Marine Genomics 11 (2013) 17–26

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

# Phylogenetic revision of the claudin gene family

David A. Baltzegar, Benjamin J. Reading, Emily S. Brune, Russell J. Borski *

*Department of Biology, North Carolina State University, Raleigh 27695-7617, NC, USA*

ARTICLE INFO

*Article history:*
Received 1 February 2013
Received in revised form 8 May 2013
Accepted 8 May 2013

*Keywords:*
Claudin
Phylogeny
Vertebrates
Zebrafish
*Danio rerio*

ABSTRACT

Claudins are four-transmembrane proteins acting to collectively regulate paracellular movement of water and ions across cellular tight junctions in vertebrate tissues. Despite the prominence of zebrafish (*Danio rerio*) as a developmental model and the existence of an annotated genome, the diversity and evolutionary history of these claudins, with respect to other vertebrate groups, is poorly described. In this study, we identify 54 zebrafish claudins, including 24 that were previously unreported, and infer homology of the encoded polypeptide sequences with other vertebrate claudin groups using Bayesian phylogenetic analysis. In this analysis, 197 vertebrate claudin and claudin-like proteins were classified into discrete 'superclades' of related proteins. Based on these groupings, an interim reclassification is proposed, which will resolve ambiguity in the present nomenclature of several vertebrate models. Fifty-two of the 54 identified claudins were detected in cDNA preparations from whole, adult zebrafish, and 43 exhibited distinct tissue expression profiles. Despite prolific expansion of the claudin gene family in teleost genomes, these claudins can still be broadly separated into two functional groups: (1) "classic" claudins that characteristically contain an equal number of opposing, charged residues in the first extracellular loop (ECL1) and (2) "non-classic" claudins that typically have an ECL1 containing a variable number of charged residues. Functional analysis of these groups indicates that 'classic' claudins may act to reduce overall paracellular permeability to water and dissolved ions, whereas 'non-classic' claudins may constitute pores that facilitate selective ion permeability.



## 1. Introduction

Claudins are transmembrane proteins governing the formation of cellular tight junctions. In the interstitial matrix, claudins interact with other claudins or tight junction proteins (e.g., occludin) to regulate paracellular permeability. The specific permeability properties are conferred both by the type and abundance of constituent claudins that span the paracellular space (Krause et al., 2008; Van Itallie and Anderson, 2006). As compartmentalization into microenvironments of distinct ionic composition is critical to biological systems, it is not surprising that claudins are now identified as vital to normal vertebrate development and homeostasis: in the embryo (Hardison et al., 2005; Münzel et al., 2011; Siddiqui et al., 2010), within lumens and the blood–brain barrier (Abdelilah-Seyfried, 2010; Cheung et al., 2011; Jeong et al., 2008; Zhang et al., 2010), components of ion and osmoregulation (Bagherie-Lachidan et al., 2008; Le Moellic et al., 2005; Nilsson et al., 2007; Ohta et al., 2006; Tipsmark et al., 2008a), and influencing carcinoma and disease (D'Souza et al., 2005; Hewitt et al., 2006; Satake et al., 2008; Winkler et al., 2009).

Zebrafish (*Danio rerio*) is a prominent vertebrate model, particularly in the fields of developmental biology and physiology, however its full complement of claudins is unknown. Ongoing genome assembly and annotation revisions have created spurious or redundant records, thus making transcriptomic profile analysis difficult despite availability of high-density array platforms. Additionally, the current nomenclature for zebrafish claudins is ambiguous, comprising both alphabetical and numerical designations, which only partially reflects homology to claudins from other taxa. This ambiguity is compounded by erroneous classifications in other groups. For example, zebrafish claudin j (*cldnj*) is essential to the formation of the otolith (hearing) during development (Hardison et al., 2005). This gene is clearly an ortholog of pufferfish (*Takifugu rubripes*) claudin 6 (*cldn6*), yet neither may be related to human claudin 6 (*CLDN6*; GenBank UGID:1293154).

