found in all three teleost species. In contrast, only zebrafish, but neither *Fugu* nor *Tetraodon,* contained a *Calb1* ortholog, suggesting a deletion of *Calb1* in the pufferfish lineage (supplementary data, Fig. S2 on web page www.molgen.mpg.de/~rodent/projects.html). However, the combined data and tree topologies strongly suggest a duplication of the Ksp-/LI-cadherin locus before the split of the tetrapod and teleost lineages, most probably by duplication of a chromosomal region or maybe an even larger genomic entity.

Discussion

Cadherin genes have been expanded throughout the evolution of metazoans [50–52]. Within the cadherin superfamily, Ksp- and LI-cadherin share many structural features distinct from all other cadherins [53]. Furthermore, Ksp- and LI-cadherin are expressed in highly resorptive epithelia of the kidney and the intestine, respectively [34, 35]. They mediate cell-cell adhesion, induce a particular cellular phenotype and are not linked via β-catenin to the actin cytoskeleton [36, 37]. To address the question of a common phylogenetic origin of both genes indicated by protein structure comparison, we analyzed in detail the mouse Ksp- and LI-cadherin genes as well as their chromosomal environment.

Comparison of the Ksp- and LI-cadherin genomic organization revealed a complete conservation of their intron positions. There is no difference in the number of introns and their phases; the biggest variation in exon length corresponds to only 7 amino acids (exon 4). Moreover, the total length of the coding region of both molecules differs only by two amino acids, although their overall amino acid identity is below 30%. Besides the conserved structural features of both cadherins, the almost identical genomic organization is additional strong support for a common phylogenetic origin of Ksp- and LI-cadherin.

We had previously shown that the intron positions within the region of the LI-cadherin gene encoding the cadherin repeats EC3 to EC7 almost perfectly match the respective regions of the classical E-, P- and N-cadherin genes encoding EC1 to EC5 [45]. The only differences are found in the second intron of EC4, which is slightly shifted and has a different phase, and in the region encoding the premembrane repeat (EC7), which contains an additional intron. Similarly, the desmoglein 1 gene (DSG1) has all intron positions and phases within the extracellular domain conserved as compared with the genes of classical cadherins, but lacks any intron in the premembrane repeat [54].

The Ksp-cadherin gene is located within the cluster of the classical cadherins on mouse chromosome 8q (from the centromere to distal: cadherin 8, OB-, VE-, Ksp-, P-, E-, T- and M-cadherin), whereas LI-cadherin lies isolated on mouse chromosome 4 (Fig. 4). This observation is confirmed by the distribution of the human orthologs of these genes on human chromosomes 16 and 8, respectively [30, 43]. The chromosomal localization of mouse Ksp-cadherin on chromosome 8, presented in this study, is in agreement with the findings of Thomson and colleagues [43].

A detailed experimental and *in silico* analysis of the genomic neighborhood of Ksp- and LI-cadherin revealed an

extensive paralogy between those regions of mouse chromosomes 8 and 4, respectively, that includes at least 15 pairs of gene paralogs. This paralogy is also conserved in the rat and human genome [55, 56]. Thus, it is likely that a duplication event during the evolution of the vertebrate genome has led to a duplication of an ancestral Ksp-/LI-cadherin. Although Ksp-cadherin is located within a cluster of 8 classical cadherins, Ksp- and LI-cadherin comprise the only cadherin gene pair within the identified paralogy. This suggests that the expansion of classical cadherins on mouse chromosome 8 took place subsequent to the duplication of a chromosomal segment that has led to the observed paralogy.

Ksp-, LI- and E-cadherin orthologs are present in the zebrafish *Danio rerio* as well as the two puffer fish species *Takifugu rubripes* and *Tetraodon fluviatilis*. Interestingly, Ksp- and E-cadherin are also located on the same zebrafish chromosome (chromosome 7), but no obvious cluster of classical cadherins is found in the neighborhood of the Ksp-cadherin gene (ENSEMBL annotation of the zebrafish genome from 01 April 2004). Comparing the gene structures of LI- and classical cadherins, we recently proposed that both cadherin families originated from a common ancestor molecule with five cadherin repeats and that the LI-cadherin ancestor separated from the five-domain precursor by partial gene duplication [45].

