Cell. Mol. Life Sci. 63 (2006) 1125–1140
1420-682X/06/101125-16
DOI 10.1007/s00018-005-5571-8


Cellular and Molecular Life Sciences

# Review

# The vertebrate connexin family

**V. Cruciani and S.-O. Mikalsen***

Institute for Cancer Research, The Norwegian Radium Hospital, 0310 Oslo (Norway), Fax: + 47 22935767,
e-mail: sveinole@ulrik.uio.no

Received 1 December 2005; received after revision 8 January 2005; accepted 31 January 2006
Online First 29 March 2006

**Abstract.** Connexins are chordate-specific transmembrane proteins that can form gap junctional channels between adjacent cells. With the progress in vertebrate genome sequencing, it is now possible to reconstruct the main lines in the evolution of the connexin family from fishes to mammals. Four connexin groups are only found in fishes. Otherwise, the differences between fishes and mammals can be explained by two gene losses (Cx39.9 and Cx43.4) after the divergence of the Reptilia, and three gene duplications (the generation of Cx26 and 30 from a preCx26/30 sequence, Cx30.3 and 31.1 from a preCx30.3/31.1 sequence, and Cx31.3 from an uncertain origin). Orthologs of most connexins can be found throughout the vertebrates from fishes to mammals. As judged from the recently defined connexins in tunicates, the original connexin might be related to the ortholog groups of Cx36, 39.2, 43.4, 45 or 47.

**Keywords.** Gap junction, pannexin, connexin, evolution, phylogeny, synteny, chromosome.

## Introduction

Gap junction plaques are clusters of transmembrane channels that allow direct contact between the cytoplasm of one cell and the cytoplasm of its neighbor. Each cell participates with one hemichannel. In vertebrates, the hemichannel is called a connexon, and each connexon is made of six protein subunits named connexins. Due to these junctions, all the cells in a communicating tissue potentially share a pool of small molecules (<1000 Da) and metabolites, such as nucleotides, amino acids, smaller saccharides and second messengers. They may also transmit other regulatory molecules of unknown identity. In excitable cells, including cardiac muscle and neurons, gap junction channels allow the transmission of electrical impulses. Approximately 80 years ago, the observation of dye transfer between stomach epithelial cells was probably the first indication of gap junctional intercellular communication (Schmidtmann [1], cited in Shibata et al. [2]), but the implications were not understood at the time. More than 30 years later, electrical conductance between neurons of crayfish was observed [3]. In retrospect, this showed the presence of electrical synapses in these cells [4]. In the same period, gap junction structures were first observed by electron microscopy [5–7]. In 1964, Loewenstein and Kanno [8, 9] discovered that a sizeable hydrophilic molecule (fluorescein) could pass from cell to cell. Soon thereafter, several other hydrophilic molecules ranging from 300 to 1000 Da were shown to pass between adjacent cells [10].

The sequences of the connexins started to emerge in the last half of the 1980s, human and rat Cx32 being the first [11, 12]. They were rapidly followed by rat Cx43 [13] and Cx26 [14]. Most of the remaining human and rodent connexins were sequenced during the 1990s and the first couple of years of the new century. The first connexins identified from nonmammalian vertebrates were Cx30 [15] and Cx38 [16] from *Xenopus laevis*. In chicken, Cx43 was the first to be cloned [17], closely followed by Cx42 and Cx45 [18]. In 1994, the sequences of the first bony fish connexins, Cx32.2 and Cx32.7 from Atlantic croaker [19], were published, soon followed by Cx35

* Corresponding author.

from skate [20], a cartilaginous fish. Thus, the presence of connexins in representatives of the major vertebrate groups had been firmly established. The number of potentially functional connexins in mammals has been established as 19–21 [21, 22], with a possibility of reaching 22 in opossum as the Cx59 ortholog has not yet been identified in this species [22]. A recent overview of *Xenopus* connexins identified 10 sequences [23], but we here increase that number considerably. Furthermore, until now, the number of connexins in chicken and fishes has not been clear, because no full overview for these groups had been presented before the submission of this review.
Although a number of nonvertebrate systems have also shown the property of direct cell-to-cell communication [8, 10, 24] and/or the presence of gap junctions [7, 25], no connexins have been found in such systems but, rather, the pannexin/innexin family seems to play a connexin-like role [26, 27]. Pannexin-like genes are also found in humans [28, 29]. This review will not discuss pannexins, and the reader is referred elsewhere for recent updates [4, 23, 30, 31].
The connexins are important in development, differentiation and growth control [32, 33]. This is also shown by the wide variety of diseases where mutations in connexin genes or aberrant regulation of connexins have been implicated [34–37]. A number of signal transduction pathways have connexins as their targets, either by the direct phosphorylation of the connexin proteins or by the regulation of connexin gene expression [38]. More information on the physiology and potential functions of connexins can be found in several recent reviews [4, 38–47]. We will present here an overview of some evolutionary aspects for the connexin family. In addition to providing general knowledge, an evolutionary analysis might be useful for understanding connexin-related diseases. For example, there are many similarities in the expression of mammalian Cx46 and 50 and their zebrafish orthologs (the same connexin in different species; see Fig. 1) [48]. These two connexins are associated with two different types of cataract in humans and mouse [49–52]. Other connexins show differences in expression even within the mammals [53, 54]. Analyzing the identities, similarities and differences in functions and sequences may help to define the parts of the proteins that are likely to be of decisive importance in their function in humans. Changes in amino acids in such important parts of the proteins may have more serious consequences, and knowledge about their molecular evolution may therefore promote the understanding of patterns of human disease-causing mutations [55, 56].

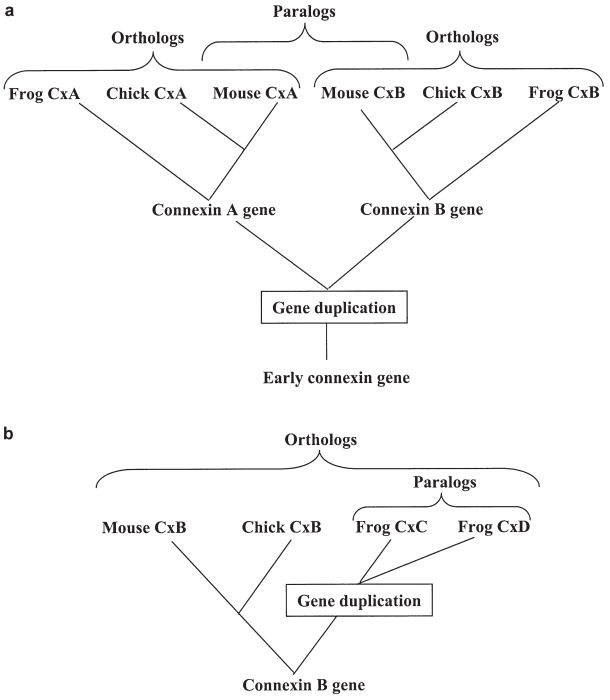


**Figure 1.** The concepts of orthology and paralogy. (*a*) The paralogs are the genes of a gene family within a single species. The orthologs can be considered as the same gene in different species; thus orthologs derive from the last common ancestral gene that was present before the speciation event took place. Modified from http://www.ncbi.nlm.gov/Education/BLASTinfo/Orthology.html. (*b*) The generation of 1:2 orthology. A gene duplication takes place in one of the evolutionary branches (in this case, the amphibian branch), but not in the other. In this case, mouse and chicken CxB are the orthologs of frog CxC and CxD, despite the fact that frog CxC and CxD are paralogs. It is important to note that the concept of orthology concerns only the genetic family tree, and does not imply functional equivalence [104].

## Conservation of sequences and gene structure

Following the cloning of the first few connexins, it became clear that certain parts of the connexin sequences were more conserved than others [57, 58], and this still holds true now that probably all functional human and mouse connexins have been identified [21, 35, 59]. The connexins in these two species have established the pattern to which connexins in all other species are compared. The connexins are integral membrane proteins with four transmembrane (TM) domains, two extracellular (EC) loops, an intracellular loop and cytoplasmic N- and C-terminal ends [60, 61]. The two EC loops have three cysteines each that are spaced in a specific manner. Most connexins display a pattern of $CX_6CX_3C$ in EC1, and $CX_4CX_5C$ in EC2. The single exception is Cx31 with a $CX_5CX_5C$ spacing in EC2 [62, 63]. Many positions in the N terminus, TM domains and EC loops also show a strong degree of conservation among a number of paralogs (the different connexins within one species; see Fig. 1), as illustrated in Figure 2a. In fact, these are the characteristics that are used to define the connexin family. On the other hand, the intracellular loop and the C-terminal tail show considerable differences among a number of orthologs, as illustrated in Figure 2b. Naturally, the ex-

**a**

```
HsCx26   MDWGTLQTILGGVNKHSTSIGKIWLTVLFIFRIMILVVAAKEVWGDEQADFVCNTLQPGC
HsCx30   MDWGTLHTFIGGVNKHSTSIGKVWITVIFIFRVMILVVAAQEVWGDEQEDFVCNTLQPGC
         ******:*::************:*:**:****:*******:******* ***********

HsCx26   KNVCYDHYFPISHIRLWALQLIFVSSPALLVAMHVAY
HsCx30   KNVCYDHFFPVSHIRLWALQLIFVSTPALLVAMHVAY
         *******:**:**************:***********
```

**b**

```
HsCx46   KKLKQGVTSRLGPDASEAPLGTADPPPLPPSSRPPAVAIGFPPYYAHTAAPLGQARAVGY
MmCx46   KKLKQGVTNHFNPDASEARHKPLDPLPAATSSGPPSVSIGFPPYYTHPACPTVQGKAIGF
         ********.::.******   . ** * ..** **:*:*******:*.*.*  *.:*:*:

HsCx46   PGAPPPAADFKMLALTEARGKGQSAKLYNGHHHLLMTEQNWANQAAERQPPALKAYPAAS
MmCx46   PGAPLSPADFTVVTLNDAQGRNHPVKHCNGHH--LTTEQNWTRQVAEQQTPASKPSSAAS
         **** ..***.:::*.:*:*:.:..*  ****  * *****:.*.**:*.** *. .***

HsCx46   TPAAPSPVGSSSPPLAHEAEAGAAPLLLDGSGSSLEGSALAGTPEEEEQAVTTAAQMHQP
MmCx46   SPDGRKG------------------LIDSSGSSLQESALVVTPEEGEQALATTVEMYSP
         :* . .                    *:*.*****: ***. **** ***::*:.:*:.*

HsCx46   PLPLGDPGRASKASRASSGRARPEDLAI
MmCx46   PLVLLDPGRSSKSS---NGRARPGDLAI
         ** * ****:**:*   .***** ****
```

**Figure 2.** Similarities among paralogs and differences between orthologs. (*a*) Alignment of the N terminus, TM1, EC1 and TM2 of human Cx26 and Cx30. A high degree of identity is found between the two paralogs. The TM domains are indicated by shading. (*b*) Alignment of the C-terminal tails of human Cx46 (HsCx46) and mouse Cx46 (MmCx46). A considerable degree of difference is found for the two orthologs.

tent of difference increases when comparing species with a greater evolutionary distance. Thus, if the conserved domains of the connexins (N terminus, TM1, EC1, TM2, TM3, EC2, TM4) are used as 'virtual probes', investigating the genomes of species that are spaced over larger evolutionary distances should be possible.

