Institute of Veterinary Pathology

Department of Veterinary Medicine

Freie Universität Berlin

Evolutionary Diversity of *CLCA* Genes Among Birds and Mammals

Thesis submitted for the fulfillment of a

Doctor of Philosophy (Ph.D.) degree in Biomedical Sciences

at the

Freie Universität Berlin

submitted by

Florian Bartenschlager

Veterinarian from Munich

Berlin 2023

Journal-No.: 4391

Printed with permission of the Department of Veterinary Medicine

of the Freie Universität Berlin

Dean:	Univ.-Prof. Dr. Uwe Rösler
Supervision:	PD. Dr. Lars Mundhenk
First Reviewer:	PD. Dr. Lars Mundhenk
Second Reviewer:	Univ.-Prof. Dr. Mahtab Bahramsoltani
Third Reviewer:	Univ.-Prof. Dr. Salah Amasheh

Descriptors (according to CAB-Thesaurus):

mammals, man, pigs, mice, cats, chickens, ostriches, quails, turkeys, genetics, genomics, biochemistry, molecular biology, genetic analysis, genetic markers, genetic models, evolution, skin, keratinocytes

Day of Doctorate: 07.02.2023

Dedicated to

my family

Contents

List of figures .... **VII**

List of abbreviations .... **VIII**

Introduction **1**

1.1 CLCA molecules in biomedical research 2

1.1.1 The role of CLCA molecules in inflammatory airway diseases 2

1.1.2 The role of CLCA molecules in cancer 4

1.1.3 The association of CLCA members with other pathophysiological conditions 6

1.2 Principles of gene homology and molecular evolution 6

1.2.1 Evolution by gene amplification due to duplication 6

1.2.2 Gene paralogues, orthologues and orthologue conjecture 8

1.2.3 Methods for analysis of molecular evolution 9

1.3 Genetic diversity of the *CLCA* gene family 9

1.3.1 The *CLCA* gene locus in mammals 11

1.3.2 *CLCA* homologous genes in other taxa 12

1.4 Cloned members, cell specific protein expression patterns and biochemical properties of mammalian CLCAs 13

1.4.1 CLCA1: the CLCA of goblet cells 13

1.4.2 CLCA2: the CLCA of keratinocytes 14

1.4.3 CLCA3: the ubiquitous CLCA 15

1.4.4 CLCA4: the CLCA of enterocytes 16

1.5 Scientific questions and hypotheses 17

1.5.1 Is the exon-intron structure of galline *CLCA1* and *2* similar to those of mammalian clusters *CLCA1-4*? 17

1.5.2 Are the protein architecture and the biochemical properties of galline CLCA1 a combination of those of mammalian CLCA1, 3 and 4? 17

1.5.3 Is the cellular expression pattern of galline CLCA1 a combination of those of mammalian CLCA1, 3 and 4? 17

1.5.4 Are the protein domain architecture and the biochemical properties conserved among galline CLCA2 and the mammalian orthologues? 17

1.5.5 Are the cell type specific expression patterns of gCLCA2 and mammalian CLCA2 conserved? 17

1.5.6 Which characteristics might hypothetical ancestral CLCA proteins have had? 17

Own research publications in scientific journals with peer-review 18

1.6 Evolutionarily conserved properties of CLCA proteins 1, 3 and 4, as revealed by phylogenetic and biochemical studies in avian homologues 18

1.7 Genomic, biochemical and expressional properties reveal strong conservation of the *CLCA2* gene in birds and mammals 45

Concluding discussion and outlook 72

1.8 A hypothetical ancestral *CLCA1* gene diversified in mammals but not in birds 72

1.9 The strong conservation of CLCA2 in keratinocytes implicates an important function in these cells 74

1.10 The chicken appears to be a promising model organism for functional CLCA2 investigations 76

1.11 Outlook and limitations of this cumulative thesis 77

1.12 Conclusions 80

Summary 81

Zusammenfassung 83

References 85

Appendix 107

List of own publications regarding this topic 111

Acknowledgements 112

Funding 114

Conflict of interest 115

Declaration of originality 116

List of figures

Figure 1. Phylogenetic relationship of putatively functional feline, human, porcine, equine, bovine, rat and murine CLCA proteins 10

Figure 2. Illustration of the CLCA gene locus from cat, human, pig, horse, cow, rat and mouse 12

Figure 3. Relationship between published rat CLCA amino acid sequences in the literature compared to those found in the NCBI Genbank 107

Figure 4. Relationship between published bovine CLCA amino acid sequences in the literature compared to those found in the NCBI Genbank 109

List of abbreviations

Aa	**A**mino **a**cid
Bp	**B**ase **p**air
Bsr	**B**eta **s**heet **r**ich
C-	**C**arboxy-
cDNA	**C**omplementary **d**eoxyribo**n**ucleic **a**cid
CF	**C**ystic **f**ibrosis
CLCA	**C**h**l**oride channel regulators, **c**alcium **a**ctivated
CpG	**C**ytosine–**p**hosphate-**g**uanine
COPD	**C**hronic **o**bstructive **p**ulmonary **d**isease
CRC	**C**olo**r**ectal **c**arcinoma
CTBP1	**C**-**t**erminal **b**inding **p**rotein **1**
DNA	**D**eoxyribo**n**ucleic **a**cid
EST	**E**xpressed **s**equence **t**ag
FAK/ERK	**F**ocal **a**dhesion **k**inase**/E**xtracellular signal-**r**egulated **k**inase
Fn III	**F**ibro**n**ectin type **III**
IL13	**I**nter**l**eukin **13**
JTT	**J**ones-**T**aylor-**T**hornton
MAPK13	**M**itogen-**a**ctivated **p**rotein **k**inase **13**
MIDAS	**M**etal **i**on **d**ependent **a**dhesion **s**ite
mRNA	**M**essenger **r**ibonucleic **a**cid
Muc5AC	**M**ucin **5AC**
N-	Amino-
NAHR	**N**on-**al**lelic **h**omologous **r**ecombination
NCBI	**N**ational **C**enter of **B**iotechnology **I**nformation
NHEJ	**N**on-**h**omologous **e**nd **j**oining
ODF2L	**O**uter **D**ense **F**iber of Sperm Tails **2**-**L**ike

PFAM	**P**rotein **Fam**ilies database
PI3/Akt	**P**hospho**i**nositide **3/Ak** strain **t**ransforming
RACE-PCR	**R**apid **a**mplification of **c**DNA-**e**nds-**p**olymerase **c**hain **r**eaction
SH3GLB1	**SH3**-Domain **G**RB2-**L**ike Endophilin **B1**
STAT6	**S**ignal **t**ransducer and **a**ctivator of **t**ranscription **6**
TMC	**T**rans**m**embrane domain in the **c**arboxy-terminal cleavage product
TMEM16A	**T**rans**mem**brane protein **16A**
Tp53	**T**umor **p**rotein **53**
VWA	**V**on **W**illebrand factor type **A**
WGD	**W**hole-**g**enome **d**uplication

Introduction

The mammalian *chloride channel regulators, calcium activated* (*CLCA*) gene family is highly complex and diverse in terms of gene amplification and inactivation events. Its homologues are often redundantly expressed in tissues and cells of many species. Based on their homology, mammalian *CLCA* genes belong to one of four distinct genetic clusters, termed *CLCA1* to *CLCA4.* All *CLCA* genes are located in a single locus, which is flanked by *Outer Dense Fiber of Sperm Tails 2-Like* (*ODF2L)* and *SH3-Domain GRB2-Like Endophilin B1* (*SH3GLB1)* genes (Plog et al. 2009; Mundhenk et al. 2018). From a protein biochemical perspective, mammalian CLCAs are cleaved glycoproteins of approximately 900 amino acids (aa) in length that share a similar domain architecture (reviewed by Patel et al. 2009). In the larger amino- (N-) terminal cleavage product, a signal sequence for entry into the endoplasmic reticulum is prepended to a proteolytically active N-CLCA (PFAM identifier: pfam08434), a von Willebrand factor type A (vWA, PFAM identifier: pfam13519) and a beta sheet rich (bsr, Patel et al. 2009) domain. A fibronectin type III domain (fn III, PFAM identifier: PF00041) is located in the smaller carboxy- (C-) terminal cleavage product. Depending on the presence of a transmembrane domain in the C-terminal cleavage product (TMC), two types of CLCA proteins can be distinguished. CLCA1 and some 3 proteins are lacking a TMC. Thus, it has been experimentally shown for CLCA1 and a murine CLCA3 but only predicted for some other CLCA3 that these proteins are fully secreted (Gibson et al. 2005; Mundhenk et al. 2006; Huan et al. 2008; Patel et al. 2009; Mundhenk et al. 2018). In contrast, members of CLCA clusters 2 and 4 possess a TMC that anchors the C-terminal cleavage product in the plasma membrane (Elble et al. 2006; Bothe et al. 2008; Braun et al. 2010; Plog et al. 2012a).

Associations with diseases, in particular with cancer and inflammatory respiratory diseases in human and veterinary medicine have been made, however, the physiological and pathophysiological functions of this gene family and their individual members are still unknown (Patel et al. 2009; Winpenny et al. 2009; Hu et al. 2019; Liu and Shi 2019). Generation of various murine knockout models regularly failed to uncover consistent (patho)physiological mechanisms and functions of these genes *in vivo* under unchallenged conditions (Robichaud et al. 2005; Long et al. 2006; Patel et al. 2006; Mundhenk et al. 2012; Erickson et al. 2015; Erickson et al. 2018; Nyström et al. 2018). Considering the complexity and redundancy of the mammalian CLCA members, in particular in the murine species with eight *CLCA* genes, mutual compensation of a knocked out homologue by another in the same anatomical location might be a reasonable explanation for a lack of phenotype in knockout mice.

But is this complexity and diversity of these genes a unique entity of mammals? And what is the evolutionary background of this complexity? To address these questions, data on CLCA obtained from evolutionarily more distant species are needed, but have not yet been available.

Recently, a *CLCA* locus with only two *CLCA* genes was predicted for the chicken (*Gallus gallus domesticus*, Mundhenk et al. 2018). The predicted protein sequence of the first galline *CLCA* (*gCLCA1*) is characterized by a close genetic distance to mammalian CLCA proteins 1, 3 and 4. In contrast, the aa sequence encoded by the second galline *CLCA2* (*gCLCA2*) gene forms a monophyletic group only with mammalian CLCA2 (Mundhenk et al. 2018).

This cumulative thesis aimed to characterize these galline CLCA members to provide evidence for the most likely evolutionary developmental scenario of the *CLCA* gene family among birds and mammals. By this, its taxonomical relationships and indispensable, conserved elements were elucidated, which can be object of comparative and functional *in vitro* investigations in the future. Furthermore, initial data were provided that assess the potential of the chicken as a model organism for functional CLCA investigations.

1.1 CLCA molecules in biomedical research

The discovery of the first *CLCA* gene family member dates back over 30 years now (Ran and Benos 1991; Zhu and Pauli 1991) and since then, these genes have been detected in the genome of various vertebrates, including humans as reviewed by Patel and colleagues (Patel et al. 2009). Interestingly, their parallel but independent discovery in two laboratories with completely distinct research focuses – namely research in airway anion channels and cancer metastasis – emphasized a pleiotropic nature of CLCA molecules from the beginning on and projects onto two main groups of diseases they are strongly associated with: inflammatory airway diseases and cancer.

Unfortunately, and similar to many other genes and gene families, the designation of early discovered *CLCA* members was according to their order of discovery or specific properties. Multiple revisions of the nomenclature in the past caused confusion and complicated the association of published information with specific members of the *CLCA* family. In this thesis, the current nomenclature as it was found in the National Center of Biotechnology Information (NCBI) Genbank (state of the information: 01.10.2020) is used for all *CLCA* homologues. If nomenclature impeded the assignment of a *CLCA* member from the literature to one of the gene clusters, phylogenetic analyses assessed their affiliation and are provided at the respective section and the appendix. The members to whom this applies are discussed in the corresponding chapters, but nomenclature is used as it was listed in the NCBI Genbank.

1.1.1 The role of CLCA molecules in inflammatory airway diseases

Among all four *CLCA* clusters, *CLCA1* seems to play the most important role in inflammatory airway diseases. On a genetic basis, it was reported that alterations in the human *CLCA1* (*hCLCA1*) gene are associated with the susceptibility of asthma (Kamada et al. 2004). Furthermore, several *hCLCA1* haplotypes were found, which are overrepresented in human

patients with chronic obstructive pulmonary disease (COPD, Hegab et al. 2004). In cystic fibrosis (CF), which is fatal mainly due to pathological changes in the lung (Ratjen et al. 2015), the human *CLCA* locus, however, appears to be primarily associated with the gastrointestinal phenotype of CF patients (Ritzka et al. 2004). In 2010, van der Doef and colleagues identified a single nucleotide polymorphism in *hCLCA1* to be overrepresented in CF patients with meconium ileus and a severe CF phenotype (Van Der Doef et al. 2010).

In terms of its role in airway inflammation, CLCA1 might be involved in the recruitment of inflammatory cells into lung tissues. For example, two different murine *CLCA1* knockout models had fewer inflammatory cells in the lung tissue compared to wild type controls under challenged conditions with ovalbumin and lipopolysaccharide or *Staphylococcus aureus* (Long et al. 2006; Dietert et al. 2014). Conversely, mice with intranasal instilled *CLCA1* plasmids and by this, overexpressed CLCA1, had an increased level of inflammatory cell infiltration in the lungs compared to control animals treated with an irrelevant plasmid (Mei et al. 2013). Besides these observations, *in vitro* experiments using murine and porcine alveolar macrophages and human alveolar macrophage like cells revealed a proinflammatory activation of these cells in the presence of CLCA1 protein (Ching et al. 2013; Erickson et al. 2018; Keith et al. 2019).