Taken together, the following scenario of the Ksp-/LI-cadherin evolution seems very likely. First, an ancestral cadherin with seven cadherin repeats evolved from a five-repeat precursor cadherin by a partial gene duplication event [45]. Second, before the divergence of land vertebrates and teleost fish, a genome or segmental chromosomal duplication involving the Ksp-/LI-cadherin ancestor gene led to the origin of an additional cadherin gene within a highly paralogous environment. Indeed, two genome duplications are proposed to have happened early in vertebrate evolution [57, 58]. Third, classical cadherin genes expanded by gene duplication forming a cluster of homologous genes. Fourth, some of these genes have undergone further changes, like the generation of T-cadherin, while others have been translocated to different chromosomes.

Since the observed paralogy involving the Ksp-/LI-cadherin genes was restricted to MMU4 and MMU8 and not detected for other cadherin gene clusters on MMU15 and MMU18, it appears to be a specific feature relating those two cadherins. We therefore conclude that large segmental duplications were not involved in the generation of the current repertoire of classical cadherin genes in general.

As the Ksp- and LI-cadherin genes survived independently after the duplication of the precursor, both genes had to gain new important functions. The tissue-specific expression of today's Ksp- and LI-cadherin in highly resorptive epithelial cells of the kidney and the intestine, respectively [40, 43], might have been responsible for their survival during evolution. Interestingly, the recently identified zebrafish LI-cadherin ortholog [59] shows an expression pattern that resembles both mouse Ksp- and LI-cadherin. However, as we show in this report, zebrafish also possess a Ksp-cadherin gene (ENSDARP00000029014). Transcripts of zebrafish LI-cadherin can be detected during development in the pronephric duct and later in the mesonephric kidney, whereas in the adult zebrafish, LI-cadherin was found only in liver and intestine [59]. So far, the expression pattern of zebrafish Ksp-cadherin has not been analyzed. In mice, temporal LI-cadherin expression was also observed during embryogenesis in the epithelia of the urogenital sinus at day 13.5 post coitus [40]. This temporally overlapping expression pattern of Ksp- and LI-cadherin at certain developmental stages might be a relict of their common phylogentic origin.

The similar protein structure and genomic organization of Ksp- and LI-cadherin supports our previously suggested classification of both cadherins in a distinct subfamily that we termed, due to their seven extracellular cadherin repeats, 7D (7 domain)-cadherins.

Acknowledgements. We thank Hugues Roest-Crollius for providing us with a yet unpublished *Tetraodon* whole-genome peptide data set and Petra Krienke for perfect technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 366, TP C2).