Individual laboratories continue to add connexin sequences from specific species in which they are interested, steadily increasing the total number of available sequences [53, 64, 65]. However, concerted genome sequencing takes place for a number of species, and some of these genomes are available in the Ensembl databases (http://www.ensembl.org), and more are being added. This is, therefore, a good source for *in silico* investigations on gene families. Thanks to the relatively simple gene structure of the connexins, this can be done without advanced gene prediction programs, and allows the analysis of connexins in genomes that have not yet become annotated. The assumption is that the gene structure has been largely conserved in evolution. The majority of connexins have an unusual gene structure, with the complete coding part of the sequence in one exon [21, 59]. Skate Cx35 was the first connexin identified that does not follow this pattern, as an intron is inserted at position 71 of the reading frame [20]. The cloning of the mammalian ortholog, Cx36, showed the presence of an intron at exactly the same position [66, 67]. Later, rodent Cx39/human Cx40.1 were also shown to have an intron in the same area [68]. Mouse Cx57 was recently shown to possess an intron at the very 3′ end of the reading frame [69], and this might extend to other mammalian orthologs, including human Cx62 [22]. Cx31.3 orthologs may also have an intron in the 3′-most part of the reading frame [21, 22]. Some fish connexins have several introns, and nearly all tunicate connexins have three to eight introns in their coding sequences [70]. This may suggest that the early connexins had introns, but most of them were lost at some point between urochordate and teleost (bony fish) divergences. The loss of introns may have occurred by retrotranscription of the processed mRNA, followed by reinsertion into the genome [71, 72]. If the reinsertion occurred by homologous recombination, the old version of the gene was lost, and replaced by an intronless version. If the reinsertion occurred elsewhere in the genome, a new connexin gene might be created, although the new sequence would probably more often become a pseudogene.

## Differences in sequences and the puzzlement

A recent review pointed out that some of the *Xenopus* connexins bridged the gap between mammalian and fish connexins [23]. On the other hand, '… for many mammalian connexins no genuine orthologs can easily be nominated in birds, fish, amphibians or tunicates on the basis of sequence characteristics only. It has become clear

that every class of vertebrates uses their own unique set of connexins to build their gap junctions' [23]. We think this citation describes the puzzlement that many researchers (for ourselves, we could even say confusion) have felt when looking at the different and ever-increasing number of sequences from an increasing number of species, now also including the nonvertebrate group of tunicates [70, 73].

A potential reason for the puzzlement could come from the fact that some of the connexins, like orthologs of Cx36 and 43, also have a high degree of conservation in the variable domains, and can therefore easily be recognized across the vertebrates [66, 67, 74, 75]. From this observation, one might assume that this should also be valid for other connexins. However, orthologs are under a wide range of selection pressures [76], and thus the degree of sequence similarity may also vary to a quite wide extent. Many of the connexins from chicken, *Xenopus* or fishes show considerable differences to all of the human and mouse connexins [23]. Zebrafish Cx52.6 and Cx55.5 have little resemblance to known connexins from other vertebrates, and it was suggested that they may be without orthologs in higher vertebrates, although some similarities to human Cx59 (also called Cx58) and Cx62/Cx57 were indicated [64, 74]. Zebrafish Cx27.5 showed a relatively high homology to mammalian Cx26 and 32 in the TM and EC domains, but no expression was found in the liver [74]. In addition to the intrinsic difficulties due to the sequence differences, several nonmammalian connexins show the highest similarity to one and the same mammalian sequence. The archetypical *Xenopus* Cx38 has some similarity with mammalian Cx37 [77], but *Xenopus* Cx41 has an even higher similarity to Cx37 [78]. Perch has two sequences very similar to mammalian Cx36, called Cx35 and Cx34.7 [79]. Atlantic croaker Cx32.2 and Cx32.7 and zebrafish Cx32.2 have their highest resemblance to chicken Cx56 and mammalian Cx46 [19, 80]. Furthermore, zebrafish Cx48.5 has a high sequence similarity with mammalian Cx46, but some of its properties are more comparable to mammalian Cx50 [65], which, on the other hand, has a higher sequence homology to zebrafish Cx44.1 [74].

Luckily, some order among all the sequences can be found. The major steps are: (i) The use of genome databases including mammals, chicken, *Xenopus* and bony fishes. The inclusion of the major groups of vertebrates may make bridging the gap [23] over the evolutionary distances easier. (ii) The extraction and interpretation of information from these genome databases, yielding the majority of connexins in each species. As indicated in Figure 2A, a number of paralogs have sequences that are closely related. It is therefore important to make the connexin family as complete as possible in each of the species to categorize the sequences correctly. Also important is to overrule computerized gene recognitions, if necessary, as they may be wrong in a number of cases, especially in the less studied genomes. One should be aware of the potential presence of both sequencing and assembly errors. Additionally, some of the most recently added genomes in the Ensembl databases may not contain computerized gene suggestions. When the knowledge on connexin sequences described above is combined, we are looking for the following pattern: (high similarity to the N terminus, TM1, EC1, TM2) – (intracellular loop with low expectancy of similarity) – (high similarity to TM3, EC2, TM4) – (C-terminal tail with unknown number of codons and low expectancy of similarity), all in one continuous reading frame. From these data, the conserved domains are extracted since they are easier to follow over larger evolutionary distances. In fact, the variable domains among orthologs might be so different that they yield very little information for species spread wide apart in evolution, and they may rather introduce 'noise' in the construction of a phylogenetic tree (e.g., creating longer branch lengths). The variable domains are of course valuable for comparison of orthologs in closely related species, but are best used as a supplementary criterion for the large-scale approach that we are discussing here (see our comments on Cx37 below). Additionally, there might be a statistical advantage in comparing sequences of similar lengths (see e.g. Pang et al. [81] and our comments on Cx59 and 62 below). (iii) The phylogenetic analysis of this information. This is essential when more than a few sequences are compared. A BLAST search automatically gives better alignments of the more conserved sequences, but it is the subsequent phylogenetic analyses that allow one to categorize the derived sequences into the different orthologous groups. There are many methods and models for phylogenetic analyses, with different underlying assumptions (e.g. nucleotide composition, transition and transversion ratios of the mutations, pattern of amino acid changes caused by the mutations). For the connexins, we obtained quite consistent results using a number of different models.

The single most important connection between these steps is to define the conserved domains in the connexins as described above.

## The easy case: mammals

Quite early on, with only a few connexin sequences established, it became clear that the connexin family contained at least two subfamilies, which were called group II or $\alpha$ connexins and group I or $\beta$ connexins [58, 82, 83]. As the number of available sequences increased, it became evident that Cx36-like genes formed a third group, at that time called $\gamma$, and that Cx45-like genes occupied a position outside all three groups [79], although somewhat closer to the $\gamma$ group. This is confirmed by our phyloge-

netic analysis (Fig. 3), and we suggest that they are called group IIIa (which includes Cx36, 31.9, 40.1, 39.2) and group IIIb (Cx45, 47, 43.4, 31.3).

Thanks to genome sequencing, human and mouse were the first two species where the potentially full complement of functional connexins was established [21, 35]. As listed by Söhl and Willecke [21], the mouse genome contains 20 members, and the human genome 21 members. This includes Cx23, a predicted gene that deviates from other connexins both in gene structure and cysteine pattern in the extracellular loops [21]. We have not included Cx23 in our analyses. Thus, we are left with 19 and 20