It is well known that CLCA1 is strongly upregulated in the airways of asthmatic patients. This trait of upregulation in asthmatics was documented for various mammalian species, including humans (Hoshino et al. 2002; Toda et al. 2002; Wang et al. 2007; Woodruff et al. 2007), horses (Anton et al. 2005; Gerber et al. 2009), cats (Erickson et al. 2020) and murine models of asthma (Nakanishi et al. 2001; Zhou et al. 2001; Hashimoto et al. 2004; Kuperman et al. 2005; Thai et al. 2005; Nakano et al. 2006; Mei et al. 2013). Allergic airway diseases like asthma are characterized by excessive mucus production and mucus plugging (Rogers 2000) in an Interleukin 13 (IL13)-rich microenvironment (Corren 2013). The mucus protein component Mucin 5AC (Muc5AC) appears to be an essential factor for this plugging (Evans et al. 2015). The upregulation of this protein is regularly observed in human asthma patients (Wang et al. 2007; Woodruff et al. 2009) and murine asthma models (Thai et al. 2005; Nakano et al. 2006; Mei et al. 2013). In this context, CLCA1 might be involved in the upregulation of Muc5AC and might play a key factor of mucus cell metaplasia in a complex, but intriguing way. Initially IL13, a key cytokine in asthma (Corren 2013), seems to upregulate CLCA1 via the signal transducer and activator of transcription 6 (STAT6) transcription factor (Kuperman et al. 2002; Nakano et al. 2006). Subsequently, CLCA1 seems to interact with transmembrane protein 16A (TMEM16A), which is an IL13 – STAT6 inducible calcium activated chloride channel (Qin et al. 2016), in two ways: First, CLCA1 appears to activate TMEM16A in an auto/paracrine manner, since recent studies found the vWA domain of the secreted N-terminal cleavage product of CLCA1 in association with TMEM16A where it activated it by its stabilization at the plasma

membrane (Sala-Rabanal et al. 2015; Berry and Brett 2020). Second, CLCA1 as well as its paralogue CLCA2 was capable of activating TMEM16A by an increase of intracellular calcium levels (Sharma et al. 2018). Plasma membrane depolarization by the active TMEM16A then seems to activate mitogen-activated protein kinase 13 (MAPK13), which finally enables Muc5AC expression (Alevy et al. 2012). Despite these promising initial findings originating from *in vitro* experiments, further evidence from *in vitro* and *in vivo* models is needed to support this hypothesized mechanism of mucus plug development in allergic airway diseases. Curiously, there are no data from murine *CLCA1* knockout models that support the function of CLCA1 as a driver of mucus cell metaplasia (Robichaud et al. 2005; Long et al. 2006; Patel et al. 2006; Mundhenk et al. 2012).

An interesting finding that has not yet been transferred to airway mucins was the capability of CLCA1 to maintain mucus homeostasis in the colon of mice by its proteolytic activity (Nyström et al. 2018). This was achieved due to proteolysis of the N-terminal part of Mucin 2, which was mediated by CLCA1 (Nyström et al. 2019). Interestingly, no significant phenotype was observed in *CLCA1* knockout mice with respect to mucin deviation or after induction of colitis. This might be explained by a potential compensation of the loss of CLCA1 by other CLCA representatives present in the gut, such as CLCA4 (Erickson et al. 2015; Nyström et al. 2019). Whether this physiological function of CLCA1 applies to other mammals and if it is transferable to other mucus covered body surfaces remains to be elucidated.

1.1.2 The role of CLCA molecules in cancer

Members of *CLCA* clusters 1, 2 and 4 have been frequently associated with cancer arising from epithelial tissues. For example in human colorectal carcinoma (CRC), a type of cancer with the third highest incidence and second highest mortality of all human cancer types worldwide (Sung et al. 2021), CLCA1, 2 and 4 were weaker expressed in neoplastic cells compared to the normal or tumor adjacent tissue (Bustin et al. 2001; Yang et al. 2013; Yu et al. 2013; Yang et al. 2015; Li et al. 2017; Wei et al. 2020). This finding seems to have an important impact on the cancer associated statistical properties, since it was shown that overall survival is positively correlated with the expression level of these three CLCAs (Yang et al. 2015; Chen et al. 2019; Pan et al. 2019). Furthermore, *in vitro* experiments support this data: a negative correlation between the expression levels and the invasiveness and migration of different CRC tumor cell lines such as SW610 and LoVo in culture was shown for CLCA1 and/or CLCA4 (Li et al. 2017; Chen et al. 2019).

In human breast cancer, CLCA2 might be the important actor. Similar to CRC, *CLCA2* was downregulated in this tumor type (Gruber and Pauli 1999; Li et al. 2004). This finding was in line with *in vitro* experiments, where CLCA2 was expressed in the non-transformed cell line

MCF10A and the nontumorigenic cell line MDA-MB-453, but not in tumorigenic cell lines MDA-MB-231, MDAMB-435, MDA-MB-468, and MCF7. Introduction of CLCA2 in MDA-MB-231 and MDA-MB-435 cells, however, decreased the invasiveness of these cells (Gruber and Pauli 1999). These findings were supported by similar experiments with comparable outcomes using the murine CLCA2 orthologue (Beckley et al. 2004). Taken together, these observations suggest a negative correlation of CLCA2 with tumor malignancy and by this, a potential function of this protein as a tumor suppressor. Supportive for this finding was an induction of *CLCA2* by the central tumor suppressor Tumor protein 53 (Tp53), detected in multiple cell lines and conditions. As a consequence, malignancy properties, such as cell migration and invasion were decreased and anti-tumorous reaction patterns like the induction of replicative senescence were demonstrated (Walia et al. 2009; Sasaki et al. 2012; Tanikawa et al. 2012).

Besides these two important types of cancer, CLCA members 1, 2 or 4 have also been associated with other tumor types and malignancies. In ovarian cancer cell lines, CLCA1 seems to be elevated (Musrap et al. 2015), but low CLCA1 expression in pancreatic duct adenocarcinoma might be associated with a poor disease-free survival (Hu et al. 2018). In cervical cancer, CLCA2 suppressed proliferation, migration and invasion of cervical cancer cell lines and CLCA2 expression seems to be correlated with tumor stage and size (Zhang et al. 2021). Furthermore, CLCA2 upregulation in cervical cancer appears to be associated with a longer overall and recurrence free survival time following surgery (Zhang et al. 2021). In men with metabolic syndrome and prostate cancer, CLCA2 repression by the transcriptional co-repressor C-terminal binding protein 1 (CTBP1) might be responsible for increased malignancy (Porretti et al. 2018). In nasopharyngeal carcinoma, CLCA2 was negatively correlated with proliferation, migration, invasion and epithelial-mesenchymal transition. The authors concluded that these findings were associated with inhibition of the Focal adhesion kinase/Extracellular signal-regulated kinase (FAK/ERK) pathway (Qiang et al. 2018). CLCA2 was proposed as a tumor marker to distinguish between squamous cell carcinoma and adenocarcinoma of the lung (Shinmura et al. 2014). In hepatocellular carcinoma, CLCA4 expression was negatively correlated with tumor size, vascular invasion and the clinical stage (Liu et al. 2018). Statistical analyses of this study suggested CLCA4 as an independent predictor of overall survival and time to recurrence (Liu et al. 2018). CLCA4 expression in head and neck squamous cell carcinoma also appears to be positively correlated with prognosis (Li et al. 2020). The authors found CLCA4 to be suppressed by a type of micro-ribonucleic acid, which was abundantly present in head and neck cancer cells (Li et al. 2020). In urinary bladder carcinoma, CLCA4 was also downregulated (Hou et al. 2017). Furthermore, this study showed that CLCA4 could reduce malignancy by suppressing the Phosphoinositide 3/Ak strain transforming (PI3/AKT) pathway (Hou et al. 2017).

1.1.3 The association of CLCA members with other pathophysiological conditions

There are correlations of CLCA with other pathophysiological conditions described in the recent literature. In the human epidermis, CLCA2 seems to act as a humidity regulated protein that was found to be regulated under hyperosmolar conditions via the p38/c-Jun N-terminal kinase–activating transcription factor 2 signaling pathway (Seltmann et al. 2018). It appears to mediate cellular adhesion and by this, prevent low humidity induced cellular death (Seltmann et al. 2018). This finding might be useful in developing new therapeutics of psoriasis, a chronic skin disease. In human cardiac fibrosis, a transdifferentiation from resident, quiescent fibroblasts to active myofibroblasts is an important aspect in the pathogenesis (Gibb et al. 2020). Shao and colleagues demonstrated that epigenetic repression of *CLCA2* seems to facilitate fibroblast-myofibroblast transdifferentiation, which could contribute to cardiac fibrosis (Shao et al. 2021).

1.2 Principles of gene homology and molecular evolution

Molecular evolution of deoxyribonucleic acid (DNA) describes the development of genetic variation by mutation, including e.g. nucleotide exchanges, insertions, inversions, deletions, recombinations, conversions, and their fixation in the genome of a population by genetic drift and/or selection. An element of this process is the duplication of genes, a phenomenon that is an important driver for evolution (Ohno 1970) and described in mammalian *CLCA* genes (Mundhenk et al. 2018).

1.2.1 Evolution by gene amplification due to duplication

Accepted models that explain the mechanism of gene duplication include polyploidization, unequal crossing over, retroposition (Zhang 2003), and duplicative transposition (Hahn 2009; Magadum et al. 2013). Polyploidization describes a multiplication of homologous chromosomes at the whole-genome level (whole genome duplication, WGD). In contrast, aneuploidy just affects individual chromosomes but also designates the lack of one or more individual chromosomes. Polyploidization is well described in various classes, including plants (Rice et al. 2019) and fungi (Kellis et al. 2004). It is also documented in animals like fish (Leggatt and Iwama 2003) and amphibians (Schmid et al. 2015). However, polyploidization appears to be rather rarely under positive selection in animals (Mable 2004). Unequal crossing over can cause genetic copy number variations leading to duplicated genes which are organized as tandem repeats (Zhang 2003). Genetic duplication by retroposition of a gene-coding messenger ribonucleic acid (mRNA) requires its reverse transcription into complementary DNA (cDNA) and a subsequent genomic integration. This often happens in a gene locus that is not associated with the “original” gene (Long et al. 2003; Zhang 2003; Magadum et al. 2013). Such a process may lead to mRNA specific properties of the duplicated gene. For example, typical regulatory sequences as well as introns might be absent in genes

duplicated by retroposition. In contrast, genes duplicated by this way may exhibit a polyadenyl tail (Zhang 2003; Kaessmann et al. 2009; Magadum et al. 2013). Duplicative transposition can be mediated by either non-allelic homologous recombination (NAHR) or non-homologous end joining NHEJ (Hahn 2009). In the former, a sequence upstream of one gene is highly homologous to another one that is located downstream of it. NAHR results in duplication or loss of the gene flanked by the homologous sequences (Lee 2013). NHEJ, as a part of the physiological DNA double strand break repair, is independent of such homologous sequences (Lieber 2010).

With respect to these models and the fact that all cloned mammalian *CLCA* members possess introns and are located in line within a highly conserved gene locus, it appears not unlikely that *CLCA* duplicates evolved by unequal crossing over events or duplicative transposition. Interestingly, Gruber and coworkers found a special type of transposable element upstream the first exon of *hCLCA1*; however in contrast to typical transposable element insertions of the type found by Gruber et al, no target site duplication surrounding it was present (Gruber et al. 1998).

Once a gene duplication has occurred in the germline of an individual, it is depending on the mode of selection if and how fast it becomes fixed in the population´s genome. In this case, population is defined as a concrete number of randomly mating diploids under constant selective pressure (Innan and Kondrashov 2010). In 2010, Innan and Kondrashov reviewed accepted models of duplicate preservation in a population (Innan and Kondrashov 2010): If the new copy is under neutral selection, which means the duplication does not affect the fitness, it is likely to develop a new function, a process termed neofunctionalization (Ohno 2013). Another outcome may be subdivision of the original function among both gene copies, which is called subfunctionalization (Hughes 1994; Force et al. 1999). The time after duplication and fixation of it in a population´s genome is termed fate-determining phase. If the new copy is under positive selection during fixation, which means the duplication has a benefit for the fitness, there are three possible scenarios on the copies fate (Innan and Kondrashov 2010). Either, the copy has a beneficial effect by an immediate increase of gene dosage. As a consequence, it will be kept functional, identical to the original gene (Innan and Kondrashov 2010). Second, the duplicated gene has a shielding function against deleterious mutations of the original one and by this provides a selective advantage. Over time, a logic consequence might be inactivation of the duplicate (Innan and Kondrashov 2010). Third, the duplication per se leads to a new function of the copy that will be immediately preserved (Innan and Kondrashov 2010). If the gene to be duplicated, however, is already present as a polymorphic allele and under positive selection, the model of adaptive radiation postulates adaptive fixation of the duplication already in the pre-duplication phase (Francino 2005). In contrast, the model

of permanent heterozygote describes the following scenario: Given that two different alleles are kept by balancing selection prior to duplication and a heterozygote for a single locus comprises the same fitness as to two loci that are homozygous for a different allele, then a duplication in a gene for which the highest fitness is achieved in the heterozygous state is immediately beneficial (Spofford 1969; Proulx and Phillips 2006). When multi-allelic diversifying selection is the model of evolution, selection targets the genetic variability. In this case, gene duplications increase fitness due to a possibility of mutation and selection. As a result, the given locus is highly variable among individuals of the population (Innan and Kondrashov 2010). Finally, the dosage balance model accounts for groups of genes with optimal dosages that are dependent on each other. Loss or duplication is only tolerated in parallel and it accounts for duplications during WGDs (Innan and Kondrashov 2010).