1 Takeichi, M. (1991) Cadherin cell adhesion receptors as a morphogenetic regulator. Science 251, 1451–1455.
2 Geiger, B. and Ayalon, O. (1992) Cadherins. Annu. Rev. Cell Biol. 8, 307–332.
3 Vleminckx, K. and Kemler, R. (1999) Cadherins and tissue formation: integrating adhesion and signaling. Bioessays 21, 211–220.
4 Tepass, U., Truong, K., Godt, D., Ikura, M. and Peifer, M. (2000) Cadherins in embryonic and neural morphogenesis. Nat. Rev. Mol. Cell Biol. 1, 91–100.
5 Gumbiner, B. M. (2005) Regulation of cadherin-mediated adhesion in morphogenesis. Nat. Rev. Mol. Cell Biol. 6, 622–634.
6 Mege, R. M., Matsuzaki, F., Gallin, W. J., Goldberg, J. I., Cunningham, B. A. and Edelman, G. M. (1988) Construction of epithelioid sheets by transfection of mouse sarcoma cells with cDNAs for chicken cell adhesion molecules. Proc. Natl. Acad. Sci. USA 85, 7274–7278.
7 Matsuzaki, F., Mege, R. M., Jaffe, S. H., Friedlander, D. R., Gallin, W. J., Goldberg, J. I., Cunningham, B. A. and Edelman, G. M. (1990) cDNAs of cell adhesion molecules of different specificity induce changes in cell shape and border formation in cultured S180 cells. J. Cell Biol. 110, 1239–1252.
8 McNeill, H., Ozawa, M., Kemler, R. and Nelson, W. J. (1990) Novel function of the cell adhesion molecule uvomorulin as an inducer of cell surface polarity. Cell 62, 309–316.
9 Birchmeier, W. and Behrens, J. (1994) Cadherin expression in carcinomas: role in the formation of cell junctions and the prevention of invasiveness. Biochim. Biophys. Acta 1198, 11–26.
10 Christofori, G. and Semb, H. (1999) The role of the cell-adhesion molecule E-cadherin as a tumour-suppressor gene. Trends Biochem. Sci. 24, 73–76.
11 Nollet, F., Kools, P. and van Roy, F. (2000) Phylogenetic analysis of the cadherin superfamily allows identification of six major subfamilies besides several solitary members. J. Mol. Biol. 299, 551–572.