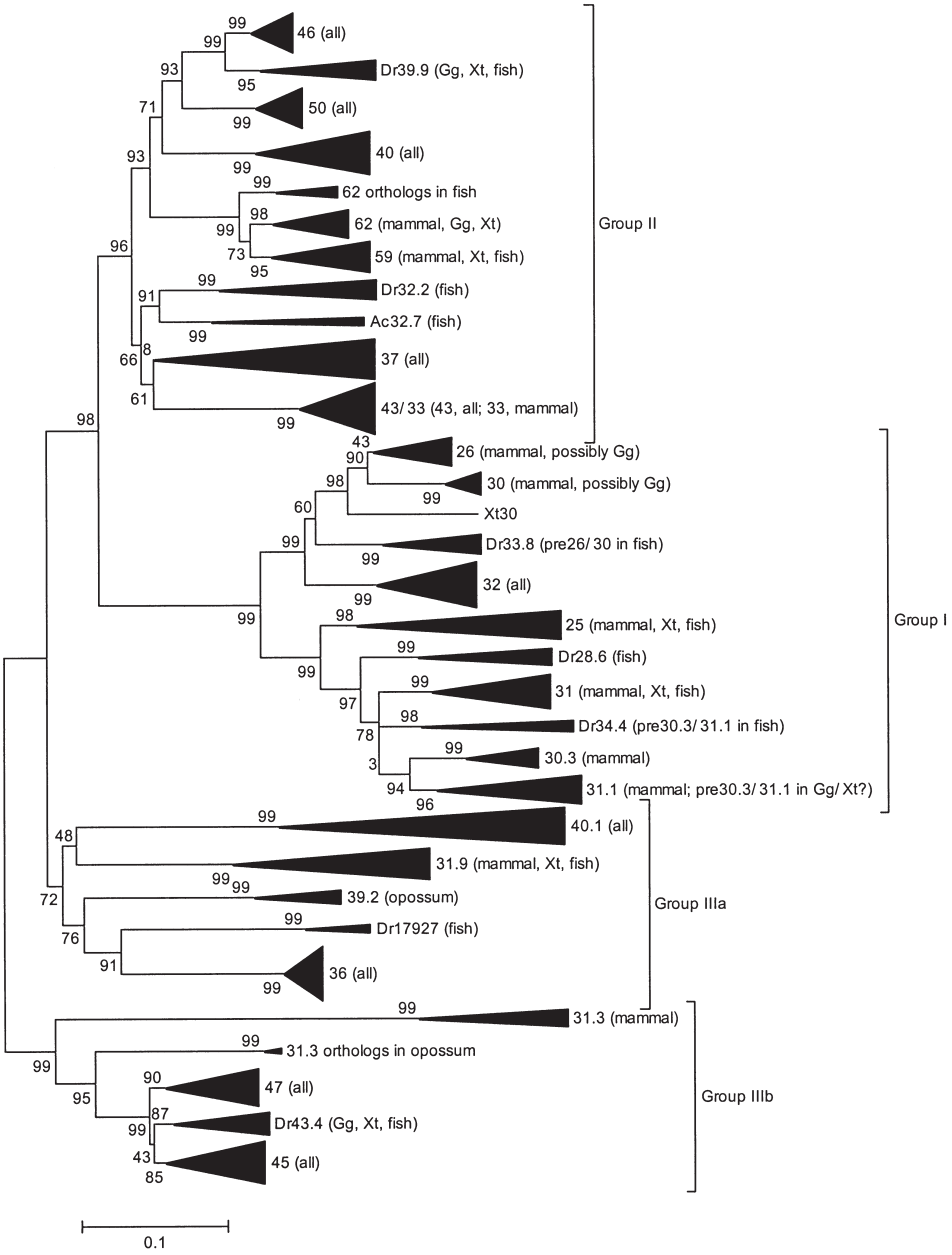


**Figure 3.** Phylogenetic tree of vertebrate connexins. The tree is based on the conserved sequences of 303 connexins from human, chimpanzee, mouse, rat, dog, cow, elephant, opossum, chicken (Gg), *Xenopus tropicalis* (Xt), zebrafish (Dr), *Fugu rubripes* and *Tetraodon nigroviridis*. The program MEGA3 was used (http://www.megasoftware.net [105]). The tree was constructed by the minimum evolution method with the Tamura 3-parameter model, using the first and second positions of the codons. Although not exactly identical, the tree is very similar when constructed on the basis of the amino acids in the conserved domains (not shown). The widths of the triangles indicate the number of sequences in the group, the lengths of the triangles indicate the variability of the included sequences. In parentheses are the evolutionary branches in which the included sequences are found. All: found in mammals, chicken, *Xenopus* and fishes. Ac32.7 indicates the orthology group where Cx32.7 from Atlantic croaker would be found. Dr17927 is an abbreviation of the Ensembl gene identification ENSDARG00000017927. The scale bar at lower left indicates the number of nucleotide substitutions per site (i.e. the scale bar indicates the length of the branch that corresponds to 10% difference in nucleotides). Interior branch statistics (500 replications) are shown at the nodes.

members in mouse and human, respectively. Among these connexins, 18 are found in both species and are considered as ortholog pairs.

The single mouse sequence not found in humans is Cx33, which is located on the X chromosome. Cx33 has a relatively high resemblance to Cx43, and belongs to the $\alpha$ group of connexins. The only other connexin found on the human or mouse X chromosome is Cx32, a $\beta$ group member. On the other hand, two human connexins, Cx25 and Cx59, were without orthologs in mouse. Thus, a few connexins were considered as species specific [21, 59]. Connexins from rat, Chinese hamster and Syrian hamster (19, 16 and 17 partial or full-length sequences, respectively) seemed to confirm the structure of the connexin family established in human and mouse [53]. Other mammals had only a few scattered full-length connexin sequences available in gene and protein databases, like GenBank or SwissProt, but none of these contradicted the connexin family structure.

The Ensembl genome databases contain the current status of genome sequences for a number of mammals that have not previously been included in a comparison of the connexin family. Our analyses included the human, chimpanzee, mouse, rat, dog, cow and opossum genomes [22]. The latter species is perhaps the most interesting among the analyzed mammals, as it diverged from the remaining mammalian branch around 170 million years ago, while most of the other main mammalian branches separated in a relatively narrow period of time 100–120 million years ago [84, 85]. Thus, opossum bridges the long gap between the divergence of the birds (which by ancestry belong to the Reptilia) around 250–310 million years ago and the main groups of mammals.

Although the results largely confirmed the presence of most connexins across all mammals, a few unexpected results also emerged [22]. It was no surprise that Cx25 and Cx59, considered as specific for humans, could also be found in the chimpanzee. However, obvious orthologs were found in dog (Cx25 and 59), cow (Cx25 and 59; the Cx59 sequence was partial as it ran into an unsequenced area) and opossum (Cx25; Cx59 was not found). Thus, the lack of Cx25 and 59 in rodents is more likely due to a rodent-specific loss of these sequences, or, perhaps more unlikely, they reside in still unsequenced areas. Some details on Cx25 chromosomal location are discussed below.

Similarly, Cx33 cannot any longer be regarded as specific for rodents. Cx33-like sequences on the X chromosome are found in humans, chimpanzee and dog. A Cx33-like sequence is also found in the cow, but since this sequence has not yet been assembled into chromosomes, its location is presently unknown. Interestingly, the Cx33-like sequences in humans, chimpanzee and cow have the characteristics of a pseudogene with several stop codons spread out into the sequences, but the dog sequence is a potentially functional sequence encoding a 35.7-kDa protein. For the time being, we suggest the name Cx43pX for these pseudogene sequences in humans, chimpanzee and cow, and Cx35.7 for the sequence in the dog.

The major surprise during the investigations of the opossum genome was the identification of a novel connexin sequence with an open reading frame encoding a protein of 39.2 kDa [22]. The sequence was therefore called Cx39.2. Cx39.2 does not pair with any other mammalian connexin, but as will be seen below, orthologs are found in the lower vertebrates. Cx39.2 groups closest to Cx36 among the mammalian connexins, but has its reading frame in a single exon.

The opossum genome has two open reading frames, dubbed Cx33.1 and Cx35, with their highest similarity to human Cx31.3 or mouse Cx29 [22], one of the most variable of all connexin ortholog groups. Opossum Cx33.1 and 35 show approximately 85% identity, and were apparently generated by a gene duplication. (Please note that throughout this review, the term duplication signifies the generation of two sequences from one ancestral sequence, and is independent of the mechanism by which the duplication occurred.) An opossum pseudogene with highest similarity to Cx35 is also present, and is consequently called Cx35p1. Interestingly, humans and chimpanzee also possess a Cx31.3-like pseudogene, Cx31.3p1 [22, 86], but the primate pseudogene must have originated independently from the opossum Cx35p1 [22]. [In the most recent version of the human genome published in the Ensembl database (release v36, November 2005), this human Cx31.3p1 sequence has been modified, and the open reading frame now apparently encodes a Cx31.3-like sequence of 27.7 kDa. Independent sequencing efforts will be needed to verify whether this sequence is a pseudogene or a potentially functional gene.]

Recently, the genomes of African elephant (*Loxodonta africana*) and Rhesus monkey (*Macaca mulatta*) were added to the Ensembl genome databases. We have also searched these genomes for connexin sequences (see Table 1 and supplementary information at http://radium.no/lamyk). In short, 22 full-length or partial sequences have been found in rhesus monkey, including two pseudogenes. In elephant, 17 full-length or partial sequences have been found. We expect the four lacking functional connexins (Cx30.3, 31.1, 31.3, 40.1) to be found upon further progress in sequencing of the elephant genome.

From what is described above and summarized in Table 1, the so-called species-specific connexins must be reconsidered. Another important lesson is the dynamics of the connexin family when looked at in evolutionary terms. During the evolution of mammals, both gene duplications and gene losses have occurred several times. Two gene duplications have occurred in the opossum lineage (Cx31.3 duplicated to give Cx33.1 and Cx35, Cx35 later gave rise to the pseudogene Cx35p1). Another gene duplication occurred in the main mammalian lineage (Cx43 gave rise to

**Table 1.** Connexin orthologs among mammals.

| Hs | Pt | Mamu | Mm | Rn | Cf | Bt | La | Md |
|---|---|---|---|---|---|---|---|---|
| 25 β7 | 25 | 25 | – | – | 25 | 25 | 25 | 25 |
| 26 β2 | 26 | 26 | 26 β2 | 26 | 26 | 26 | 26 | 26 |
| 30 β6 | 30 | 30[1] | 30 β6 | 30 | 30 | 30 | 30 | 30 |
| 30.3 β4 | 30.3 | 30.3 | 30.3 β4 | 30.3 | 30.3 | 30.3 | NF | 30.3 |
| 31 β3 | NF | 31 | 31 β3 | 31 | 31 | 31 | 31 | 31 |
| 31.1 β5 | 31.1 | 31.1 | 31.1 β5 | 31.1 | 31.1 | 31.1 | NF | 31.1 |
| 31.3 δ1 | 31.3 | 31.3[1] | 29 δ1 | 29 | 31.3[1] | 31.3[1] | NF | 33.1[2]<br>35[2] |
| 31.3p1[3] | 31.3p1 | 31.3p1 | – | – | – | – | – | – |
| – | – | – | – | – | – | – | – | 35p1[2] |
| 31.9 δ3 | 31.9 | 31.9 | 30.2 δ3 | 30.2 | 31.9 | 31.9 | 31.9[1] | 31.9 |
| 32 β1 | 32 | 32 | 32 β1 | 32 | 32 | 32 | 32[1] | 32 |
| 43pX | 43pX | 43pX | 33 α6 | 33 | 35.7 | 43pX | 35.7 | – |
| 36 δ2 | 36 | 36 | 36 δ2 | 36 | 36 | 36 | 36[1] | 36[1] |
| 37 α4 | 37 | 37 | 37 α4 | 37 | 37 | 37 | 37 | 37 |
| – | – | – | – | – | – | – | – | 39.2 |
| 40 α5 | 40 | 40 | 40 α5 | 40 | 40 | 40 | 40[1] | 40 |
| 40.1 δ4 | 40.1 | 40.1 | 39 δ4 | 39 | 40.1[1] | 40.1 | NF | 40.1 |
| 43 α1 | 43 | 43 | 43 α1 | 43 | 43 | 43 | 43 | 43 |
| 43p1 α2 | 43p1 | | – | – | – | – | – | – |
| 45 γ1 | 45 | 45 | 45 γ1 | 45[1] | 45 | 45 | 45 | 45 |
| 46 α3 | 46 | 46 | 46 α3 | 46 | 46 | 44 | 46[1] | 46 |
| 47 γ2 | 47 | 47[1] | 47 γ2 | 47 | 47 | 47 | 47 | 47 |
| 50 α8 | 50 | 50 | 50 α8 | 50 | 50 | 50 | 50 | 50 |
| 59[4] α9 | 59 | 59 | – | – | 59 | 59[1] | 59[1] | NF |
| 62 α10 | 62 | 62 | 57 α10 | 57 | 62 | 62 | 62[1] | 62 |

We have as far as possible followed the human nomenclature, although some sequences might have masses deviating from the corresponding human sequence. The Greek nomenclature in human and mouse is according to Söhl and Willecke [21]. Note that Cx31.3 has two orthologs, Cx33.1 and Cx35, in opossum.