1.2.2 Gene paralogues, orthologues and orthologue conjecture

When a gene duplication has occurred and was fixed and preserved in a population´s genome, the population may eventually be exposed to conditions that give rise to a new species. The formation of a new species is a process that is termed speciation. Depending on time of the speciation and duplication event, the duplicated genes are related to each other within and among species in defined relationships, for which a specific terminology is used. In general, gene homology describes the similarity of two or more genes due to a common ancestor. Orthologues are genes that share a single ancestral gene in the last common ancestor of two distinct species (Fitch 1970). Paralogous genes are related to each other via duplication within one lineage (Fitch 1970). More specifically, the term co-orthologues defines orthology of one or more genes of one lineage to one or more genes of another lineage due to a lineage specific duplication (Sonnhammer and Koonin 2002). Outparalogues are paralogous genes that exist as a consequence of gene duplication preceding a speciation event (Sonnhammer and Koonin 2002). In contrast, inparalogues are linage specific paralogues, which result as the consequence of gene duplication after a speciation event (Sonnhammer and Koonin 2002).

If gene sequence homology is considered in terms of possible gene functions, it is assumed that orthologous genes have a higher chance of functional equivalence to each other than paralogues do (Tatusov et al. 1997; Dolinski and Botstein 2007; Fang et al. 2010; Gabaldón and Koonin 2013; Stamboulian et al. 2020). This idea was termed orthologue conjecture (Nehrt et al. 2011; Altenhoff et al. 2012) but has been the challenged by recent findings (Nehrt et al. 2011). Especially when it comes to the prediction of gene function in newly sequenced genomes, shortcomings in the concept of orthologue conjecture can have fatal consequences for the prediction.

1.2.3 Methods for analysis of molecular evolution

Most approaches for analysis of molecular evolution are genocentric, which means they almost exclusively refer to *in silico* analyses of gene and protein sequence relationships among species and/or individuals. For example, molecular phylogeny utilizes various algorithms, based on either genetic distances (Sokal 1958; Felsenstein 1984; Saitou and Nei 1987), Maximum Parsimony (Czelusniak et al. 1990), Maximum Likelihood (Felsenstein 1981), or Bayesian Inference (Huelsenbeck et al. 2002) methods. As a result, a phylogenetic tree is generated and phylogenetic relationships among the input sequences are displayed. Besides these techniques, the comparative analysis of chromosomal gene localizations, termed synteny, is a tool for unveiling evolutionary relationships. Furthermore, a bundle of algorithms for the detection of phylogenetic signals within nucleotide or amino acid sequences that arose from selection pressures over time is available at the Datamonkey server (https://www.datamonkey.org/, Weaver et al. 2018).

However, many important traits, such as mRNA and protein expression patterns in tissues and cells or biochemical properties, have rarely been considered in such analyses. The lack of information specifying these properties might contribute to the failure of functional predictions based on the concept of orthologue conjecture in some cases and should be an integral part of any evolutionary investigation.

1.3 Genetic diversity of the *CLCA* gene family

Due to the establishment of high-throughput sequencing methods, the number of available annotated whole-genome sequences from various animal species is rising steadily and with increasing rates. As a result, numerous species with predicted *CLCA* gene sequences can be identified in the current genome databases (Kent et al. 2002; Clark et al. 2016; Cunningham et al. 2022). Despite that, the following section will exclusively cover those species, for which representatives of the *CLCA* gene family have either been cloned and fully sequenced or otherwise verified in the literature. This selection was justified by the availability of experimentally validated data from at least some of these *CLCA*s. A phylogenetic tree generated by the Maximum Likelihood method displays all putatively functional CLCA members and their phylogenetic relationship among these species based on their amino acid sequences (figure 1).

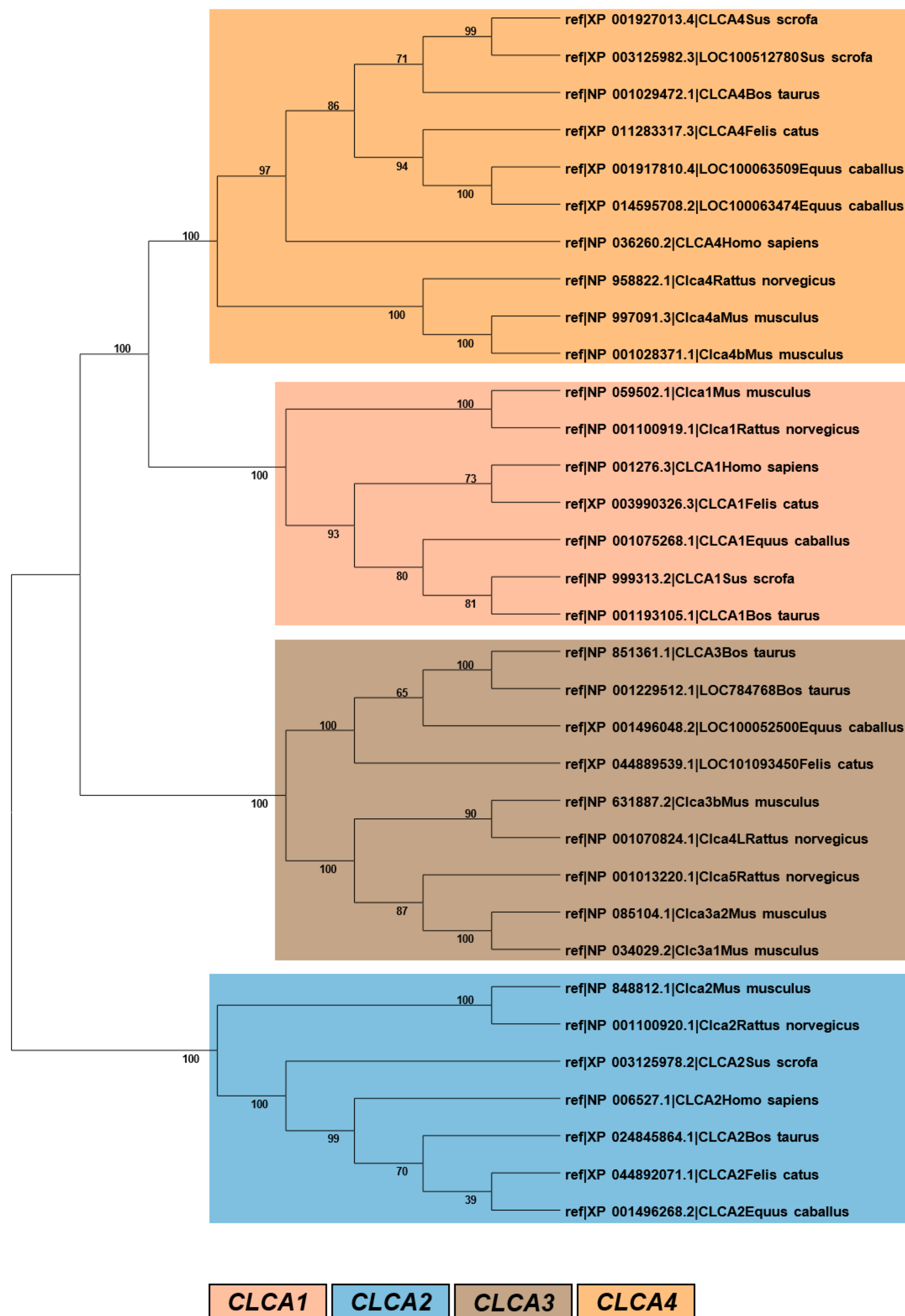


Figure 1. Phylogenetic relationship of putatively functional feline, human, porcine, equine, bovine, rat and murine CLCA proteins

The phylogenetic tree was generated using the Maximum Likelihood method and Jones-Taylor-Thornton (JTT) matrix-based model on amino acid sequences by the MEGA X software package. This analysis involved 33 amino acid sequences. CLCA sequences from predicted pseudogenes were excluded. The tree with the highest log likelihood (-17794.11) is shown. The percentage of trees in 100 bootstraps, in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.5110)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 10.21% sites). All positions containing gaps and missing data were eliminated. There were a total of 705 positions in the final dataset. Accession numbers to the NCBI Genbank sequences are provided left to the protein and scientific species name. It becomes obvious that all CLCAs can be designated to one of four distinct clusters: CLCA1 (pink), CLCA2 (blue), CLCA3 (brown) and CLCA4 (orange).

1.3.1 The *CLCA* gene locus in mammals

Based on the literature (Plog et al. 2009; Mundhenk et al. 2018) and the actual version of the NCBI Genbank (date of retrieval: 01.10.2022), there are five taxonomic orders including seven mammalian species with available experimental data for members of the *CLCA* gene family. These contain the order primates with humans, the order rodents with mice and rats, the order carnivores with cats, the order perissodactyls with horses and the order artiodactyls with pigs and cows. In all these species, the *CLCA* gene locus is flanked by the *ODF2L* and *SH3GLB1* genes and embraces up to eight *CLCA* homologues, depending on the species. The cat is the only mammal that has one putatively functional representative of the gene clusters *CLCA1* to *CLCA4* (Mundhenk et al. 2018). All other species have a single *CLCA1* and a single *CLCA2*, but differ in the copy number and intactness of *CLCA3* and *CLCA4* representatives (Plog et al. 2009; Plog et al. 2015; Clark et al. 2016; Mundhenk et al. 2018). In humans and other primates, *CLCA3* is a pseudogene, while *CLCA4* is present as a single intact gene (Mundhenk et al. 2018). Like in humans, pigs lack an intact *CLCA3*, but have a duplicated *CLCA4* (Plog et al. 2009). An interesting finding in the genome of a porcine subpopulation is the pseudogenization of one CLCA4 representative (Plog et al. 2015). Instead, horses have an intact *CLCA3* and two intact *CLCA4* genes but one putative *CLCA4* pseudogene (Clark et al. 2016). Cows and rats possess two intact *CLCA3* and one intact *CLCA4* (Mundhenk et al. 2018). Mice are the species with the most numerous *CLCA* representatives. They possess three *CLCA3* and three *CLCA4* copies, whereby one *CLCA4* is predicted to be inactivated (Patel et al. 2009). For a better understanding, the respective genotypes are illustrated in figure 2.

Prototypic mammalian *CLCA* locus

Species specific variations of the *CLCA* locus according to the NCBI Genbank

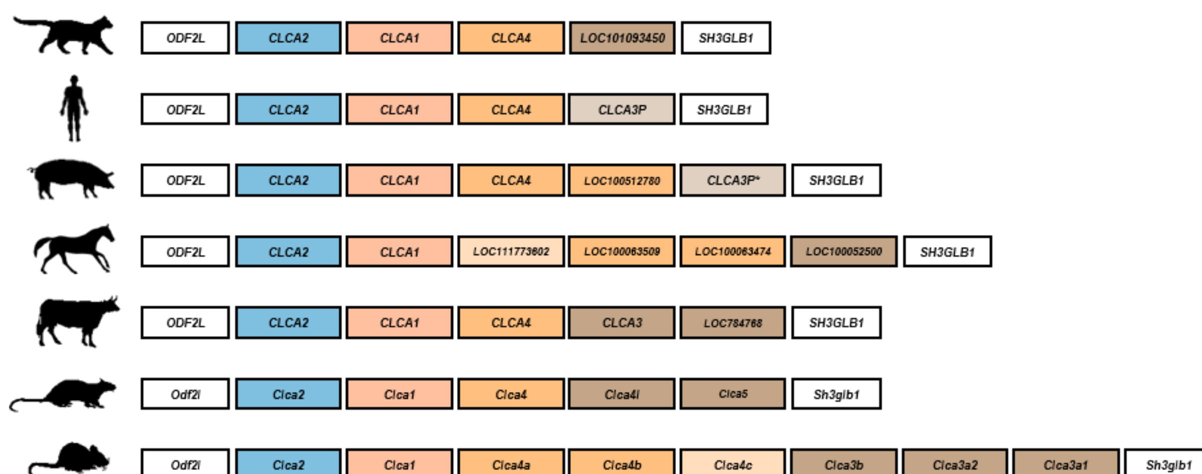


Figure 2. Illustration of the CLCA gene locus from cat, human, pig, horse, cow, rat and mouse

This figure illustrates the CLCA gene locus as it was found in the NCBI Genbank (date of retrieval: 01.10.2022). Faded orange or brown colors and/or the letter "P" indicate predicted pseudogenes. Porcine CLCA3 was not listed in the NCBI Genebank, but was included in this figure as published by Plog et al. in 2012 (*). Pink color represents the CLCA1 cluster, blue color highlights the CLCA2 cluster, brown indicates the CLCA3 cluster and orange labels the CLCA4 cluster.

1.3.2 *CLCA* homologous genes in other taxa

To date, only a single *CLCA* gene has been experimentally isolated and sequenced from a non-mammalian species (Lee et al. 2011). In this study, a 2940 base pair (bp, 940 aa) cDNA from the African clawed frog (*Xenopus laevis*) was reconstructed using rapid amplification of cDNA-ends-polymerase chain reaction (RACE-PCR). Expression analysis revealed a high abundance of mRNA in the small intestine and colon and a low abundance of mRNA in the liver (Lee et al. 2011). *In silico* analyses predicted multiple glycosylation sites and a highly hydrophobic amino acid sequence towards the C-terminus, insinuating the presence of a transmembrane domain (Lee et al. 2011).