12 Nagar, B., Overduin, M., Ikura, M. and Rini, J. M. (1996) Structural basis of calcium-induced E-cadherin rigidification and dimerization. Nature 380 360–364.
13 Pertz, O., Bozic, D., Koch, A. W., Fauser, C., Brancaccio, A. and Engel, J. (1999) A new crystal structure, Ca^{2+} dependence and mutational analysis reveal molecular details of E-cadherin homoassociation. EMBO J. 18, 1738–1747.
14 Boggon, T. J., Murray, J., Chappuis-Flament, S., Wong, E., Gumbiner, B. M. and Shapiro, L. (2002) C-cadherin ectodomain structure and implications for cell adhesion mechanisms. Science 296, 1308–1313.
15 Pokutta, S., Herrenknecht, K., Kemler, R. and Engel, J. (1994) Conformational changes of the recombinant extracellular domain of E- cadherin upon calcium binding. Eur. J. Biochem. 223, 1019–1026.
16 Koch, A. W., Bozic, D., Pertz, O. and Engel, J. (1999) Homophilic adhesion by cadherins. Curr. Opin. Struct. Biol. 9, 275–281.
17 Pouliot, Y. (1992) Phylogenetic analysis of the cadherin superfamily. Bioessays 14, 743–748.
18 Kemler, R. (1992) Classical cadherins. Semin. Cell Biol. 3, 149–155.
19 Kemler, R. (1993) From cadherins to catenins: cytoplasmic protein interactions and regulation of cell adhesion. Trends Genet. 9, 317–321.
20 Ranscht, B. (1994) Cadherins and catenins: interactions and functions in embryonic development. Curr. Opin. Cell Biol. 6, 740–746.
21 Gumbiner, B. M. (1996) Cell adhesion: the molecular basis of tissue architecture and morphogenesis. Cell 84, 345–357.
22 Drees, F., Pokutta, S., Yamada, S., Nelson, W. J. and Weis, W. I. (2005) Alpha-catenin is a molecular switch that binds E-cadherin-beta-catenin and regulates actin-filament assembly. Cell 123, 903–915.
23 Yamada, S., Pokutta, S., Drees, F., Weis, W. I. and Nelson, W. J. (2005) Deconstructing the cadherin-catenin-actin complex. Cell 123, 889–901.
24 Reynolds, A. B., Daniel, J., McCrea, P. D., Wheelock, M. J., Wu, J. and Zhang, Z. (1994) Identification of a new catenin: the tyrosine kinase substrate p120cas associates with E-cadherin complexes. Mol. Cell Biol. 14, 8333–8342.
25 Buxton, R. S., Cowin, P., Franke, W. W., Garrod, D. R., Green, K. J., King, I. A., Koch, P. J., Magee, A. I., Rees, D. A., Stanley, J. R. et al. (1993) Nomenclature of the desmosomal cadherins. J. Cell Biol. 121, 481–483.
26 Koch, P. J., Franke, W. W., Aberle, H., Schwartz, H. and Kemler, R. (1994) Desmosomal cadherins: another growing multigene family of adhesion molecules. Curr. Opin. Cell Biol. 6, 682–687.
27 Frank, M. and Kemler, R. (2002) Protocadherins. Curr. Opin. Cell Biol. 14, 557–562.
28 Yagi, T. and Takeichi, M. (2000) Cadherin superfamily genes: functions, genomic organization, and neurologic diversity. Genes Dev. 14, 1169–1180.
29 Gallin, W. J. (1998) Evolution of the 'classical' cadherin family of cell adhesion molecules in vertebrates. Mol. Biol. Evol. 15, 1099–1107.
30 Kremmidiotis, G., Baker, E., Crawford, J., Eyre, H. J., Nahmias, J. and Callen, D. F. (1998) Localization of human cadherin genes to chromosome regions exhibiting cancer-related loss of heterozygosity. Genomics 49, 467–471.
31 Walsh, F. S., Barton, C. H., Putt, W., Moore, S. E., Kelsell, D., Spurr, N. and Goodfellow, P. N. (1990) N-cadherin gene maps to human chromosome 18 and is not linked to the E-cadherin gene. J. Neurochem. 55, 805–812.
32 Hunt, D. M., Sahota, V. K., Taylor, K., Simrak, D., Hornigold, N., Arnemann, J., Wolfe, J. and Buxton, R. S. (1999) Clustered cadherin genes: a sequence-ready contig for the desmosomal cadherin locus on human chromosome 18. Genomics 62, 445–455.
33 Wu, Q. and Maniatis, T. (1999) A striking organization of a large family of human neural cadherin-like cell adhesion genes. Cell 97, 779–790.
34 Thomson, R. B., Igarashi, P., Biemesderfer, D., Kim, R., Abu-Alfa, A., Soleimani, M. and Aronson, P. S. (1995) Isolation and cDNA cloning of Ksp-cadherin, a novel kidney-specific member of the cadherin multigene family. J. Biol. Chem. 270, 17594–17601.
35 Berndorff, D., Geßner, R., Kreft, B., Schnoy, N., Lajous-Petter, A. M., Loch, N., Reutter, W., Hortsch, M. and Tauber, R. (1994) Liver-intestine cadherin: molecular cloning and characterization of a novel Ca(2+)-dependent cell adhesion molecule expressed in liver and intestine. J. Cell Biol. 125, 1353–1369.
36 Kreft, B., Berndorff, D., Bottinger, A., Finnemann, S., Wedlich, D., Hortsch, M., Tauber, R. and Geßner, R. (1997) LI-cadherin-mediated cell-cell adhesion does not require cytoplasmic interactions. J. Cell Biol. 136, 1109–1121.
37 Wendeler, M. W., Praus, M., Jung, R., Hecking, M., Metzig, C. and Geßner, R. (2004) Ksp-cadherin is a functional cell-cell adhesion molecule related to LI-cadherin. Exp. Cell Res. 294, 345–355.
38 Ozawa, M. and Kemler, R. (1990) Correct proteolytic cleavage is required for the cell adhesive function of uvomorulin. J. Cell Biol. 111, 1645–1650.
39 Thomson, R. B. and Aronson, P. S. (1999) Immunolocalization of Ksp-cadherin in the adult and developing rabbit kidney. Am. J. Physiol. 277, F146–156.
40 Angres, B., Kim, L., Jung, R., Geßner, R. and Tauber, R. (2001) LI-cadherin gene expression during mouse intestinal development. Dev. Dyn. 221, 182–193.
41 Sanger, F., Nicklen, S. and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463–5467.
42 Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, M. J., Lincoln, S. E. and Newburg, L. (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1, 174–181.
43 Thomson, R. B., Ward, D. C., Quaggin, S. E., Igarashi, P., Muckler, Z. E. and Aronson, P. S. (1998) cDNA cloning and chromosomal localization of the human and mouse isoforms of Ksp-cadherin. Genomics 51, 445–451.
44 Mount, S. M. (1982) A catalogue of splice junction sequences. Nucleic Acids Res. 10, 459–472.
45 Jung, R., Wendeler, M. W., Danevad, M., Himmelbauer, H. and Geßner, R. (2004) Phylogenetic origin of LI-cadherin revealed by protein and gene structure analysis. Cell. Mol. Life Sci. 61, 1157–1166.
46 Whyte, D. A., Li, C., Thomson, R. B., Nix, S. L., Zanjani, R., Karp, S. L., Aronson, P. S. and Igarashi, P. (1999) Ksp-cadherin gene promoter. I. Characterization and renal epithelial cell-specific activity. Am. J. Physiol. 277, F587–F598.
47 Rhodes, M., Straw, R., Fernando, S., Evans, A., Lacey, T., Dearlove, A., Greystrong, J., Walker, J., Watson, P., Weston, P. et al. (1998) A high-resolution microsatellite map of the mouse genome. Genome Res. 8, 531–542.
48 Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J. M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A. et al. (2002) Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. Science 297, 1301–1310.
49 Jaillon, O., Aury, J. M., Brunet, F., Petit, J. L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A. et al. (2004) Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate protokaryotype. Nature 431, 946–957.
50 Hynes, R. O. and Zhao, Q. (2000) The evolution of cell adhesion. J. Cell Biol. 150, F89–96.
51 Hutter, H., Vogel, B. E., Plenefisch, J. D., Norris, C. R., Proenca, R. B., Spieth, J., Guo, C., Mastwal, S., Zhu, X., Scheel, J.