Hs, *Homo sapiens*; Pt, *Pan troglodytes* (chimpanzee); Mamu, *Macaca mulatta* (rhesus monkey); Mm, *Mus musculus* (mouse); Rn, *Rattus norvegicus* (rat); Cf, *Canis familiaris* (dog); Bt, *Bos taurus* (cow); La, *Loxodonta africana* (elephant); Md, *Monodelphis domestica* (opossum). – the sequence is likely to be absent from the genome; NF, not found, i.e. this sequence is likely to be present in this genome, but has not been sequenced/identified yet.

[1] This sequence is presently partial.

[2] The original Cx31.3 ortholog in the marsupial lineage appears to have duplicated to generate two functional sequences, Cx33.1 and Cx35. Later, Cx35 duplicated to generate a pseudogene, Cx35p1. As the latter event was separate from the event that generated the primate Cx31.3p1, these two pseudogenes are not orthologs.

[3] In contrast to previous versions, the present version of the human genome in the Ensembl databases (v36) suggests that this sequence could have an open reading frame encoding a 27.7-kDa protein.

[4] Cx59 is also called Cx58.

Cx43pX/Cx33/Cx35.7), and a third and fourth duplication occurred in the primate lineage (Cx31.3 gave rise to Cx31.3p1; Cx43 gave rise to Cx43p1, another Cx43 pseudogene that has been known for some time [87]). Gene losses may have occurred at least twice. Cx39.2 was lost from the main mammalian lineage some time after the divergence of the marsupials. Cx25 and Cx59 appear to be lost only in the rodent lineage. Whether the loss of the two sequences occurred simultaneously, or in two events, is not known.

It should not come as a surprise that the connexin family also shows some evolutionary dynamics when studied in a wider vertebrate perspective.

## The making of order: chicken, *Xenopus* and fishes

We have recently extended the analysis of the connexin family into nonmammalian vertebrates, including chicken, *X. tropicalis* and three species of fish. By using the conserved domains of the connexins, one can extend the structure of the connexin family across the vertebrates, despite the seemingly bewildering sequence differences. Here, we will only present a schematic overview, except for a look at some chromosomes (below), and leave the details for a forthcoming publication.

Connexins from chicken (*Gallus gallus* [88]), *X. tropicalis* and the three fishes zebrafish (*Danio rerio*), puffer-

fish (*Fugu rubripes* [89]) and green spotted pufferfish (*Tetraodon nigroviridis* [90]) were identified by searches in GenBank and in the Ensembl genome databases. In total, 16 connexins were found in chicken, 23 in *X. tropicalis*, 36 in zebrafish, 37 in *Fugu* and 38 in *Tetraodon*, most of them containing the complete conserved domain. At first sight, the high number of connexins in fishes was very surprising. An initial analysis indicated that a conspicuously high number of the sequences occurred in pairs where the two sequences were very similar, as indicated above for perch Cx35 and Cx34.7 [79]. This is fully consistent with a fish-specific genome duplication 300–350 million years ago [91, 92]. After the submission of this manuscript, Eastman et al. [93] published an evolutionary analysis of connexins in humans, mouse and zebrafish. They write that the large number of zebrafish connexins is not due to a simple whole-genome duplication followed by a small loss of connexins. We fully agree with Eastman et al. [93] that a number of gene duplications have occurred in the fish genome, but in our interpretation, these have occurred after the genome duplication. The contrasting conclusions on genome duplication are due to the inclusion of *Fugu* and *Tetraodon* in our analysis. Six zebrafish connexins that Eastman et al. [93] found as 'singlets' show up as duplicates in *Fugu* and/or *Tetraodon* (Cx52.8, 35.4, 33.8, 48.5, 39.9, 46.8; using the naming of Eastman et al. [93]), and their chromosomal locations are consistent with a genome duplication. As the sequencing of the three fish genomes is still ongoing, even more duplicates may be found in the future. Thus, when all three fish genomes are taken into consideration, the data are consistent with an early genome duplication in fish, followed by loss of some connexin sequences, and further gene duplications of other sequences.

A minor, but interesting point, is the spacing of the cysteines in the extracellular domains. The Cx31-specific spacing in EC2, $CX_5CX_5C$, is in fact a mammalian-specific spacing. While the chicken Cx31 ortholog is still lacking (see below; note that the sequence called chicken Cx31 is closely related to mammalian Cx26/30 [94]), the orthologs in *Xenopus* and the three fishes do not have the extra amino acid between the cysteines in EC2, and show the classic spacing, $CX_4CX_5C$. On the other hand, there is a single zebrafish sequence, ENSDARG00000041792 (this is the gene identification of the Ensembl database; the sequence belongs in the fish-specific Cx32.2 group; see below), which possesses a $CX_4CX_6C$ spacing in EC2.

When the conserved domains of the vertebrate connexins were phylogenetically analyzed, the mammals and the lower vertebrates were found to display a relatively consistent structure of the gene family (Fig. 3). The differences can easily be explained by a few gene losses and gene duplications at certain stages in evolution, as will be schematically described below. The (presumably) functional connexins can roughly be classified into the following categories.

1) Four groups of connexins are only found in fishes. These are the groups of zebrafish Cx28.6, Cx32.2 [95] (this is a multimember group that includes at least four sequences in zebrafish), Dr41782 (abbreviation for ENSDARG00000041782; this is the ortholog of Atlantic croaker Cx32.7 [19]) and Dr17927 (these orthologs constitute a Cx36-like group whose genes do not contain an intron in the open reading frame). For these connexins, we cannot yet conclude whether they are indeed specific for bony fishes (i.e. that the sequences have originated only in the bony fish lineage), or if they have been lost in the remaining vertebrate lineage at some point between the divergence of bony fishes and amphibians. If the latter is the case, at least some of these connexins should be found in cartilaginous fishes.
2) Two groups of connexins are found in fishes, *Xenopus* and/or chicken, but not in mammals, and must therefore have been lost before the origin of mammals. These groups are the orthologs of zebrafish Cx39.9 (NM_21826) and Cx43.4 [96, 97]. The latter group also includes zebrafish Cx44.2 (NM_131810). Cx43.4 has been suggested as an ortholog of Cx45 [96], but most phylogenetic models cluster the Cx43.4 group close to, but outside, both Cx45 and Cx47 groups. These are therefore closely related connexins. Partial sequences identified as mouse and rat Cx43.4 were recently submitted to GenBank (DQ149608, DQ149609, DQ125485), but we have not been able to substantiate these sequences as connexins in searches in the Ensembl mouse and rat genome databases.
3) Only two orthologous groups of functional connexins are specific to mammals. The first group is Cx33/35.7/43pX, where a duplication has occurred within the mammals as described above. The second group is Cx29/31.3 (33.1/35 in opossum), one of the most variable connexins, where a duplication has occurred at some point of time between the divergence of the birds and the divergence of the opossum. Judged from the phylogenetic trees, Cx29/31.3 may originate from the Cx45/47/43.4 groups, but this is somewhat uncertain. It is conspicuous that Cx43.4 (described above) disappeared approximately at the same time as Cx29/31.3 originated, but there is, as yet, no phylogenetic evidence that nonmammalian Cx43.4 transformed into Cx29/31.3 in mammals.
4) Two pairs of mammalian paralogs, Cx26/30 and Cx30.3/31.1, appear to have been generated as separate genes some time between the divergence of bony fishes and the divergence of opossum. A more detailed discussion on these connexins is provided below, where we take their chromosomal locations into consideration.