In 2013, an *in silico*-based database research study identified the protease motif containing N-CLCA domain as a structural component of proteins in numerous bacterial and metazoan species (Lenart et al. 2013). Interestingly, the kingdoms Fungi and Plantae (plants) lacked proteins with such a domain, a finding that also applied to the genus Amoeba. The authors of this study designated all proteins containing such an N-CLCA domain as CLCA proteins, which lead to the identification of myriad CLCAs with a highly diverse protein architecture across the

tree of life. According to the study, the placozoan species *Trichoplax adhaerens* was the most ancient species with a CLCA protein domain architecture that is comparable to the known mammalian CLCAs. From N- to C-terminus, this predicted CLCA protein comprised an N-CLCA domain, a vWA domain and a fibronectin type III domain. Since the study was based on sequence retrieval from databases and the bsr domain is not assigned to a proper accession number, a bsr domain was not predicted for this representative. These findings implicate the presence of *CLCA* genes coding for a mammalian type CLCA protein since the Proterozoic eon (542 – 2500 million years ago), where the placozoan clade likely started to form (Kumar et al. 2022).

1.4 Cloned members, cell specific protein expression patterns and biochemical properties of mammalian CLCAs

Experimental data specifying the expression of CLCA protein on the cellular level, as well as *in vitro* biochemical properties are available for some of the CLCA members of the aforementioned seven mammalian species. The following sections will cover the current knowledge on cell type specific protein expressions of CLCAs. It will become obvious that most of the CLCAs are restricted to a low amount of cell types with a slight species dependent variance. Members of clusters CLCA1, CLCA2 and CLCA4 belong to this group. In contrast, others are widely expressed by many cell types, such as members of CLCA cluster 3. All references on which these findings are based will be provided in detail in the corresponding sections. Since the cell composition of organs in higher vertebrates is often highly complex, different CLCA homologues, which were expressed by distinct cells that are found in a single organ. This can cause an overlapping overall expression pattern. For example, in the feline trachea, CLCA1 is found in goblet cells and CLCA3 is expressed by tracheal epithelial cells and cells of submucosal glands (Erickson et al. 2020).

It has to be emphasized at this point that the studies on protein expression that are cited in the following chapters varied in terms of the study setups. As a consequence, systematical and comprehensive data regarding the expression in certain anatomical localizations are lacking for some of the species.

1.4.1 CLCA1: the CLCA of goblet cells

The first mammalian *CLCA1* member, *hCLCA1*, was isolated by Gruber and coworkers in 1998 from a human genomic library (Gruber et al. 1998). Since then, orthologues from mouse (Komiya et al. 1999), pig (Gaspar et al. 2000), horse (Anton et al. 2005) and cat (Erickson et al. 2020) were cloned. In summary, CLCA1 protein was predominantly detected in goblet cells of various mucus membrane surfaces in pigs, mice, cats, humans, and horses (Hoshino et al. 2002; Leverkoehne and Gruber 2002; Anton et al. 2005; Range et al. 2007; Plog et al. 2009;

Erickson et al. 2020). Interestingly, CLCA1 protein was also found in airway submucosal glands of pigs, mice and horses (Leverkoehne and Gruber 2002; Anton et al. 2005; Plog et al. 2009), but not in cats (Erickson et al. 2020). Submucosal glands in other organs, like those of the murine uterus (Leverkoehne and Gruber 2002) or those of the porcine alimentary tract (Plog et al. 2009) stained positive for CLCA1 using immunohistochemistry. No data for the cell type specific protein expression of CLCA1 was found in the literature for rats and cows. *In vitro* experiments investigating biochemical properties of the human and murine orthologues demonstrated that CLCA1 was a fully secreted protein (Gibson et al. 2005; Mundhenk et al. 2006). The secretion of the N-terminal cleavage product was shown for the porcine, feline, and equine orthologue, however, experiments investigating the fate of the carboxyterminal cleavage product are lacking (Range et al. 2007; Plog et al. 2009; Erickson et al. 2020). Furthermore, it was experimentally shown for human (Gruber et al. 1998), murine (Mundhenk et al. 2006), porcine (Plog et al. 2009) and equine (Range et al. 2007) CLCA1 orthologues that these proteins were cleaved and glycosylated. For the cat, cleavage was also suggested (Erickson et al. 2020). Experiments demonstrating the full secretion of both CLCA1 cleavage products of the rat and bovine orthologues are lacking. This also applies for cleavage and glycosylation of the rat and bovine or glycosylation of the feline orthologues.

1.4.2 CLCA2: the CLCA of keratinocytes

Only one year after *hCLCA1* was cloned, *hCLCA2*, the first member of the CLCA2 cluster was isolated from a expressed sequence tag (EST) database (Agnel et al. 1999). *CLCA2* orthologues from mouse (Evans et al. 2004) and pig (Plog et al. 2012b) were subsequently cloned, however no such cloning was reported for rat, cat, horse, and cow. CLCA2 protein was mainly found in cutaneous and/or mucocutaneous keratinocytes of various organs in humans (Connon et al. 2004;Seltmann et al. 2018), mice (Braun et al. 2010), rats (Hamalainen et al. 2021), cats (Erickson et al. 2020) and pigs (Plog et al. 2012b). In addition, a slight species-specific variation of the expressing cell types was described. In humans, CLCA2 was detected in corneal epithelial cells (Connon et al. 2006) and lung vascular endothelium (Abdel-Ghany et al. 2001). Murine CLCA2 protein was also found in Hasall´s body of thymus (Braun et al. 2010) and mammary epithelium (Hughes et al. 2016). In pigs, expression of the CLCA2 protein was additionally described in cells of the inner hair follicle root sheath (Plog et al. 2012b). In the airways, CLCA2 protein was found in submucosal glands in humans, mice and pigs (Dietert et al. 2015). Furthermore, mice express the CLCA2 protein in non-ciliated epithelial cells and in goblet cells of a certain niche in the bronchial bifurcation (Dietert et al. 2015). The expressing cell type for equine and bovine species was not experimentally investigated yet. Protein biochemical properties of CLCA2 included cleavage of human (Gruber et al. 1999), murine (Evans et al. 2004), porcine (Plog et al. 2012b), and glycosylation of human (Gruber et al.

1999) and murine (Braun et al. 2010) orthologues. Glycosylation was not experimentally tested in pigs and cleavage plus glycosylation was not tested in rat, feline, equine and bovine CLCA2. The N-terminal, but not the C-terminal cleavage product is secreted, as it was shown for human (Elble et al. 2006) and murine orthologue (Braun et al. 2010). This suggests the presence of a TMC.

1.4.3 CLCA3: the ubiquitous CLCA

All members of the third cluster of *CLCA* genes have been cloned from cow (Cunningham et al. 1995; Elble et al. 1997), mouse (Gandhi et al. 1998; Lee et al. 1999; Elble et al. 2002) and rat (Jeong et al. 2005; Yoon et al. 2006). In humans and pigs, *CLCA3* appears to be a pseudogene (Mundhenk et al. 2018). The horse and the cat are the only two of five species with a putatively functional CLCA3 (Patel et al. 2009; Mundhenk et al. 2018). However, no cloned member was described yet in both of these species. Since the nomenclature of the cloned *CLCA3* representatives of cow and rat in the original publications is confusing and the nomenclature given by NCBI Genebank is not indicative of the associated cluster, phylogenetic trees indicating the designation of these sequences are listed in figure 1 and in the appendix figures 3 and 4. The cellular CLCA3 protein expression spectrum described in the literature was the widest among all mammalian CLCA members. In the mouse, CLCA3a protein was expressed in pancreatic acinar cells, acinar and duct cells of the parotid gland, gastric parietal cells, goblet cells and crypt-enterocytes of small intestine, proximal and distal tubular epithelial cells (Roussa et al. 2010), endothelial cells (Abdel-Ghany et al. 2002) and keratinocytes in cell culture (Seltmann et al. 2018). Furthermore, the results of a study by Furuya and coworkers suggested that Clca3a might be expressed by stromal cells of the splenic red pulp, lymphatic endothelium and thymic stromal epithelium (Furuya et al. 2010). No study described the cell type specific protein expression of murine CLCA3b yet. Rat CLCA3 protein was detected in olfactory sensory neurons (Gonzalez-Silva et al. 2013; Mura et al. 2017), ductal epithelial cells of the submandibular gland (Yamazaki et al. 2005; Ishibashi et al. 2006; Yamazaki et al. 2013) and keratinocytes (Yamazaki et al. 2013). Feline CLCA3 protein was found in cells of submucosal glands and/or ciliated epithelial cells of the airways (Mundhenk et al. 2018; Erickson et al. 2020) and esophageal keratinocytes (Mundhenk et al. 2018). The bovine CLCA3 protein was detected in vascular endothelial cells of the lung and in respiratory epithelial cells of bronchi and trachea (Elble et al. 1997). Cell types expressing CLCA3 protein in the horse are still unknown. Murine CLCA3a and rat CLCA3 are cleaved (Roussa et al. 2010) and glycosylated (Gandhi et al. 1998; Yamazaki et al. 2005; Roussa et al. 2010; Lee et al. 2016) proteins. Results from *in vitro* experiments by Elble et al. and Cunningham and coworkers targeting the bovine CLCA3 also implied cleavage and glycosylation of the bovine CLCA3 proteins (Cunningham et al. 1995; Elble et al. 1997). Whether the putatively functional

CLCA3 members are fully secreted proteins in each species is contradictory and remains to be determined experimentally. *In silico* analysis predicted no TMC for the feline orthologue (Mundhenk et al. 2018). A secretion was experimentally demonstrated for a murine CLCA3 member (Huan et al. 2008). In contrast, Yoon et al. predicted a transmembrane domain for rat CLCA3 (Yoon et al. 2006). Furthermore, an *in vitro* study of a rat CLCA3 found no evidence of heterologous expressed protein in the cell culture medium (Lee et al. 2016).

1.4.4 CLCA4: the CLCA of enterocytes

Concurrently with *hCLCA2*, the sequence of human *CLCA4*, the first cloned member of the CLCA4 cluster, was published in 1999 (Agnel et al. 1999). A few years later, and also at the same time with *mCLCA2*, the murine *CLCA4a* (Evans et al. 2004) was isolated from the intestine. The rat *CLCA4* was later cloned from the uterus (Song et al. 2009). Finally, porcine *CLCA4a* (Plog et al. 2012a) and *CLCA4b* (Plog et al. 2015) were cloned in 2012 and 2015, respectively. Murine *CLCA4b*, and the *CLCA4* orthologues from cat, cow and horse have not been isolated yet. CLCA4 protein members were predominantly found in enterocytes. For example, human CLCA4 protein was detected in epithelial cells of colonic glands (Chen et al. 2019; Mo et al. 2020). The murine orthologue was expressed in enterocytes along the intestine (Bothe et al. 2008) and the porcine CLCA4a in enterocytes on the small intestine villus tips (Plog et al. 2012a), while CLCA4b protein was detected in small and large intestinal crypt epithelial cells (Plog et al. 2015). Interestingly, the porcine CLCA4a protein seemed to be expressed also in tracheal and bronchial epithelial cells (Plog et al. 2012a). The expressing cell type of the murine CLCA4b and the rat, feline, equine, and bovine orthologues is still unknown. Cleavage and glycosylation was shown for murine CLCA4a (Evans et al. 2004; Bothe et al. 2008) and porcine CLCA4a (Plog et al. 2012a), while only cleavage was tested and demonstrated for porcine CLCA4b (Plog et al. 2015) proteins. The release of the N-terminal cleavage product into the medium was verified for murine CLCA4a (Bothe et al. 2008) and porcine CLCA4a (Plog et al. 2012a), while the C-terminal fragment of CLCA4 orthologues was found in association with the cell membrane in murine CLCA4a (Evans et al. 2004; Bothe et al. 2008) and porcine CLCA4a (Plog et al. 2012a). These findings suggested the presence of a TMC. Biochemical properties of other CLCA4 members still remain unknown and have to be elucidated in the future.

The numerous, descriptive data obtained from mammalian CLCAs provide an excellent basis for comparative molecular and expressional analyses. By comparison of these data to those of evolutionary more distant species, such as the chicken, conserved and newly acquired traits can be identified and help to infer the most likely evolution of this gene family.

1.5 Scientific questions and hypotheses

1.5.1 Is the exon-intron structure of galline *CLCA1* and *2* similar to those of mammalian clusters *CLCA1-4*?

The known mammalian *CLCA* genes are encoded by 14 exons. It was hypothesized that this is a conserved trait among galline and mammalian *CLCA* genes.

1.5.2 Are the protein architecture and the biochemical properties of galline CLCA1 a combination of those of mammalian CLCA1, 3 and 4?

Due to the close genetic distance of gCLCA1 and mammalian CLCA1, 3 and 4 demonstrated by Mundhenk and coworkers (Mundhenk et al. 2018), it was hypothesized that a similar protein domain architecture and biochemical properties of mammalian CLCA1, 3 and 4 apply for gCLCA1.

1.5.3 Is the cellular expression pattern of galline CLCA1 a combination of those of mammalian CLCA1, 3 and 4?

Since mammalian CLCA clusters 1, 3 and 4 are expressed by different cell types, it was hypothesized that gCLCA1 combines this property without being expressed just by a single cell type.

1.5.4 Are the protein domain architecture and the biochemical properties conserved among galline CLCA2 and the mammalian orthologues?

Due to the close genetic distance to mammalian CLCA2 and due to the concept of orthologue conjecture, it was hypothesized that gCLCA2 has similar protein domains and biochemical properties as mammalian CLCA2.

1.5.5 Are the cell type specific expression patterns of gCLCA2 and mammalian CLCA2 conserved?

Due to the close genetic distance to mammalian CLCA2, it was hypothesized that gCLCA2 has a similar cell type specific expression pattern.