and Hedgecock, E. M. (2000) Conservation and novelty in the evolution of cell adhesion and extracellular matrix genes. Science 287, 989–994.
52 Hill, E., Broadbent, I. D., Chothia, C. and Pettitt, J. (2001) Cadherin superfamily proteins in *Caenorhabditis elegans* and *Drosophila melanogaster*. J. Mol. Biol. 305, 1011–1024.
53 Geßner, R. and Tauber, R. (2000) Intestinal cell adhesion molecules. Liver-intestine cadherin. Ann. N. Y. Acad. Sci. 915, 136–143.
54 Puttagunta, S., Mathur, M. and Cowin, P. (1994) Structure of DSG1, the bovine desmosomal cadherin gene encoding the pemphigus foliaceus antigen. Evidence of polymorphism. J. Biol. Chem. 269, 1949–1955.
55 Parmentier, M., Passage, E., Vassart, G. and Mattei, M. G. (1991) The human calbindin D28k (CALB1) and calretinin (CALB2) genes are located at 8q21.3----q22.1 and 16q22----q23, respectively, suggesting a common duplication with the carbonic anhydrase isozyme loci. Cytogenet. Cell. Genet. 57, 41–43.
56 Ahn, A. H., Freener, C. A., Gussoni, E., Yoshida, M., Ozawa, E. and Kunkel, L. M. (1996) The three human syntrophin genes are expressed in diverse tissues, have distinct chromosomal locations, and each bind to dystrophin and its relatives. J. Biol. Chem. 271, 2724–2730.
57 Ohno, S. (1993) Patterns in genome evolution. Curr. Opin. Genet. Dev. 3, 911–914.
58 Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A. J., Herwig, R., Vingron, M. and Lehrach, H. (2003) New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes. Genome Res. 13, 1056–1066.
59 Horsfield, J., Ramachandran, A., Reuter, K., LaVallie, E., Collins-Racie, L., Crosier, K. and Crosier, P. (2002) Cadherin-17 is required to maintain pronephric duct integrity during zebrafish development. Mech. Dev. 115, 15–26.