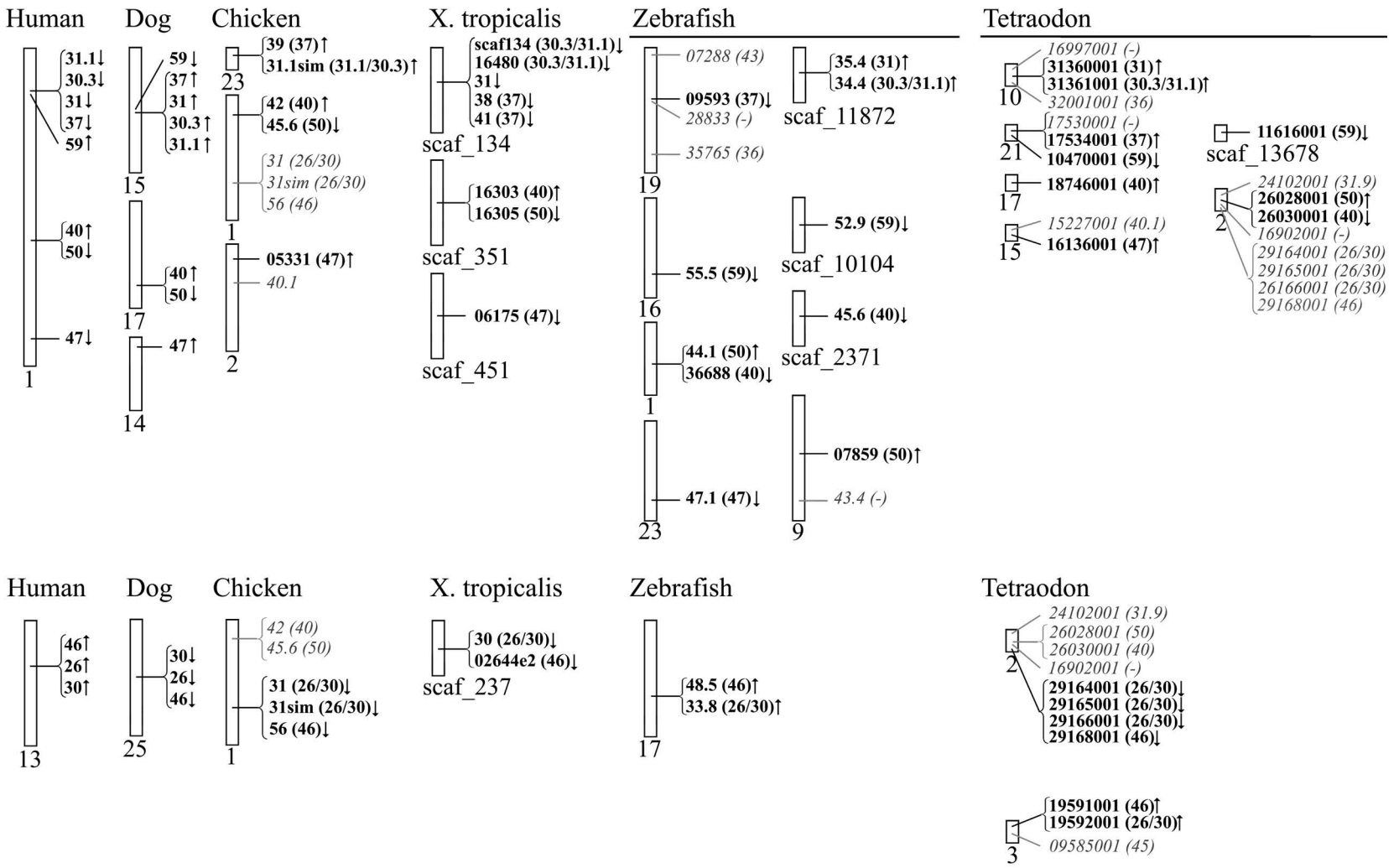


**Figure 4.** Interspecies comparison of chromosomal locations of orthologs of connexins found on human chromosomes 1 (upper) and 13 (lower). Adjacent connexin genes that are less than $10^5$ nucleotides distant from each other are enclosed in brackets. The transcriptional direction (according to http://www.ensembl.org) is indicated by arrows. The ortholog human connexin number is given in parentheses if the naming is not identical to the human connexin. Many sequences are named by an abbreviation of the Ensembl gene identification as indicated in Table 2 and in the text. A number of the sequences have not yet been assembled into chromosomes. For these sequences, the scaffold number (scaf_XXX) is indicated instead of the chromosome number. The sizes of the chromosomes/scaffolds are not proportional. Sequences not orthologous to human connexins located on chromosome 1 or 13 are indicated in gray italics.

**Table 2.** Connexin orthologs among vertebrates (the mammal is a fusion of the mammalian sequences of Table 1).

| Mammal | Gg | Xt | Dr | Fr | Tn |
|---|---|---|---|---|---|
| 25 | NF | 18051 | 43971 (28.8) | 153654 | 35526001 |
| 26 | 31sim | 30 | 33.8 | | 29165001<br>29164001 |
| 30 | 31 | | | 137409<br>124726 | 29166001<br>19592001 |
| 30.3 | 31.1sim | scaf134 | 34.4 | 129456 | 31361001 |
| 31.1 | | 16480 | | 158780 | |
| 31 | NF | 31 | 35.4 | 157125<br>137254 | 31360001 |
| 31.3 | – | – | – | – | – |
| – | – | – | 28.6<br>28833 (30.9) | 143712<br>125568 | 16997001[1]<br>17530001 |
| 31.9 | – | 24391 | – | 133739 | 24102001 |
| 32 | 32 | 32 | 37518 (31.7)<br>27.5 | 144150<br>157320 | 17091001<br>22232001 |
| 33/43pX | – | – | – | – | – |
| 36 | 36 | 12912<br>02781 | 35.1<br>35340 (34.1)<br>35765[1] (–) | 137912<br>150812<br>130034 | 32001001<br><br>26779001[1] |
| – | – | – | 17927 (35.8) | 123845 | 05142001 |
| – | – | – | 28.9<br>41792 (28.1)<br>32.2<br>32.3 | <br><br>158218 | 35462001<br><br>35459001 |
| – | – | – | 41782[2] (34.5) | 159440 | 35458001 |
| 37 | 39 | 41<br>38 | 09593 (39.4) | 125574 | 17534001[1] |
| 39.2 | NF | 09451 | 09334 (40.5) | 142334 | 22263001 |
| 40 | 42 | 16303 | 45.6<br>36688 (41.8) | 137519<br>129429 | 26030001<br>18746001 |
| 40.1 | 40.1 | scaf137 | 45445[1] (46.8) | 139615<br>150243[1] | 16594001[1]<br>15227001[1] |
| 43 | 43 | 43 | 43<br>07288 (40.8) | 130494 | 13946001 |
| – | 11233[1] | 14395 | 44.2<br>43.4 | 122421<br>121008 | 01153001<br>16902001 |
| 45 | 45 | 03391 | 38313 (52.8) | 145344<br>148221 | 09585001<br>19395001 |
| 46 | 56 | 02644e2 | 48.5 | 149277 | 29168001<br>19591001 |
| – | 39.9[1] | scaf482 | 39.9 | 141787<br>144151 | 22231001<br>17092001 |

**Table 2.** (Continued).

| Mammal | Gg | Xt | Dr | Fr | Tn |
|---|---|---|---|---|---|
| 47 | 05331 | 06175 | 47.1 | 136128 | 16136001 |
| 50 | 45.6 | 16305 | 44.1<br>07859 (50.5) | <br>scaf387 | 26028001 |
| 59 | NF | 05268 | 55.5<br>52.9 | scaf797<br>139587 | 11616001<br>10470001 |
| 62 | 15606 | 19125 | 52.6<br>XM695977[3] (52.7) | 149617 | 06204001 |

Gg, *Gallus gallus* (chicken); Xt, *Xenopus tropicalis*; Dr, *Danio rerio* (zebrafish); Fr, *Fugu rubripes* (pufferfish); Tn, *Tetraodon nigroviridis* (green spotted pufferfish). We have refrained from giving any previously unannotated sequences a connexin name, but have, rather, used an abbreviated version of the Ensembl gene identification (see below). For zebrafish, the recent naming of connexins by Eastman et al. [93] has been put in parentheses. The abbreviated version of the Ensembl gene identification is used in the following way: chicken, 11233 should be read as ENSGALG00000011233; *Xenopus*, 18051 should be read as ENSXETG00000018051; zebrafish, 43971 should be read as ENSDARG00000043971; *Fugu*, 153654 should be read as SINFRUG00000153654 [note that the new version of the Ensembl databases (release v36) has added the prefix NEW to all gene identifications, in addition to changing the gene identification numbers for some of them; for convenience releases v33, v34 or v35 of the Ensembl databases should be used]; *Tetraodon*, 35526001 should be read as GSTENG00035526001. All nonmammalian sequences will be released together with a forthcoming publication, or they can be found in the Ensembl databases. Sequences named as scafXXX have not been recognized by the databanks as a potential gene. We have named these sequences according to their scaffold number.

[1] The sequence is partial.

[2] This is the ortholog group of Atlantic croaker Cx32.7.

[3] The sequence can be found in GenBank under this accession number.

5) The large majority of connexins, 16 orthologous groups [in human nomenclature: Cx25, 31, 31.9, 32, 36, 37, 39.2 (opossum), 40, 40.1, 43, 45, 46, 47, 50, 59, and 62] can be defined across the vertebrate groups from bony fishes to mammals, even if some of these connexins are lacking in one or more of the included species.

The results of Eastman et al. [93] are in accordance with most of our conclusions even though we used partially different approaches to the evolutionary analyses, but there are also some differences. Most of the differences can be ascribed to their inclusion of only three species (human, mouse, zebrafish), while we have included a wider range of mammals (opossum in particular), *X. tropicalis*, chicken and two other fishes. For example, Eastman et al. [93] conclude that the group of Cx43.4 is specific to fishes. It is not, as it is also found in *Xenopus* and chicken. The same is valid for Cx39.9 orthologs. Furthermore, their DreCx40.5 is the ortholog of opossum Cx39.2. Similar to our results, Eastman et al. [93] found no zebrafish ortholog of human Cx31.9, but orthologs are found in *Fugu* and *Tetraodon*. We can also be more specific on when the duplication of the preCx26/30 and preCx30.3/31.1 sequences occurred (this is further discussed below).

We believe that other differences can be explained by our use of the conserved domains, while Eastman et al. [93] used full-length sequences, but we have not analyzed this in detail. For example, they conclude that DreCx40.8, 34.1, 28.8 and 47.1 are closely related, but nonorthologous, to Cx43, 36, 25 and 47, respectively. We conclude that they are the respective orthologs. This is because many phylogenetic models give identical results, because we have included additional species, and because we are in several cases able to achieve an improved statistics by excluding the variable domains. On the other hand, our suggestion that mammalian Cx37 has orthologs in fishes is not supported by strong statistics. However, most of the models locate DreCx39.4 and the corresponding orthologs in *Fugu* and *Tetraodon* close to the mammalian, chicken and *Xenopus* orthologs of Cx37. Furthermore, there is conservation among the six C-terminal-most amino acids of the full-length protein. Such short stretches of conservation in the C-terminal tail are often observed among orthologs of the larger connexins, even though the remaining parts of the C-terminal tails are totally different. This property may therefore be used as a supplementary criterion for deciding the orthology relationships.

In our view, the major difference between the work of Eastman et al. [93] and ourselves concerns the Cx59 and Cx62 orthologous groups. They locate these groups as a separate clade more closely related to the $\beta$ group connexins. We always find Cx59 and 62 to locate together with Cx40, 46 and 50, although the exact topology varies with the model used. The results of Eastman et al. [93] are likely due to 'noise' introduced by the long variable domains of the Cx59 and 62 orthologs. Alignment of the conserved domains of all human connexin proteins

clearly shows that Cx59 and 62 are more closely related to the $\alpha$ connexins than to the $\beta$ connexins, having the highest identities with Cx46 and 50.