1.5.6 Which characteristics might hypothetical ancestral CLCA proteins have had?

Own research publications in scientific journals with peer-review

1.6 Evolutionarily conserved properties of CLCA proteins 1, 3 and 4, as revealed by phylogenetic and biochemical studies in avian homologues

Authors: Florian Bartenschlager, Nikolai Klymiuk, Christoph Weise, Benno Kuropka, Achim D Gruber, Lars Mundhenk

Year: 2022

Journal: PLoS ONE

DOI: doi.org/10.1371/journal.pone.0266937

Bibliographic Source: Bartenschlager, F., Klymiuk, N., Weise, C., Kuropka, B., Gruber, A. D., Mundhenk, L. (2022): Evolutionarily conserved properties of CLCA proteins 1, 3 and 4, as revealed by phylogenetic and biochemical studies in avian homologues. PloS One 17: e0266937.

No permission by *PLoS ONE* is required for inclusion in this thesis as all content published by *PLoS ONE* is distributed under the terms of the Creative Commons Attribution License. This permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Declaration of own portion of work in this research publication:

Contributions of F. Bartenschlager:

1. Development and conceptualization of the study design including planning, preparation, and acquisition and processing of tissues. Project administration including experimental planning, preparation, performance, analysis, evaluation, validation and data curation of *in silico* gene, genome and protein analyses as well as cloning, RT-qPCR, cell culture, western blot, immunocytochemistry, immunohistochemistry and immunofluorescence experiments. Experimental planning and performance of sample preparation for mass spectrometry analyses.
2. Visualization of experimental results and subsequent writing of original draft including supplementary files except the sections covering phylogenetic analyses on domain level and parts of the mass spectrometry.

Contributions of the other authors:

All co-authors participated in the development of study design, evaluation of experimental results and the setup and review of the manuscript. Chris Weise and Benno Kuropka assisted

in the sample preparation, performed analysis of the mass spectrometry experiments and participated in writing the section on mass spectrometry. Nikolai Klymiuk prepared and performed phylogenetic analyses and wrote the manuscript section covering this part.

Declaration on ethics:

All animal tissues used in this study were by-products from slaughtered animals. The animals had been bred, housed, and slaughtered in the Albrecht Daniel Thaer-Institute of Agricultural and Horticultural Sciences of the Humboldt-Universität zu Berlin, Germany (State Office of Health and Social Affairs Berlin, approval number IC 114-ZH70).

The publisher´s version can be openly accessed under the following link:

https://doi.org/10.1371/journal.pone.0266937

PLOS ONE

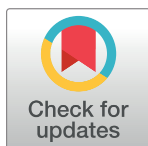
RESEARCH ARTICLE

Evolutionarily conserved properties of CLCA proteins 1, 3 and 4, as revealed by phylogenetic and biochemical studies in avian homologues

Florian Bartenschlager[1], Nikolai Klymiuk[2,3], Christoph Weise[4], Benno Kuropka[4], Achim D. Gruber[1], Lars Mundhenk[1]*

1 Faculty of Veterinary Medicine, Department of Veterinary Pathology, Freie Universität Berlin, Berlin, Germany, **2** Large Animal Models in Cardiovascular Research, Internal Medical Department I, Technical University of Munich, Munich, Germany, **3** Center for Innovative Medical Models, Ludwig-Maximilians University Munich, Munich, Germany, **4** Institute of Chemistry and Biochemistry, Core Facility BioSupraMol, Freie Universität Berlin, Berlin, Germany

* Lars.mundhenk@fu-berlin.de



OPEN ACCESS

Citation: Bartenschlager F, Klymiuk N, Weise C, Kuropka B, Gruber AD, Mundhenk L (2022) Evolutionarily conserved properties of CLCA proteins 1, 3 and 4, as revealed by phylogenetic and biochemical studies in avian homologues. PLoS ONE 17(4): e0266937. https://doi.org/10.1371/journal.pone.0266937

Editor: Cyril Charles, Ecole Normale Supérieure de Lyon, FRANCE

Received: May 20, 2021

Accepted: March 30, 2022

Published: April 13, 2022

Copyright: © 2022 Bartenschlager et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting information files.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Species-specific diversities are particular features of mammalian chloride channel regulator, calcium activated (*CLCA*) genes. In contrast to four complex gene clusters in mammals, only two *CLCA* genes appear to exist in chickens. *CLCA2* is conserved in both, while only the galline *CLCA1* (*gCLCA1*) displays close genetic distance to mammalian clusters 1, 3 and 4. In this study, sequence analyses and biochemical characterizations revealed that gCLCA1 as a putative avian prototype shares common protein domains and processing features with all mammalian CLCA homologues. It has a transmembrane (TM) domain in the carboxy terminal region and its mRNA and protein were detected in the alimentary canal, where the protein was localized in the apical membrane of enterocytes, similar to CLCA4. Both mammals and birds seem to have at least one TM domain containing CLCA protein with complex glycosylation in the apical membrane of enterocytes. However, some characteristic features of mammalian CLCA1 and 3 including entire protein secretion and expression in cell types other than enterocytes seem to be dispensable for chicken. Phylogenetic analyses including twelve bird species revealed that avian CLCA1 and mammalian CLCA3 form clades separate from a major branch containing mammalian CLCA1 and 4. Overall, our data suggest that gCLCA1 and mammalian CLCA clusters 1, 3 and 4 stem from a common ancestor which underwent complex gene diversification in mammals but not in birds.

Introduction

The ***ch**l**oride channel regulator, **c**alcium **a**ctivated* (*CLCA*) gene family has been implicated in the pathophysiology of inflammatory airway diseases, such as asthma [1–3] or cystic fibrosis [4–6], tumor progression and metastasis [7–11]. In this context, different CLCA proteins have

been linked to cellular functions including mucus processing [12,13], activation of airway macrophages [14–16], the modulation of the anion channel TMEM16A [17,18], cellular differentiation such as mucus cell metaplasia [1,19–21], cellular senescence and apoptosis [7,22–27], or as adhesion molecules for metastatic tumor cells [28,29]. Due to their proposed biomedical relevance, CLCA molecules have been studied in healthy and diseased humans [2,4,5,26,30–37] and mammalian model organisms including mouse [38–44], pig [45–49], horse [3,50], and cat [51,52].

The mammalian genome contains multiple *CLCA* gene copies, which can be classified into four distinct clusters based on genetic distances [51]. All *CLCA* genes are located in a single, conserved gene locus flanked by the two genes "*outer dense fibre of sperm tails 2-like*"(*ODF2L*) and "*SH3-domain GRB2-like endophilin B1*" (*SH3GLB1*). Clusters 1 and 2 are conserved in mammals, and each species carries only one functional *CLCA1* and *CLCA2* homologue. In contrast, clusters 3 and 4 are diverse, characterized by inactivation and duplication events in several mammalian species. In pigs, sheep, and dry-nose primates including humans, the *CLCA3* gene underwent independent mutational inactivation including several frameshift and nonsense mutations as well as the introduction of an additional exon [51]. In contrast to this pseudogenization, the gene is duplicated in cattle and mice leading to two or three functional orthologues, respectively. In other species such as cats, cluster 3 consists of only a single functional orthologue. In cluster 4, the *CLCA4* gene is duplicated in pigs and mice with two or three representatives, respectively. In humans, cats, and cattle only one functional *CLCA4* gene exists in this cluster. Taken together, the *CLCA* gene family is characterized by conserved clusters, such as 1 and 2, and by diverse clusters, such as 3 and 4 in mammals [51].

This genetic heterogeneity is also reflected in the cellular expression patterns of the proteins. CLCA1 was detected in mucin-producing cells of different mammal species [3,30,49,50,52,53] and the CLCA2 proteins are commonly expressed in keratinocytes of stratified epithelia [27,47,52,54–56]. In contrast, the expression patterns of the cluster 3 proteins are diverse in mammalian species. A single intact CLCA3 has so far only been found and characterized to date in cats. It is expressed in mucus-producing submucosal cells, ciliated epithelial cells, and esophageal keratinocytes [51]. In species with duplicated and intact *CLCA3* genes such as mouse and cattle (*Bos taurus*), their gene products were localized to keratinocytes, secretory epithelial cells, smooth muscle cells, and vascular endothelial cells [38,40,42,57–61]. In mammals, the CLCA4 proteins are predominantly expressed in enterocytes [46,48,62]. The duplication products of CLCA4 may lead to different cellular niches. In pigs, for example, one was localized to enterocytes of the villus tips of the small intestine while the other duplicate was found in crypt epithelial cells of small and large intestine [46,48]. In humans and pigs and in contrast to mice, CLCA4 members are also expressed in the airways [33,46,63]. In cluster 3 and 4, the duplication events seem to broaden the tissue and cellular expression pattern of the CLCA proteins with a species-specific evolution. This duplication may be associated with a sub- or neofunctionalization, which is unknown to date.

Despite this diversity at the genomic level and in the expression pattern of the particular CLCA clusters in mammals, a common feature of their functional protein products is the entry into the secretory pathway and an autocatalytic posttranslational cleavage of the precursor protein into a larger amino (N)-terminal and a smaller carboxy (C)-terminal subunit [64]. The N-terminal subunit encompasses a cysteine-rich CLCA domain with a HExxH motif (PFAM identifier: pfam08434), a von Willebrand factor type A (vWA) domain (PFAM identifier: pfam13519) and a β-sheet rich domain [64]. The C-terminal cleavage fragment has a fibronectin type III domain (PFAM identifier: PF00041) [64]. Additionally, based on biochemical properties, the mammalian CLCA clusters can be divided into two groups: the proteins of cluster 1/3 are soluble proteins, which are secreted in their entirety by the cells [51,52,64–66],

whereas the proteins of cluster 2/4 are anchored in the plasma membrane via a C-terminal transmembrane domain [47,62,64,67]. Only one study on the phylogenetic distribution of the zinc-binding amino-acid (aa) motif HExxH of the CLCA domain has been published so far [68]. Furthermore, no CLCA homologues of other vertebrates have been investigated in detail to date, which limits assertions about novel acquired and conserved traits of CLCA proteins in the mammal lineage. Recently, it was reported that the *CLCA* genes in chicken are flanked by *ODF2L* and *SH3GLB1* identically to mammals; however, only two homologues were found in this gene locus [51]. One galline homologue clustered with the conserved mammalian *CLCA2* genes. In contrast, the second *CLCA* homologue, *gCLCA1*, clustered not only with the mammalian cluster 1, but also with the diverse clusters 3 and 4.

Here, we test the hypothesis that gCLCA1 combines expressional and protein-biochemical characteristics of mammalian CLCA1, 3, and 4, which suggests a common ancestor with an independent expansion in the early mammalian lineage. We analyzed the genomic organization and protein structure of gCLCA1 *in silico*, characterized its protein-biochemical properties and expression patterns in comparison to mammalian CLCA1, 3, and 4. The data may help to elucidate the properties and putative functions of early ancestral members of this gene family and infer novel properties of the diverse CLCA clusters in mammals.

Materials and methods

Characterization of the *gCLCA* gene locus and phylogenetic analyses

In addition to the architecture of the g*CLCA* locus [51] detailed gene positions and sizes and exon-intron boundaries were extracted from the NCBI (https://www.ncbi.nlm.nih.gov/) database and Plog et al. [49]. In brief, CLCA protein sequences were obtained from the NCBI and ensembl (http://www.ensembl.org/index.html) databases and aligned. Phylogenetic trees were generated using the phylip package (Phylogeny Inference Package 3.6. J. Felsenstein. Department of Genome Sciences, University of Washington, Seattle, WA, USA). First, three independent trees were generated by the maximum likelihood (ml), most parsimony (mp), and 100 neighbor joining (nj) methods. The final tree was based on the ml tree, with branch nodes occurring also in mp. Mammalian and avian CLCA2 were used as an outgroup. For better identification of distinct evolutionary speed in different protein domains, separate trees were calculated for the N-CLCA domain which includes a catalytic domain (cat, corresponding to aa45-200 in chicken CLCA1) and a cysteine-rich domain (cys, aa201-289), the von Willebrand factor A domain (vWA, aa297-483), a β-sheet rich domain (bsr, aa484-698), and fibronectin type III domain (fn3, aa754-874).

In silico amino-acid sequence analysis of gCLCA1 and generation of antibodies

The aa sequence of the gCLCA1 was analyzed via NCBI Conserved Domain Database [69], EMBL-EBI HMMER web server [70], SUPERFAMILY 2 database [71,72], Phobius webserver [73], SOSUI [74], CCTOP [75], and SignalP 5.0 [76] algorithms to identify putative protein domains. Asparagine (N)-linked glycosylation sites were predicted using NetNGlyc 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/). The putative proteolytic cleavage site was identified by interspecies comparison with aa sequences of mammalian CLCA proteins. Anti-gCLCA1 antibodies were generated as described [49]. In brief, according to immunogenicity predictions, two oligopeptides were synthesized corresponding to aa92 to 105 (KKNSTYSRLKTESY, gCN1) and aa821 to 834 (ASVPSDDEGNTSDG, gCC1). *Limulus polyphemus* hemocyanin-conjugated peptides were used for immunization of two rabbits each. Specific IgG-antibodies

were purified from the antisera using a cyanogen bromide immunization peptide-coupled sepharose column. The immunopurified polyclonal antibodies generated against the N-terminal and C-terminal part of the gCLCA1 protein were named gC1-N1 and gC1-C1, respectively.