Conventionality of gene nomenclature is essential, albeit dynamic and rooted on the persistent accumulation of data (Povey et al., 2001). The same safeguards that prevent variability and confusion of gene nomenclature, however also promote obsolescence. As examples, Kollmar et al. (2001) identified 11 zebrafish claudins with no ortholog in mammals, described as claudins *a–k*. In 2004, genome analysis of the pufferfish yielded a staggering 56 claudins, presumably the result of gene and genome-wide duplication. Those with no clear homology to mammalian claudins were classified within new numeric groups (Loh et al., 2004). The largest expansion occurred at the claudin 4 loci: human *CLDN4* putatively shares a common ancestor to 13 discrete fish claudins (*cldn27a–d*, *cldn28a–c*, *cldn29a–b*,

* Corresponding author at: Department of Biology, North Carolina State University, Campus Box 7617, Raleigh 27695-7617, NC, USA. Tel.: +1 919 515 8105.
*E-mail address:* russell_borski@ncsu.edu (R.J. Borski).

1874-7787/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.margen.2013.05.001

*cldn30a–d*) (Loh et al., 2004). Currently, the same theoretical ortholog from a "non-model" fish could be classified as *cldn4*, *cldnd*, or *cldn29a*. From an evolutionary perspective, the expansion of the claudin gene family provides a unique opportunity to model the possible fates of duplicated genes. Yet, this task remains difficult without comprehensive reclassification of vertebrate claudins.

In this study, we identified 54 zebrafish claudins, of which 24 were previously undescribed (novel) in the annotated assembly (*Zv9*). Shared evolutionary history was inferred using human, mouse (*Mus musculus*), frog (*Xenopus tropicalis*), and pufferfish claudins using Bayesian phylogenetic inference and genome synteny. The mRNA expression of identified claudins was verified in preparations from whole zebrafish. Tissue specific expression of claudin genes in zebrafish is also reported. Since standardization of gene nomenclature is requisite to a unified knowledge base derived from both model and non-model organisms, we propose an interim reclassification scheme for all major claudin groups, reflective of common evolutionary descent. This reclassification provides a much needed reference for the study of claudin function across taxa.

## 2. Materials and methods

### 2.1. Identification of putative claudins

Candidate zebrafish claudins were identified by review of accessioned records available through NCBI *Gene* (www.ncbi.nlm.nih.gov/gene; search criteria: *Danio rerio* claudin, in March of 2010). These records contained both validated and provisional claudins, hypothetical loci containing *claudin-like* domains (*pfam00822*: PMP22_Claudin; PMP-22/EMP/MP20/Claudin family), and other closely related genes (claudin domain containing 1 [*cldnd1*], peripheral myelin protein 22 [*pmp22*], lens intrinsic membrane protein [*lim*], calcium channel voltage-dependent gamma subunits 1–8 [*cacng*]). Our preliminary search was refined by BLASTp search (of the translated peptide sequences) of the *T. rubripes* and *M. musculus* non-redundant protein databases (NCBI). Sequences with a Claudin BLASTp match or containing the *pfam00822* PMP/Claudin domain were selected for further study. Final candidate genes were selected by peptide sequence alignment using *ClustalX* (Thompson et al., 1997) and by preliminary phylogeny analysis with human, mouse, *X. tropicalis*, and *T. rubripes* claudin, *cldnd1*, *pmp22*, and *cacng2* sequences (parameters: generations = 2 million, sample frequency = 2000, among-site variation = equal (fixed), amino acid rate matrix = Poisson, burnin = 950). The complete list of sequences selected for analysis (total = 197) is provided as Supplemental information (Table S2).

### 2.2. Phylogeny and genomic synteny comparisons

Bayesian phylogenetic analysis was performed using four models for amino acid substitution: *Poisson* (Bishop and Friday, 1987), *Blosum62* (Henikoff and Henikoff, 1992), *WAG* (Whelan and Goldman, 2001), and the *Equalin* model, an F81 model variant (Felsenstein, 1981). All other analysis parameters were held constant (generations = 50 million, sample frequency = 10,000, among-site rate variation = equal [fixed], burnin = 1250). Analysis was performed using MrBayes (v3.1.2) on TeraGrid computing accessible through the Cyberinfrastructure for Phylogenetic Research (CIPRES) portal, available online at http://www.phylo.org/portal2/home.action (Huelsenbeck and Ronquist, 2001; Miller et al., 2010). Consensus trees from the Bayesian analysis were tested as "user trees" for significant differences in log likelihood using the Shimodaira–Hasegawa (S–H) test for alternate evolutionary hypotheses in TREE-PUZZLE (Schmidt et al., 2002; Shimodaira and Hasegawa, 1999). The S–H test for log likelihood testing was performed with the following models: *Blosum62*, *WAG*, *Dayhoff* (Dayhoff et al., 1978), and *JTT* (Jones et al., 1992).