Compared with the changes that have occurred within mammals (described above), the number of losses and duplications that have occurred in the vertebrate lineage leading to mammals is low. If we exclude the four fish-specific connexin groups, the differences between the common ancestor of bony fishes and mammals on the one side, and mammals on the other side can be explained by three duplications (Cx26/30, Cx30.3/31.1, and the generation of Cx31.3) and two gene losses (Cx39.9 and Cx43.4; considering that Cx39.2 was lost within mammalian evolution).

Thus, the last common ancestor of bony fishes and mammals had at least as many connexins as the mammals have today, especially if we assume that the fish-specific connexins were present. Using the human and zebrafish names, the connexins in this old pre-fish vertebrate were Cx25, pre26/30, pre30.3/31.1, 28.6, 31, 31.9, 32, 32.2, 32.7 (Atlantic croaker name), 36, a 36-related sequence (dubbed Dr17927 in Fig. 3), 37, 39.2 (opossum), 39.9, 40, 40.1, 43, 43.4, 45, 46, 47, 50, 59, and 62 (with mammalian orthologs underlined).

## A look at some vertebrate chromosomes

The distribution of connexins on the chromosomes appears to be nonrandom [21]. Human chromosome 1 contains eight connexins, chromosomes 6 and 13 have three connexins each, chromosome 17 has two connexins, and chromosomes 7, 10, 15 and X have one each (pseudogenes are not included in this count). On chromosomes 1 and 13, some of the connexins cluster closely together, with less than $10^5$ nucleotides between adjacent sequences. The closely linked sequences on human chromosome 1 are Cx31.1, 30.3, 31 and 37 in one cluster, and Cx40 and 50 in another cluster (Fig. 4). On chromosome 13, Cx30, Cx26 and Cx46 are closely linked. When we look across the vertebrates, the orthologs of these connexins are usually found clustered in similar manners with their relative transcription direction conserved (Fig. 4). Such comparisons may give support to the predictions made independently by the phylogenetic analysis, and may also give rise to new predictions. Let us consider the largest connexin cluster, Cx31.1, 30.3, 31 and 37, on human chromosome 1. In chicken, the cluster has been reduced to two sequences, Cx39 [98] and Cx31.1sim (abbreviation for 'similar to Cx31.1'; accession number XM_425784). The former is clearly an ortholog of human Cx37, the latter is related to Cx30.3 and 31.1. Two questions arise from these observations: (i) Is chicken Cx31.1sim the common ortholog of mammalian Cx30.3 and 31.1, or is it the ortholog to only one of the two mammalian sequences, implying that there could still be an unidentified Cx30.3/31.1-like sequence in chicken? (ii) An ortholog of mammalian Cx31 is lacking in the chicken. In mammals and *Xenopus*, the Cx31 ortholog is positioned in close vicinity to the Cx30.3/31.1 and 37 orthologs. Is chicken lacking the Cx31 ortholog, or is it still unidentified?

At present, these questions cannot be answered with certainty, but we will here provide some speculations. All phylogenetic models positioned chicken Cx31.1sim closer to Cx31.1 than Cx30.3. If we assume that chicken Cx31.1sim is indeed the 1:1 ortholog of mammalian Cx31.1, we would expect two chicken connexins to be positioned in the chromosomal segment between chicken Cx39 and chicken Cx31.1sim, a stretch of approximately 8000 nucleotides. A detailed search did not reveal any connexin-like sequences in this segment, but there are some unsequenced stretches where a connexin could potentially hide. We believe it is unlikely that, by coincidence, two connexins have not been detected in such a short stretch. Therefore, our guess is that only one connexin will eventually be found in this area, and that this will be the Cx31 ortholog. Thus, chicken Cx31.1sim may represent the preCx30.3/31.1. If this is correct, the two Cx30.3/31.1-like sequences (scaf134 and 16480) found in *X. tropicalis* are due to an amphibian-specific gene duplication.

A similar problem concerns Cx26/30 orthologs. The Cx26/30-like sequences from fishes are always found outside the mammalian Cx26 and Cx30 groups in the phylogenetic trees. This suggests that there was one common preCx26/30 in the common ancestor of fishes and other vertebrates. If so, when did the separate Cx26 and Cx30 sequences originate? *X. tropicalis* Cx30 in most models splits off from the common root of the mammalian Cx26 and Cx30 groups (Fig. 3), and when inside, it generates a trifurcating topology with all mammalian Cx26 on one side and all Cx30 on the other. Our present view is that *X. tropicalis* Cx30 (and *X. laevis* Cx30; BC043797) is also the common ortholog of both mammalian Cx26 and Cx30, thus representing the preCx26/30. In chicken, there are two Cx26/30-like sequences, chicken Cx31 [94] and chicken Cx31sim (XM_425641). In the phylogenetic analysis, they never distribute to Cx26 and Cx30 in a 1:1 fashion, but always group together, usually closer to Cx26, but in a few models also outside the mammalian Cx26/30 groups. When the full-length amino acid sequences are aligned, mammalian Cx30 and chicken Cx31 have some similarities in the part of the C-terminal tail that extends beyond the position corresponding to the stop codons in mammalian Cx26 and chicken Cx31sim. When the chromosomal locations of the sequences are considered, chicken Cx31-Cx31sim-Cx56 have the expected transcription direction and very similar spacings as mammalian Cx30-Cx26-Cx46 (Fig. 4). Altogether,

these observations may suggest that the duplication of preCx26/30 to separate Cx26 and Cx30 sequences occurred shortly before the divergence of Reptilia, so that the bird sequences did not fully acquire the characteristics that later appeared in the mammalian lineage for each of these two sequences. These points might be further elucidated by investigating reptile connexins.

As pointed out above, rodents lack two connexins, Cx25 and 59, found in other vertebrates from mammals to fishes. In humans, Cx25 and Cx62 are located on chromosome 6, with a spacing of approximately 2.5 million nucleotides. The murine ortholog of Cx62, Cx57, is located on chromosome 4. The Ensembl genome databases indicate synteny (literally 'on the same thread,' i.e. there is a preserved order of genes) between human chromosome 6 and mouse chromosome 4 for the entire stretch between Cx62 and Cx25. Nearly all genes found in this stretch have their corresponding orthologs in the other species. Even the genes in the immediate surroundings of Cx25 find their orthologs in mouse. On the one side of human Cx25, human ENSG00000111850 (a gene without any name) is orthologous with ENSMUSG00000028295; on the other side, a zinc finger protein is found, human ENSG00000188994 and mouse ENSMUSG00000039967. Thus, there has been a very specific loss of Cx25 in rodents that apparently cannot be explained, for example, by DNA breakage and interchromosomal exchange.

## Close to the origin: tunicates

The presence of connexins in the tunicate heart was indicated some years ago [99]. With genomic sequencing of a tunicate, *Ciona intestinalis* [100], firm evidence of connexin-like genes in urochordates was established [73]. This was further confirmed by the sequencing and functional studies of a connexin, called Cx47, from another tunicate, *Halocynthia pyriformis* [70]. When *Halocynthia* Cx47 was expressed in *Xenopus* oocytes, its electrophysiological properties were reminiscent of mammalian Cx36 orthologs [70]. Among the human connexins, the highest degree of identity of the conserved domains of *Halocynthia* Cx47 was obtained with Cx36, 45 and 47 [70]. BLAST analyses of *Ciona* connexins found in most cases mammalian Cx45 as the best hit [73, 101]. When the conserved domains of the *Ciona* connexins are compared among themselves, the average identity is very similar to the average identity for the conserved domains of the human connexins. However, while the human connexins show a range of identities from approximately 15% (Cx25 vs 31.3) to nearly 80% (Cx30.3 vs 31.1), the range is considerably smaller in *Ciona*, stretching from 35 to 67% identity. This suggests that the tunicates have not developed connexin subfamilies to the same extent as the vertebrates. This is supported by detailed phylogenetic analyses, because all tunicate connexins become located together with the group III connexins. Some tunicate connexins locate themselves within or in the vicinity of the Cx36/39.2 groups, other tunicate connexins locate themselves as intermediates between the Cx36/39.2 and Cx45/47/43.4 groups, and still others branch out from the common root of the Cx36/39.2/45/47/43.4 groups. Thus, this may suggest that the original connexin(s) are more closely related to these vertebrate connexins, and that the classic $\alpha$ and $\beta$ connexins developed later.

The skate Cx36 ortholog is very similar to the mammalian one, but no other connexins have been sequenced from skate, or any of the evolutionary branches (e.g. amphioxus and lamprey) that are supposed to have split off in the period between the urochordate (tunicate) and teleost (bony fish) branches. One or two genome duplications probably occurred during this period [92, 102, 103], and our guess is that the major part of the shaping of the 'modern' connexin family, both with regard to the number of connexins, the subfamily structure and their gene structure, took place during this period.

## Conclusions and perspectives

The structure of the connexin family is quite conserved among vertebrates, in the sense that a few gene duplications or gene losses can explain all changes that have taken place along the main evolutionary branch from the pre-fish vertebrate to mammals. With a high degree of certainty, the common ancestor of bony fishes and mammals possessed the orthologs of mammalian Cx25, pre26/30, pre30.3/31.1, 31, 31.9, 32, 36, 37, 39.2, 40, 40.1, 43, 45, 46, 47, 50, 59 and 62. In addition, it also had some connexin genes (two to six?) that were lost before the mammals emerged. The most original connexins appear to be more closely related to the connexins that have been difficult to arrange into the $\alpha$ and $\beta$ subfamilies, i.e. Cx36, 39.2, 43.4, 45 and 47.

Although we are confident that we have included the majority of connexins in chicken, *Xenopus* and the three fishes in our analyses, a few more sequences are likely to be detected as the sequencing progresses, and different errors are corrected. The inclusion of a few new connexins might change some smaller details, but we believe that the main pattern and structure outlined here and in the work of Eastman et al. [93] will not change to any large extent. The chance of changes is largest in very closely related groups, e.g. the Cx43.4, 45 and 47 groups, or other groups where some sequences are still partial or other sequences are lacking.

However, there are still some major points in connexin evolution where previous work, at best, can only give weak hints. How was the 'modern' connexin family shaped during the genome duplication(s) in the early vertebrates?