Animals and tissue processing

44 tissues each (S1 File) from ten-week old female chickens (Hampshire x White Leghorn, n = 3) and the gonads of male chickens (Hampshire x White Leghorn, n = 3) of the same age were snap-frozen on dry ice subsequent to a brief immersion in 2-methylbutane. For immunohistology or immunofluorescence, tissues were fixed in 4% neutral-buffered formaldehyde for 48 h and embedded in paraffin (FFPE-tissues). All tissues were by-products from slaughtered animals intended for human consumption. The animals had been bred, housed, and slaughtered in the Albrecht Daniel Thaer-Institute of Agricultural and Horticultural Sciences of the Humboldt-Universität zu Berlin, Germany (State Office of Health and Social Affairs Berlin, approval number IC 114-ZH70). Weight at harvest was 1–1.2 kg (females) and 1.3–1.5 kg (males). The animals were raised on miscanthus litter in groups of 25, with infrared heat lamps offered until week five. The animals were fed with fledgling rearing feed until week eight and pullet feed afterwards. Females were harvested in the morning and males in the morning of the following day.

Molecular cloning of *gCLCA1* and generation of *gCLCA1* mutants

gCLCA1 was cloned and mutants were generated as described in detail in S2 File. In brief, the *gCLCA1* open reading frame (ORF) was amplified from reverse transcribed RNA extracted from the proctodeum and tagged with the enhanced yellow fluorescent protein (*EYFP*) at the C-terminus by cloning into the p*EYFP*-N1 vector (*gCLCA1WT*). For analysis of an EQ mutation in the zinc-binding aa motif in accordance with Pawlowski et al. [77] and Bothe et al. [78], the wild type motif was replaced by a synthesized gene fragment containing a substitution of glutamic acid (E) with glutamine (Q) at position 164 (*gCLCA1E164Q*). As none of the specific anti-gCLCA1 antibodies were applicable for immunoblotting (see Immunoblotting section), a *gCLCA1*-fragment was substituted by a synthesized gene fragment containing a recognition site for murine N-terminal CLCA1 antibodies (α-p3b2, [53]) in order to enable for the detection of the N-terminal cleavage product of the murinized wild type (*gCLCA1Nmabc1)* or murinized EQ mutant (*gCLCA1Nmabc1E164Q) in vitro*.

Tissue expression pattern of *gCLCA1* mRNA

The *gCLCA1* mRNA organ-specific expression pattern was analyzed by RT-qPCR as described in detail in S2 File. In brief, total RNA was isolated from tissues (S1 File), reverse transcribed, and the synthesized cDNA was diluted to a final concentration of 1 ng/μl. Specific, exon-boundaries spanning primers were used to detect *gCLCA1* or the reference gene phosphoglycerate kinase 1(*PGK1*) [79] using a SYBR green qPCR assay. *gCLCA1*-mRNA was considered to be expressed, when C_t-values of 35 or below were detected in at least two of three samples tested.

Transient transfection of HEK293 cells

HEK293 cells were transfected as described in detail in S2 File. In brief, the cells were grown in six-well plates in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% heat-inactivated fetal calf serum (FCS), 1% HEPES, and 1% penicillin/streptomycin. When the

cells were approximately 80–90% confluent, they were transfected using 2 μg of a plasmid (*EYFP*-mock, pcDNA3.1^{+}-mock, *gCLCA1WT*, *gCLCA1E164Q*, *gCLCA1Nmabc1*, *gCLCA1N-mabc1E164Q*, *mCLCA1* [78], *mCLCA1E157Q* [78], *mCLCA4a* [80], or *mCLCA4aE157Q* [80]) and 8 μl polyethylenimine (PEI, 1 mg/ml) per well. 12 h after transfection, the cells were washed with phosphate buffered saline (PBS) and serum-free DMEM was added to the cells. 72 h after transfection, the cell culture medium was removed and centrifuged at 14000 x g for one h at 4˚C. The supernatant was concentrated using Vivaspin 2 protein concentrator spin columns (Sartorius, Göttingen, Germany) with a 10 kDa molecular weight cutoff. The cells of each well were lysed using 500 μl radioimmunoprecipitation assay (RIPA) buffer supplemented with a protease inhibitor cocktail. The protein concentration of supernatants and cell lysates were quantified using the bicinchoninic acid (BCA) method prior to freezing at –20˚C.

Endoglycosidase treatment

For glycosylation analysis, lysates from *gCLCA1Nmabc1*-transfected cells were deglycosylated by incubation with 25 U/ml endo H, 50 U/ml PNGase F, or left untreated at 37˚C overnight according to the manufacturer´s protocol (New England Biolabs, Ipswich, Massachusetts, USA) and consequently immunoblotted as described in the Immunoblotting section.

Immunoblotting

Cell lysates and supernatants of transfected cells were analyzed by immunoblotting as described in detail in S2 File. In brief, samples of cell lysates or concentrated cell culture supernatant were reduced in 1,4-dithiothreitol (DTT) and separated in a 10% acrylamide gel. The proteins were transferred onto a polyvinylidene fluoride (PVDF)-membrane and blocked with 5% non-fat milk. The membranes were probed with gC1-N1 and gC1-C1 both at a three-fold dilution series from 5 μg/ml to 0.05 μg/ml, or mouse monoclonal anti-YFP (cat. G163, abm, Vancouver, Canada) at 1:500, or rabbit polyclonal anti-mCLCA1 [53] at 1:500, or rabbit polyclonal anti-mCLCA4a [62], or mouse monoclonal anti beta-actin (A5441, MilliporeSigma, Saint Louis, Missouri, USA) antibodies at 1:1000. The membranes were incubated with horseradish peroxidase-conjugated goat anti-rabbit (115-035-068, Jackson Immuno Research Laboratories, Inc., West Grove, Pennsylvania, USA) or goat anti-mouse (111-035-144, Jackson Immuno Research Laboratories, Inc.) secondary antibodies and developed using enhanced chemiluminescence. The immunoblotted gCLCA1 protein was not detectable using antibodies gC1-N1 or gC1-C1 at any dilution used (S3 File).

Immunocytochemistry of transfected HEK293 cells

Procedures are described in detail in S2 File. In brief, cells were grown on 8-well tissue chamber slides and transfected with *gCLCA1WT* or *EYFP*-mock plasmids. 72 h after transfection, the cells were briefly fixed in ice-cold methanol (100%) for five min followed by a fixation in 4% paraformaldehyde for ten min. The cells in each well were permeabilized with 0.1% Triton X-100 in PBS and blocked with 10% goat normal serum (GS) and 0.05% Tween 20 in PBS. After blocking, the cells were probed with untreated or pre-absorbed gC1-N1, or gC1-C1, or irrelevant affinity-purified rabbit polyclonal (anti-pig CFTR [81]) antibodies (each used at 1 μg/ml). Alexa fluor 568-conjugated goat anti-rabbit (AB_143157, Invitrogen, Carlsbad, California, USA) were used as secondary antibodies followed by 4′,6-diamidino-2-phenylindole (DAPI) nuclear counterstain for each experiment. No specific signals were detected using the gC1-N1 antibody (S3 File).

Tissue expression pattern of gCLCA1 protein using immunofluorescence and immunohistochemistry

Methods are described in detail in S2 File. In brief, all tissues from the three animals that showed a *gCLCA1*-specific Ct-value below 35 were analyzed using immunofluorescence. FFPE-tissues were cut at 3 μm thickness, mounted on adhesive glass slides, and dewaxed. For immunohistochemistry, endogenous peroxidase was blocked by 0.5% H_2O_2. Antigens were retrieved using 1 mg/ml recombinant protease from *Streptomyces griseus* or microwave heating (600 W) for 15 min in 10 mM citric acid, pH 6.0, containing 0.05% Triton X-100. Slides were blocked with 10% Roti-ImmunoBlock and 20% GS in PBS and probed with gC1-C1 at 1 μg/ml or rabbit monoclonal anti-villin (ab130751, Abcam, Cambridge, United Kingdom) at 1:400 or irrelevant affinity-purified rabbit polyclonal (anti-pig CFTR [81]) at 1 μg/ml antibodies. 3,3´-Diaminobenzidine (DAB) stain was added after incubation with goat anti-rabbit biotinylated secondary antibodies (BA-1000, Vector Laboratories, Burlingame, California, USA) at 1:200 and an avidin-biotin complex solution. Slides were counterstained with 1% alcian blue in HCl at pH 1.0 and 0.1% nuclear fast red-aluminum sulfate.

For immunofluorescence analysis, tissue sections were permeabilized with 0.1% Triton X-100 in PBS and blocked with 10% GS and 0.05% Tween 20 in PBS. The slides were incubated with the gC1-C1 antibody at 1 μg/ml or an irrelevant affinity-purified rabbit polyclonal (anti-pig CFTR) antibody [81] using the same concentration. Alexa fluor 568-conjugated goat anti-rabbit (AB_143157, Invitrogen) secondary antibodies at 1:200 and a DAPI nuclear counterstain were used for immunofluorescence.

Unless otherwise noted, all *in vitro* studies were performed in a minimum of three independent cell batches where all transfections were conducted in parallel.

Results

The galline *gCLCA1* gene shares characteristics with its mammalian orthologues and encodes a putatively functional protein

The galline *CLCA* locus is located on chromosome 8 flanked by the genes *ODF2L* and *SH3GLB1* and it is shorter than mammalian *CLCA* loci ([49], Fig 1A). Compared to the human, porcine, and murine genes of cluster 1 ([49], Fig 1A and 1B), the *gCLCA1* gene is shorter, however, it is composed of 14 exons just as all intact mammalian *CLCA* genes (Fig 1B). The *gCLCA1* ORF consists of 2808 base pairs (bp, 936 aa), which is longer than those of mammalian homologues, such as human *CLCA1* (2742 bp, 914 aa) or murine *CLCA4b* (2775 bp, 925 aa) [64,82]. The predicted galline *CLCA1* gene does not contain premature stop codons or frameshift mutations as it had been identified for mammalian *CLCA3* [51] and therefore may encode a functional protein. Compared to the reference sequence of the NCBI Genbank (XM_422360.6, *Gallus gallus*), our *gCLCA1* ORF contained four synonymous and four non-synonymous single nucleotide polymorphisms (SNPs, S4 File).

The gCLCA1 protein contains canonical CLCA domains

All known mammalian CLCA proteins enter a secretory pathway and contain an N-CLCA (PFAM identifier: pfam08434) and a vWA (PFAM identifier: pfam13519) domain. For gCLCA1, the SignalP 5.0 algorithm also predicted a signal peptide between aa1 and 29. Furthermore, the NCBI Conserved Domain Database (CDD), EMBL-EBI HMMER and SUPERFAMILY 2 databases identified an N-CLCA (aa31 to 293) and a vWA domain (aa315 to 482, Fig 2). A cysteine-rich domain within the N-terminal cleavage product, which has been described for mammalian CLCA proteins [77], was also identified in gCLCA1 (Fig 2).

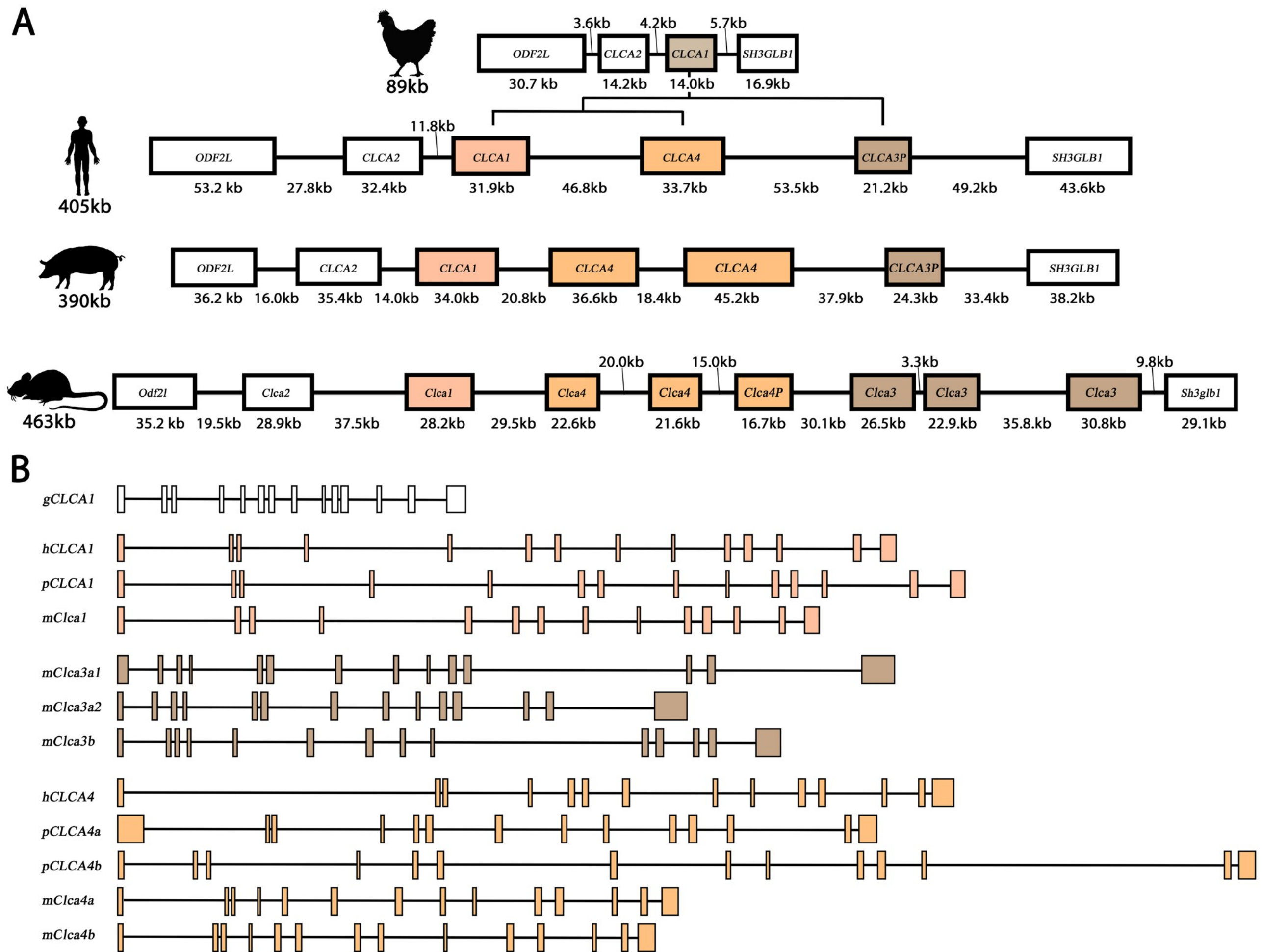


Fig 1. *gCLCA1* gene locus and structure as compared with mammalian homologues. (A) Comparisons of *CLCA* gene loci of the chicken and of the three mammalian representatives human, pig, and mouse. The galline *CLCA* locus consists of two *CLCA* genes, *gCLCA1* and *gCLCA2*, and is shorter than that of mammals. Genes are depicted as boxes and noncoding, intergenic segments are represented as black lines. The chicken locus is scaled 2-fold larger for illustrative purposes. P = pseudogene, kb = kilobases. (B) Comparative gene structure of *gCLCA1* and functional mammalian *CLCA1*, 3 and 4 genes. Like the mammalian *CLCA* genes of clusters 1, 3, and 4, *gCLCA1* is encoded by 14 exons (vertical boxes). Due to short intronic segments (horizontal black lines), the *gCLCA1* gene is shorter than its orthologues of the mammal cluster 1, 3, and 4.