Genomic synteny comparisons were performed using the following assemblies available through *Ensembl* (collaboration of *EMBL-EBI* and the Wellcome Trust Sanger Institute, available at http://www.ensembl.org): Fugu [*T. rubripes*]—International Fugu Genome Consortium version 4 [June, 2005]; human [*Homo sapiens*]—Genome Reference Consortium assembly version 37 [February, 2009]; mouse [*M. musculus*]—Mouse Genome Sequencing Consortium version 37 [April, 2007]; Western clawed frog [*X. tropicalis*]—Joint Genome Institute version 4.2 [November, 2009]; zebrafish [*D. rerio*]—Sanger Institute assembly version 9 [*Zv9*; April 2010]. Orthology of non-claudin genes was determined using the *Ensembl* annotated "orthologs" database and by local alignment search tools (BLAST/BLAT) available through NCBI and *Ensembl*.

### 2.3. RNA extraction and tissue expression

Zebrafish males (Tuebingen longfin strain) were used to examine mRNA expression of identified claudins. The following tissues were collected and pooled from 5 fish: eye, whole brain, gill, heart, kidney (whole), spleen, skin (whole left side, including the lateral line system), and testes. Additional males were used for preparation of whole fish total RNA. Collected tissue was preserved in RNAlater (Ambion) at 4 °C overnight before bead homogenization with RNAzol RT (Molecular Research Center) buffer. Total RNA was extracted by manufacturer's protocol (MRC). Following extraction, DNA contamination was removed by DNAse-I treatment using a Turbo DNA-free Kit (Ambion). Before cDNA synthesis, total RNA from all tissues was quantified by $A_{260}$ absorbance using a NanoDrop (ND1000) spectrophotometer. RNA quality was assessed by 18S and 28S ribosomal band integrity after gel electrophoresis. One microgram of total RNA for each tissue (or from whole fish) was reverse-transcribed with random primers using a High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). The cDNA reactions were diluted 1:6 prior to PCR amplification.

Primer pairs for 54 putative *D. rerio* claudins were designed from accessioned NCBI *Gene* sequences using Vector NTI software (Lu and Moriyama, 2004). Additionally, the housekeeping gene β-actin 1 (*bactin1*) was amplified as a positive control. A complete list of primers and annealing temperatures is provided in Table S4 (Supplemental information): amplicon size range = 400–500 bp, annealing temperature range = 54–63 °C. All PCR reactions were performed using *Taq* DNA polymerase and 10× buffer (Fisher Scientific). Reactions (25 μL) contained the following: 1× Buffer A, 0.2 mM dNTP mix, forward and reverse primers (0.4 μM each), DNA polymerase (0.3 U/μL), 1.3 μL of diluted cDNA (~11 ng), and nuclease-free water (Sigma-Aldrich; to volume). The PCR cycling parameters were as follows: (1 cycle) 95 °C for 2 min; (40 cycles) 95 °C for 30 s, 54–63 °C for 30 s, 72 °C for 45 s; (1 cycle) 72 °C for 5 min, 4 °C holding.

To validate successful PCR of the targeted claudin, amplification was first performed using whole-fish cDNA and these were submitted for sequencing. All PCR reactions were cleaned by Qiaquick PCR purification columns (Qiagen) and concentrated as a 2× elution. Samples were submitted to the University of Chicago CRC-DNA sequencing facility with the forward primer (Applied Biosystems 3730XL 96-capillary sequencer). Sequence chromatograms were identified by BLASTn search to accessioned *D. rerio* claudins (non-redundant database; Organism = *Danio rerio* [txid7955]; August 20th, 2010; e-values: 0.0–$7e^{-35}$; % identity = 76–100). Two poor-quality sequences, LOC557209 and LOC794676, were identified by alignment using Blast2align (NCBI; 77–82% identity).

For the tissue expression profile, amplification was performed simultaneously with cDNA derived from discrete tissue preparations. All templates were normalized by starting total RNA concentration (1 μg) and verified by *bactin1* amplification. The reaction and cycling parameters were as stated previously. After amplification, the