Were there one, or more, pre-vertebrate connexins that formed the basis for the vertebrate connexins? A search for connexins in skate, lamprey and amphioxus genomes would bring us closer to the answers to these questions. Even more intriguing is the question why the pannexins were not sufficient for gap junctional communication in chordates, and which gene gave rise to the very first connexin.

*Acknowledgements.* V. C. is supported by The Norwegian Cancer Society.

1 Schmidtmann M. (1925) Über die intracellulare Wasserstoffionenkonzentration unter physiologischen and einigen pathologischen Bedingungen. Z. Ges. Exp. Med. **45:** 714–742
2 Shibata Y., Kumai M., Nishii K. and Nakamura K. (2001) Diversity and molecular anatomy of gap junctions. Med. Electron Microsc. **34:** 153–159
3 Furshpan E. J. and Potter D. D. (1959) Transmission at the giant motor synapses of the crayfish. J. Physiol. **145:** 289–325
4 Söhl G., Maxeiner S. and Willecke K. (2005) Expression and functions of neuronal gap junctions. Nat. Rev. Neurosci. **6:** 191–200
5 Robertson J. D. (1957) New observations on the ultrastructure of the membranes of frog peripheral nerve fibers. J. Biophys. Biochem. Cytol. **3:** 1043–1051
6 Karrer H. E. (1960) Cell interconnections in normal human cervical epithelium. J. Biophys. Biochem. Cytol. **7:** 181–185
7 Hama K. (1959) Some observations on the fine structure of the giant nerve fibers of the earthworm, *Eisenia foetida*. J. Biophys. Biochem. Cytol. **6:** 61–66
8 Kanno L. and Loewenstein W. R. (1964) Intercellular diffusion. Science **143:** 959–960
9 Loewenstein W. R. and Kanno Y. (1964) Studies on an epithelial (gland) cell junction. I. Modifications of surface membrane permeability. J. Cell Biol. **22:** 565–586
10 Kanno Y. and Loewenstein W. R. (1966) Cell-to-cell passage of large molecules. Nature **212:** 629–630
11 Kumar N. M. and Gilula N. B. (1986) Cloning and characterization of human and rat liver cDNAs coding for a gap junction protein. J. Cell Biol. **103:** 767–776
12 Paul D. L. (1986) Molecular cloning of cDNA for rat liver gap junction protein. J. Cell Biol. **103:** 123–134
13 Beyer E. C., Paul D. L. and Goodenough D. A. (1987) Connexin43: a protein from rat heart homologous to a gap junction protein from liver. J. Cell Biol. **105:** 2621–2629
14 Zhang J.-T. and Nicholson B. J. (1989) Sequence and tissue distribution of a second protein of hepatic gap junctions, Cx26, as deduced from its cDNA. J. Cell Biol. **109:** 3391–3401
15 Gimlich R. L., Kumar N. M. and Gilula N. B. (1988) Sequence and developmental expression of mRNA coding for a gap junction protein in *Xenopus*. J. Cell Biol. **107:** 1065–1073
16 Ebihara L., Beyer E. C., Swenson K. I., Paul D. L. and Goodenough D. A. (1989) Cloning and expression of a *Xenopus* embryonic gap junction protein. Science **243:** 1194–1195
17 Musil L. S., Beyer E. C. and Goodenough D. A. (1990) Expression of the gap junction protein connexin43 in embryonic chick lens: molecular cloning, ultrastructural localization, and post-translational phosphorylation. J. Membr. Biol. **116:** 163–175
18 Beyer E. C. (1990) Molecular cloning and developmental expression of two chick embryo gap junction proteins. J. Biol. Chem. **265:** 14439–14443
19 Yoshizaki G., Patiño R. and Thomas P. (1994) Connexin messenger ribonucleic acids in the ovary of Atlantic croaker: molecular cloning and characterization, hormonal control, and correlation with appearance of oocyte maturational competence. Biol. Reprod. **51:** 493–503
20 O'Brien J., Al-Ubaidi M. R. and Ripps H. (1996) Connexin 35: a gap-junctional protein expressed preferentially in the skate retina. Mol. Biol. Cell **7:** 233–243
21 Söhl G. and Willecke K. (2003) An update on connexin genes and their nomenclature in mouse and man. Cell Commun. Adhes. **10:** 173–180
22 Cruciani V. and Mikalsen S.-O. (2005) The *connexin* gene family in mammals. Biol. Chem. **386:** 325–332
23 de Boer T. P. and Van der Heyden M. A. G. (2005) *Xenopus* connexins: how frogs bridge the gap. Differentiation **73:** 330–340
24 Fraser S. E., Green C. R., Bode H. R. and Gilula N. B. (1987) Selective disruption of gap junctional communication interferes with a patterning process in *Hydra*. Science **237:** 49–55
25 Dewey M. M. and Barr L. (1964) A study of the structure and distribution of the nexus. J. Cell Biol. **23:** 553–585
26 Curtin K. D., Zhang Z. and Wyman R. J. (1999) *Drosophila* has several genes for gap junction proteins. Gene **232:** 191–201
27 Phelan P. and Starich T. A. (2001) Innexins get into the gap. BioEssays **23:** 388–396
28 Panchin Y., Kelmanson I., Matz M., Lukyanov K., Usman N. and Lukyanov S. (2000) A ubiquitous family of putative gap junction molecules. Curr. Biol. **10:** R473–R474
29 Bruzzone R., Hormuzdi S. G., Barbe M. T., Herb A. and Monyer H. (2003) Pannexins, a family of gap junction proteins expressed in brain. Proc. Natl. Acad. Sci. USA **100:** 13644–13649
30 Panchin Y. V. (2005) Evolution of gap junction proteins – the pannexin alternative. J. Exp. Biol. **208:** 1415–1419
31 Phelan P. (2005) Innexins: members of an evolutionarily conserved family of gap-junction proteins. Biochim. Biophys. Acta **1711:** 225–245
32 Loewenstein W. R. (1979) Junctional intercellular communication and the control of growth. Biochim. Biophys. Acta **560:** 1–65
33 Lo C. W. (1996) The role of gap junction membrane channels in development. J. Bioenerg. Biomembr. **28:** 379–385
34 Kelsell D. P., Di W. L. and Houseman M. J. (2001) Connexin mutations in skin disease and hearing loss. Am. J. Hum. Genet. **68:** 559–568
35 Söhl G. and Willecke K. (2004) Gap junctions and the connexin protein family. Cardiovasc. Res. **62:** 228–232
36 Gerido D. A. and White T. W. (2004) Connexin disorders of the ear, skin, and lens. Biochim. Biophys. Acta **1662:** 159–170
37 Nakase T. and Naus C. C. (2004) Gap junctions and neurological disorders of the central nervous system. Biochim. Biophys. Acta **1662:** 149–158
38 Cruciani V. and Mikalsen S.-O. (2002) Connexins, gap junctional intercellular communication and kinases. Biol. Cell **94:** 433–443
39 Vine A. L. and Bertram J. S. (2002) Cancer chemoprevention by connexins. Cancer Metastasis Rev. **21:** 199–216
40 Sáez J. C., Berthoud V. M., Brañes M. C., Martínez A. D. and Beyer E. C. (2003) Plasma membrane channels formed by connexins: their regulation and functions. Physiol. Rev. **83:** 1359–1400
41 Goodenough D. A. and Paul D. L. (2003) Beyond the gap: functions of unpaired connexon channels. Nat. Rev. Mol. Cell Biol. **4:** 285–294
42 King T. J. and Lampe P. D. (2005) Temporal regulation of connexin phosphorylation in embryonic and adult tissues. Biochim. Biophys. Acta **1719:** 24–35
43 Wei C.-J., Xu X. and Lo C. W. (2004) Connexins and cell signaling in development and disease. Annu. Rev. Cell Dev. Biol. **20:** 811–838
44 Theis M., Söhl G., Eiberger J. and Willecke K. (2005) Emerging complexities in identity and function of glial connexins. Trends Neurosci. **28:** 188–195