https://doi.org/10.1371/journal.pone.0266937.g001

Similar to the transmembrane proteins of the mammalian cluster 4, the *in silico* analysis consistently predicted a single hydrophobic transmembrane domain between aa897 (SOSUI, Phobius) / 898 (CCTOP) and aa919 (SOSUI) / 922 (CCTOP, Phobius) (Fig 2) for gCLCA1. To verify the presence of the predicted transmembrane domain, the cellular location of the C-terminal cleavage product of the gCLCA1 protein was experimentally analyzed. A green signal of the C-terminal EYFP-tag was detected along the plasma membrane of cells transfected with the *gCLCA1WT* plasmid (Fig 3A). This is in contrast to the diffuse green, cytoplasmic autofluorescent EYFP signal as detected in cells transfected with the EYFP-mock plasmid (S6 File). pcDNA3.1+ transfected HEK293 cells failed to show a specific green autofluorescent signal (S6 File). A red signal along the plasma membrane was found by immunostaining of the cells with the gC1-C1 antibody which binds to the C-terminal subunit of gCLCA1 (Fig 3B, S3 File). These results suggest the presence of a C-terminal transmembrane domain in gCLCA1 similar to mammalian CLCA4 proteins.

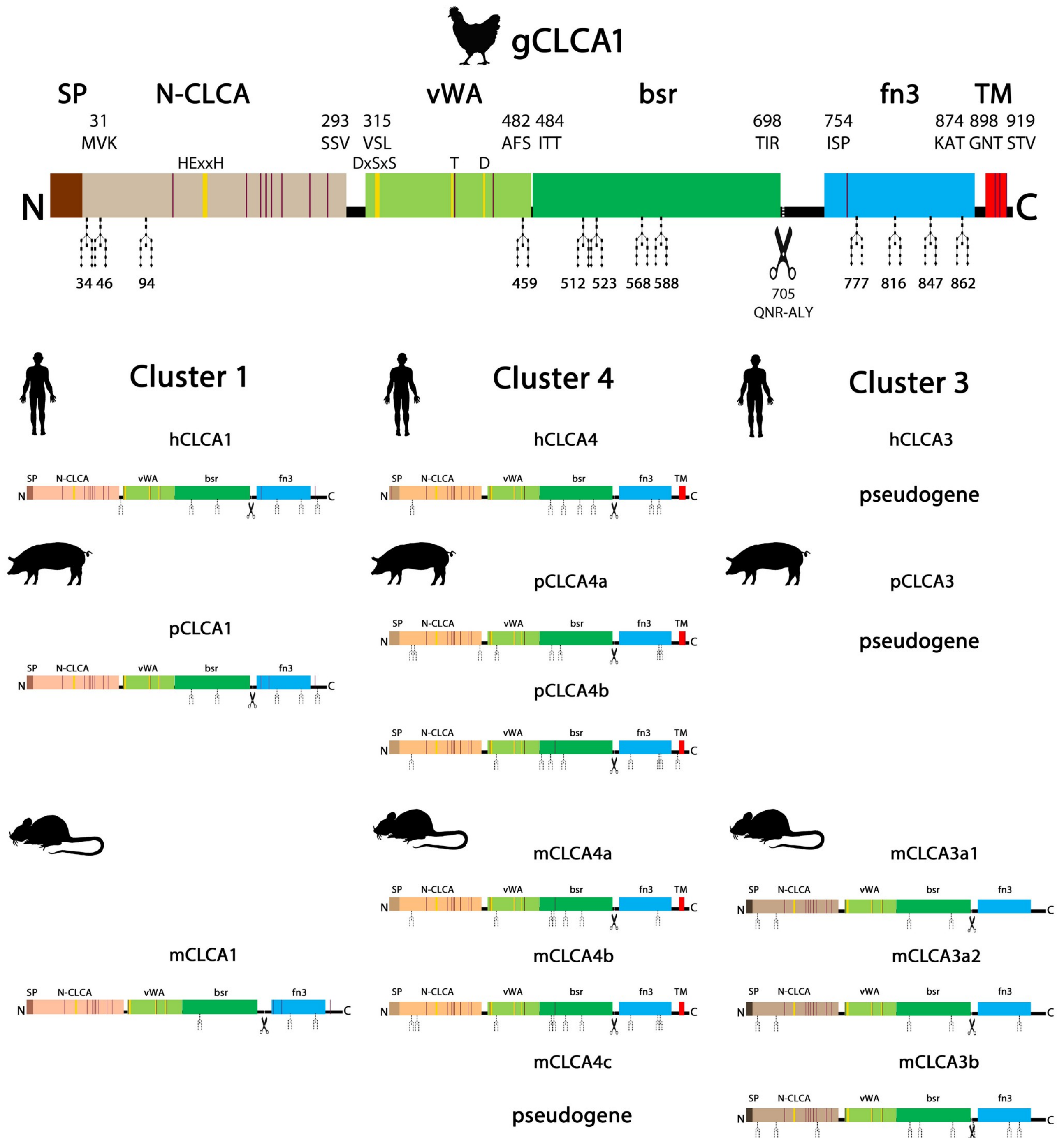


Fig 2. The basic functional domains of mammalian CLCA proteins are conserved in gCLCA1. Schematic depiction of the gCLCA1 protein and of CLCA proteins of clusters 1, 3, and 4 in three mammalian representatives (human, pig and mouse). A cleavable N-terminal signal peptide (dark brown box), an N-CLCA domain (N-CLCA, light brown box), a vWA-domain (vWA, light green box), a β-sheet rich domain (bsr, dark green box) and a fibronectin type III domain (fn3, light blue box) were predicted by *in silico* analyses and manual sequence alignments. The N-CLCA domain comprises an intact zinc-dependent metalloprotease motif (HExxH, vertical yellow bar). The vWA domain contains a MIDAS site (DxSxS, T, D, vertical yellow bar). Vertical dark red bars indicate cysteine residues, predominantly in the N-CLCA domain. The gCLCA1 protein is putatively cleaved after the amino acids QNR at aa position 705 (scissor). gCLCA1 has a putative C-terminal transmembrane domain (TM, light red box) similarly to the CLCA representatives of cluster 4. The gCLCA1 protein is an N-linked glycoprotein with 12 predicted glycosylation sites. The number of glycosylation sites is higher than in the CLCA proteins of cluster 4, followed by the proteins of cluster 3 and 1. Putative glycosylation sites are indicated by -icons. Numbers display the absolute position, for protein domains further specified by the first or last three amino acids of each domain, respectively.

https://doi.org/10.1371/journal.pone.0266937.g002

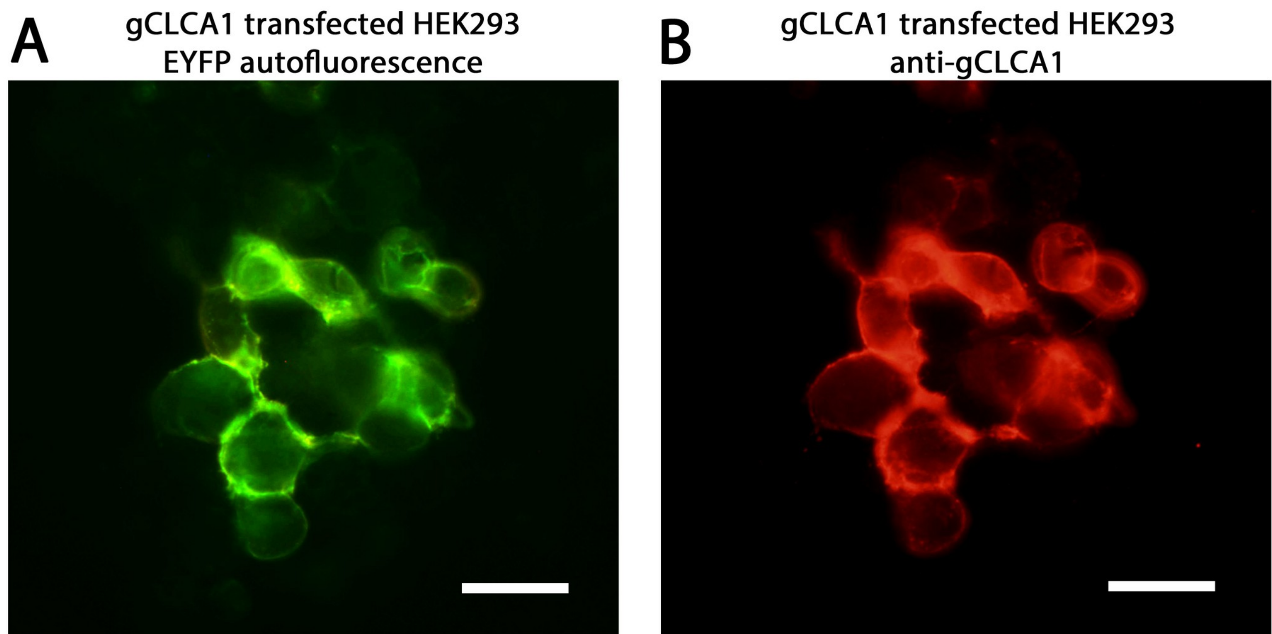


Fig 3. Localization of the C-terminal cleavage product of gCLCA1 in the plasma membrane. (A) The C-terminal cleavage product of the gCLCA1EYFP fusion protein was detected at the plasma membrane of transiently transfected HEK293 cells via autofluorescence of the EYFP-tag (green). (B) A corresponding red signal was reproducible by immunofluorescence using the gC1-C1 (anti-gCLCA1 C-terminal) antibody. Alexa fluor 568-conjugated secondary antibodies. Bars indicate 20 μm. Representative images of three independent experiments.

https://doi.org/10.1371/journal.pone.0266937.g003

The gCLCA1 protein has biochemical properties similar to those of mammalian CLCAs Posttranslational Cleavage

The zinc-binding aa motif HExxH of the CLCA domain, which is necessary for autocatalytic cleavage of a precursor protein into a larger N- and a smaller C-terminal cleavage product [64], is a common feature of all known mammalian CLCA proteins [77,83]. An intact HExxH motif (HEWAH, aa163–167) and a putative proteolytic cleavage site (QNR/ALY, after R705) were also identified in the gCLCA1 protein (Fig 2). It has been reported that mutagenic disruption of the mammalian HExxH motif abrogates the cleavage of mammalian CLCA1 or only impairs that of CLCA4 proteins [78,80]. To evaluate a putative HExxH-dependent cleavage of gCLCA1, lysates from cells transfected with the gCLCA1 wild type and the gCLCA1E164Q mutant were immunoblotted with anti-EYFP antibodies. For cells expressing the wild-type protein, a weak band of the precursor protein at ~154 kilodalton (kDa) and a strong band of the C-terminal cleavage product at ~64 kDa were detected (Fig 4). In contrast, no cleavage product was visible in lysates from cells transfected with the gCLCA1E164Q mutant. A single strong band of the uncleaved, mature glycosylated protein at ~166 kDa and the weak band of the immature glycosylated precursor protein at ~154 kDa (Fig 4) were detected. These findings are consistent with an autocatalytic cleavage of gCLCA1 as previously described for mammalian CLCAs. Additionally, the EQ mutation of gCLCA1 leads to a total abrogation of the cleavage event similar to mammalian CLCA1 (Fig 5).