45 Harris A. L. (2001) Emerging issues of connexin channels: biophysics fills the gap. Q. Rev. Biophys. **34:** 325–472
46 Jiang J. X. and Gu S. (2005) Gap junction- and hemichannel-independent actions of connexins. Biochim. Biophys. Acta **1711:** 208–214
47 Chanson M., Derouette J. P., Roth I., Foglia B., Scerri I., Dudez T. et al. (2005) Gap junctional communication in tissue inflammation and repair. Biochim. Biophys. Acta **1711:** 197–207
48 Cheng S., Christie T. and Valdimarsson G. (2003) Expression of connexin48.5, connexin44.1, and connexin43 during zebrafish (*Danio rerio*) lens development. Dev. Dyn. **228:** 709–715
49 Shiels A., Mackay D., Ionides A., Berry V., Moore A. and Bhattacharya S. (1998) A missense mutation in the human connexin50 gene (*GJA8*) underlies autosomal dominant 'zonular pulverulent' cataract, on chromosome 1q. Am. J. Hum. Genet. **62:** 526–532
50 White T. W., Goodenough D. A. and Paul D. L. (1998) Targeted ablation of connexin50 in mice results in microphthalmia and zonular pulverulent cataracts. J. Cell Biol. **143:** 815–825
51 Mackay D., Ionides A., Kibar Z., Rouleau G., Berry V., Moore A. et al. (1999) Connexin46 mutations in autosomal dominant congenital cataract. Am. J. Hum. Genet. **64:** 1357–1364
52 Gong X., Li E., Klier G., Huang Q., Wu Y., Lei H. et al. (1997) Disruption of $\alpha_3$ connexin gene leads to proteolysis and cataractogenesis in mice. Cell **91:** 833–843
53 Cruciani V., Heintz K.-M., Husøy T., Hovig E., Warren D. J. and Mikalsen S.-O. (2004) The detection of hamster connexins: a comparison of expression profiles with wild-type mouse and the cancer-prone *Min* mouse. Cell Commun. Adhes. **11:** 155–171
54 Söhl G., Nielsen P. A., Eiberger J. and Willecke K. (2003) Expression profiles of the novel human connexin genes hCx30.2, hCx40.1, and hCx62 differ from their putative mouse orthologues. Cell Commun. Adhes. **10:** 27–36
55 Miller M. P. and Kumar S. (2001) Understanding human disease mutations through the use of interspecific genetic variation. Hum. Mol. Genet. **10:** 2319–2328
56 Miller M. P., Parker J. D., Rissing S. W. and Kumar S. (2003) Quantifying the intragenic distribution of human disease mutations. Ann. Hum. Genet. **67:** 567–579
57 Willecke K., Hennemann H., Dahl E., Jungbluth S. and Heynkes R. (1991) The diversity of connexin genes encoding gap junctional proteins. Eur. J. Cell Biol. **56:** 1–7
58 Bennett M. V., Barrio L. C., Bargiello T. A., Spray D. C., Hertzberg E. and Sáez J. C. (1991) Gap junctions: new tools, new answers, new questions. Neuron **6:** 305–320
59 Willecke K., Eiberger J., Degen J., Eckardt D., Romualdi A., Güldenagel M. et al. (2002) Structural and functional diversity of connexin genes in the mouse and human genome. Biol. Chem. **383:** 725–737
60 Hertzberg E. L., Disher R. M., Tiller A. A., Zhou Y. and Cook R. G. (1988) Topology of the Mr 27,000 liver gap junction protein: cytoplasmic localization of amino- and carboxyl termini and a hydrophilic domain which is protease-hypersensitive. J. Biol. Chem. **263:** 19105–19111
61 Falk M. M., Kumar N. M. and Gilula N. B. (1994) Membrane insertion of gap junction connexins: polytopic channel forming membrane proteins. J. Cell Biol. **127:** 343–355
62 Hoh J. H., John S. A. and Revel J.-P. (1991) Molecular cloning and characterization of a new member of the gap junction gene family, connexin-31. J. Biol. Chem. **266:** 6524–6531
63 Hennemann H., Schwarz H. J. and Willecke K. (1992) Characterization of gap junction genes expressed in F9 embryonic carcinoma cells: molecular cloning of mouse connexin31 and -45 cDNAs. Eur. J Cell Biol. **57:** 51–58
64 Zoidl G., Bruzzone R., Weickert S., Kremer M., Zoidl C., Mitropoulou G. et al. (2004) Molecular cloning and functional expression of zfCx52.6: a novel connexin with hemichannel-forming properties expressed in horizontal cells of the zebrafish retina. J. Biol. Chem. **279:** 2913–2921
65 Cheng S., Shakespeare T., Mui R., White T. W. and Valdimarsson G. (2004) Connexin 48.5 is required for normal cardiovascular function and lens development in zebrafish embryos. J. Biol. Chem. **279:** 36993–37003
66 Condorelli D. F., Parenti R., Spinella F., Salinaro A. T., Belluardo N., Cardile V. et al. (1998) Cloning of a new gap junction gene (Cx36) highly expressed in mammalian brain neurons. Eur. J. Neurosci. **10:** 1202–1208
67 Söhl G., Degen J., Teubner B. and Willecke K. (1998) The murine gap junction gene connexin36 is highly expressed in mouse retina and regulated during brain development. FEBS Lett. **428:** 27–31
68 Maltzahn J. von, Euwens C., Willecke K. and Söhl G. (2004) The novel mouse *connexin39* gene is expressed in developing striated muscle fibers. J. Cell Sci. **117:** 5381–5392
69 Hombach S., Janssen-Bienhold U., Söhl G., Schubert T., Büssow H., Ott T. et al. (2004) Functional expression of connexin57 in horizontal cells of the mouse retina. Eur. J. Neurosci. **19:** 2633–2640
70 White T. W., Wang H., Mui R., Litteral J. and Brink P. R. (2004) Cloning and functional expression of invertebrate connexins from *Halocynthia pyriformis*. FEBS Lett. **577:** 42–48
71 Roy S. W. and Gilbert W. (2005) The pattern of intron loss. Proc. Natl. Acad. Sci USA **102:** 713–718
72 Jeffares D. C., Mourier T. and Penny D. (2006) The biology of intron gain and loss. Trends Genet. **22:** 16–22
73 Sasakura Y., Shoguchi E., Takatori N., Wada S., Meinertzhagen I. A., Satou Y. et al. (2003) A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. X. Genes for cell junctions and extracellular matrix. Dev. Genes Evol. **213:** 303–313
74 Dermietzel R., Kremer M., Paputsoglu G., Stang A., Skerrett I. M., Gomes D. et al. (2000) Molecular and functional diversity of neural connexins in the retina. J. Neurosci. **20:** 8331–8343
75 Van der Heyden M. A. G., Van Eijk M., Wilders R., de Bakker J. M. T. and Opthof T. (2004) Connexin43 orthologues in vertebrates: phylogeny from fish to man. Dev. Genes Evol. **214:** 261–266
76 Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome. Nature **420:** 520–562
77 Willecke K., Heynkes R., Dahl E., Stutenkemper R., Hennemann H., Jungbluth S. et al. (1991) Mouse connexin37: cloning and functional expression of a gap junction gene highly expressed in lung. J. Cell Biol. **114:** 1049–1057
78 Yoshizaki G. and Patiño R. (1995) Molecular cloning, tissue distribution, and hormonal control in the ovary of Cx41 mRNA, a novel *Xenopus* connexin gene transcript. Mol. Reprod. Dev. **42:** 7–18
79 O'Brien J., Bruzzone R., White T. W., Al-Ubaidi M. R. and Ripps H. (1998) Cloning and expression of two related connexins from the perch retina define a distinct subgroup of the connexin family. J. Neurosci. **18:** 7625–7637
80 Chatterjee B., Chin A. J., Valdimarsson G., Finis C., Sonntag J. M., Choi B. Y. et al. (2005) Developmental regulation and expression of the zebrafish connexin43 gene. Dev. Dyn. **233:** 890–906
81 Pang H., Tang J., Chen S. S. and Tao S. (2005) Statistical distributions of optimal global alignment scores of random protein sequences. BMC Bioinform. **6:** 257
82 Risek B., Guthrie S., Kumar N. and Gilula N. B. (1990) Modulation of gap junction transcript and protein expression during pregnancy in the rat. J. Cell Biol. **110:** 269–282
83 Haefliger J.-A., Bruzzone R., Jenkins N. A., Gilbert D. J., Copeland N. G. and Paul D. L. (1992) Four novel members of

the connexin family of gap junction proteins. Molecular cloning, expression, and chromosome mapping. J. Biol. Chem. **267:** 2057–2064
84 Kumar S. and Hedges S. B. (1998) A molecular timescale for vertebrate evolution. Nature **392:** 917–920
85 O'Brien S. J., Eisenberg J. F., Miyamoto M., Hedges S. B., Kumar S., Wilson D. E. et al. (1999) Genome maps 10. Comparative genomics. Mammalian radiations. Wall chart. Science **286:** 463–478
86 Söhl G., Eiberger J., Jung Y. T., Kozak C. A. and Willecke K. (2001) The mouse gap junction gene connexin29 is highly expressed in sciatic nerve and regulated during brain development. Biol. Chem. **382:** 973–978
87 Fishman G. I., Eddy R. L., Shows T. B., Rosenthal L. and Leinwand L. A. (1991) The human connexin gene family of gap junction proteins: distinct chromosomal locations but similar structures. Genomics **10:** 250–256
88 Hillier L. W., Miller W., Birney E., Warren W., Hardison R. C., Ponting C. P. et al. (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature **432:** 695–716
89 Aparicio S., Chapman J., Stupka E., Putnam N., Chia J. M., Dehal P. et al. (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. Science **297:** 1301–1310
90 Jaillon O., Aury J. M., Brunet F., Petit J. L., Stange-Thomann N., Mauceli E. et al. (2004) Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate protokaryotype. Nature **431:** 946–957
91 Taylor J. S., Braasch I., Frickey T., Meyer A. and Van de Peer Y. (2003) Genome duplication, a trait shared by 22000 species of ray-finned fish. Genome Res. **13:** 382–390
92 Vandepoele K., De Vos W., Taylor J. S., Meyer A. and Van de Peer Y. (2004) Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. Proc. Natl. Acad. Sci. USA **101:** 1638–1643
93 Eastman S. D., Chen T. H., Falk M. M., Mendelson T. C. and Iovine M. K. (2006) Phylogenetic analysis of three complete gap junction gene families reveals lineage-specific duplications and highly supported gene classes. Genomics **87:** 265–274
94 Heller S., Sheane C. A., Javed Z. and Hudspeth A. J. (1998) Molecular markers for cell types of the inner ear and candidate genes for hearing disorders. Proc. Natl. Acad. Sci. USA **95:** 11400–11405
95 Iovine M. K., Higgins E. P., Hindes A., Coblitz B. and Johnson S. L. (2005) Mutations in *connexin43* (*GJA1*) perturb bone growth in zebrafish fins. Dev. Biol. **278:** 208–219
96 Desplantez T., Marics I., Jarry-Guichard T., Veteikis R., Briand J. P., Weingart R. et al. (2003) Characterization of zebrafish Cx43.4 connexin and its channels. Exp. Physiol. **88:** 681–690
97 Landesman Y., Postma F. R., Goodenough D. A. and Paul D. L. (2003) Multiple connexins contribute to intercellular communication in the *Xenopus* embryo. J. Cell Sci. **116:** 29–38
98 Cicirata F., Nicotra A., Cicero D., Parenti R. and Zappalà A. (2004) Cloning and expression pattern of connexin39, a new member of the gap junction gene family isolated from the neural tube of chicken embryos. Gene **328:** 121–126
99 Becker D. L., Cook J. E., Davies C. S., Evans W. H. and Gourdie R. G. (1998) Expression of major gap junction connexin types in the working myocardium of eight chordates. Cell Biol. Int. **22:** 527–543
100 Dehal P., Satou Y., Campbell R. K., Chapman J., Degnan B., De Tomaso A. et al. (2002) The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. Science **298:** 2157–2167
101 Okamura Y., Nishino A., Murata Y., Nakajo K., Iwasaki H., Ohtsuka Y. et al. (2005) Comprehensive analysis of the ascidian genome reveals novel insights into the molecular evolution of ion channel genes. Physiol. Genom. **22:** 269–282
102 Gu X., Wang Y. and Gu J. (2002) Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. Nat. Genet. **31:** 205–209
103 Panopoulou G., Hennig S., Groth D., Krause A., Poustka A. J., Herwig R. et al. (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
104 Theissen G. (2002) Secret life of genes. Nature **415:** 741
105 Kumar S., Tamura K. and Nei M. (2004) MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief. Bioinform. **5:** 150–163