Secretion of the N-terminal cleavage product

It has been shown that the N-terminal subunit of mammalian CLCA4 is secreted [46,62]. To test for secretion of the N-terminal cleavage product of the gCLCA1, conditioned cell culture media and lysates from cells transfected with *gCLCA1* were analyzed. The ~102 kDa N-

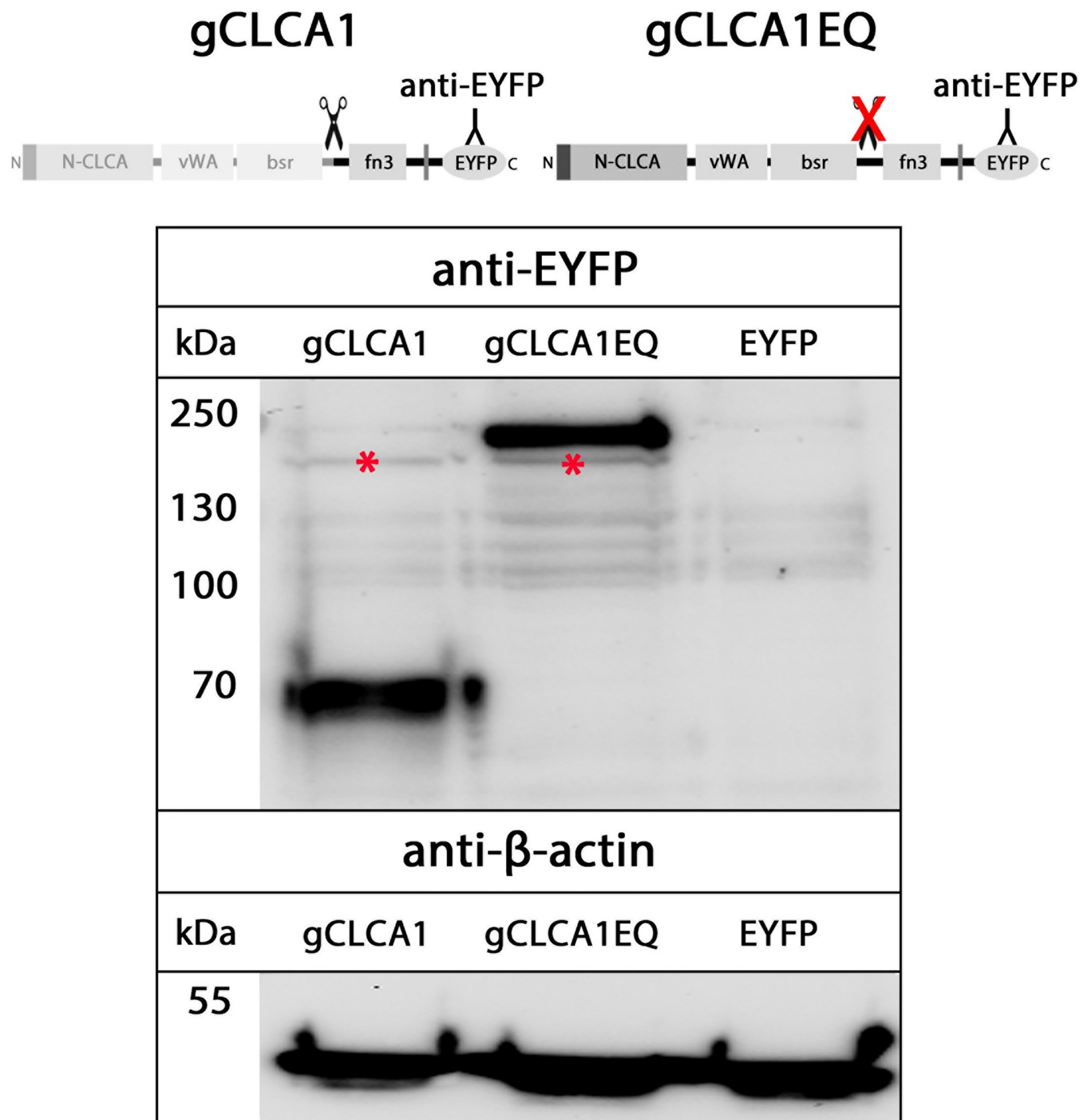


Fig 4. gCLCA1 is a HExxH-dependent cleavable protein. Immunoblot of cell lysates from HEK293 cells transiently transfected with the *EYFP*-mock plasmid (EYFP), the *gCLCA1WT* (gCLCA1) plasmid, and the *gCLCA1E164Q* plasmid containing an EQ substitution at position two of the catalytic active HExxH motif (gCLCA1EQ) is shown. A C-terminal cleavage product of gCLCA1 and its immature glycosylated precursor protein were identified at ~64 kDa or at ~154 kDa (*), respectively. Cleavage was prevented by the EQ substitution in the HExxH motif as no cleavage product was detectable; however, a strong band at ~166 kDa was identified, which putatively represents the uncleaved, mature glycosylated full-length protein. Identical to cells transfected with the *gCLCA1WT* plasmid, the immature glycosylated precursor protein at ~154 kDa was detected in cell lysate from *gCLCA1E164Q* transfected cells. To control for equal total protein loading the samples were identically immunoblotted with primary anti-beta-actin antibodies. Representative images of three independent experiments.

https://doi.org/10.1371/journal.pone.0266937.g004

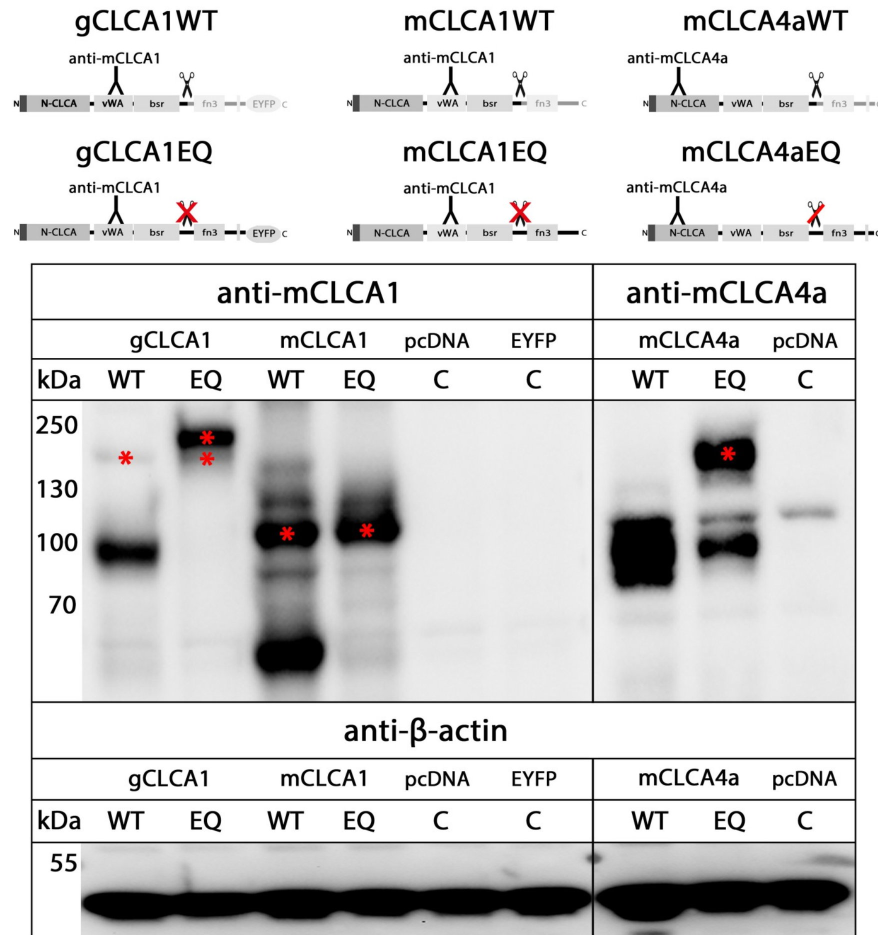


Fig 5. The gCLCA1EQ mutant eliminated autocatalytic cleavage similar to CLCA proteins of mammalian cluster 1. Immunoblotting of HEK293 cell lysates. The left panel illustrates immunoblots from murinized galline CLCA1, the central panel from murine CLCA1 and the right panel from murine CLCA4a constructs. The proteins of the three homologues differ in their molecular weights. The mutation of the HExxH motif of gCLCA1 eliminated the cleavage of the ~154 kDa precursor protein (*) similarly to the murine CLCA1 protein. In contrast, the cleavage of the murine CLCA4aEQ mutant was only impaired, but not totally absent. Asterisks (*) indicate the uncleaved protein of the respective CLCA homolog. To control for equal total protein loading, the samples were identically immunoblotted with primary anti-beta-actin antibodies. WT = HExxH wild type motif, EQ = EQ mutation in the HExxH motif, C = mock-transfected control. Representative images of three independent experiments.

https://doi.org/10.1371/journal.pone.0266937.g005

terminal cleavage product was detected in the supernatant and the cell lysate while the ~64 kDa C-terminal cleavage product was exclusively found in the cell lysate (Fig 6), which further supports the presence of a C-terminal transmembrane domain. Consistently, only tryptic peptides from the N-terminal cleavage product were found in the supernatant from cells transfected with the gCLCA1 using LC-ESI-MS/MS (S5 File). Thus, similar to mammalian CLCA4 proteins, the N-terminal cleavage product of gCLCA1 was secreted.

N-glycosylation and cleavage in the medial Golgi

Mammalian CLCA proteins of clusters 1, 3, and 4 are N-linked glycoproteins, with different number of glycosylation sites [46,49,51,52,62,66]. Consistent with these results, the NetNGlyc 1.0 algorithms predicted 12 potential N-linked glycosylation sites for gCLCA1, located at aa

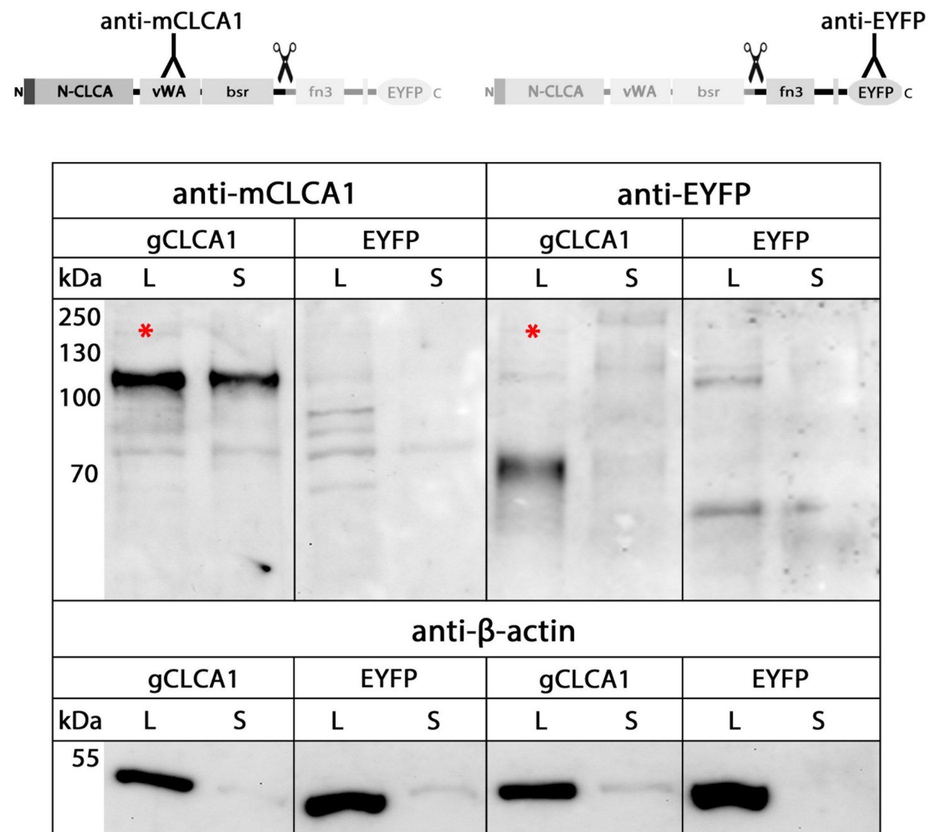


Fig 6. Secretion of only the N-terminal cleavage product of gCLCA1 into the cell culture supernatant. Immunoblotting of lysates (L) and cell culture supernatants (S) from HEK293 cells transfected with the *gCLCA1Nmabc1* (gCLCA1) and *EYFP*-mock plasmid (EYFP). The N-terminal cleavage product was detected at ~102 kDa in cell lysate and supernatant. In contrast, the C-terminal, transmembrane domain containing cleavage product was exclusively detected at ~64 kDa in the cell lysate, but not in the supernatant. A weak band of the immature glycosylated gCLCA1 precursor protein was detectable in the cell lysate at ~154 kDa (*). The same amounts of proteins for the respective lysates and supernatants from transfected cells were analyzed in these experiments. To control for equal total protein loading of cell lysates and for lack of contamination of the supernatants by cell debris, the samples were identically immunoblotted with primary anti-beta-actin antibodies. Representative images of three independent experiments.

https://doi.org/10.1371/journal.pone.0266937.g006

positions 34, 46, 94, 459, 512, 523, 568, 588, 777, 816, 847, and 862 (Fig 2). To corroborate these predictions, cell lysate from *gCLCA1Nmabc1*-transfected cells was treated with endoglycosidases endo H and PNGase F for identification of the kind and extent of the glycosylation. The ~154 kDa precursor protein was sensitive to endo H and PNGase F (Fig 7) resulting in a size shift from ~154 kDa to ~130 kDa which shows an immature high mannose-type glycosylation pattern. In contrast, the N- and C-terminal cleavage products were resistant to endo H but sensitive to PNGase F treatment, shown by a reduction in size from ~102 kDa to ~78 kDa (Fig 7) and ~64 kDa to ~52 kDa (Fig 7), respectively. This indicates that the majority of the predicted eight consensus glycosylation sites in the N-terminal subunit and four glycosylation sites in the C-terminal cleavage product may be used for glycosylation, when a molecular weight of ∼3 kDa per site is assumed [84]. Furthermore, the complex, high mannose-rich glycosylation pattern of the N- and C-terminal cleavage products, in contrast to the precursor protein, suggests a cleavage of gCLCA1 in the medial Golgi.

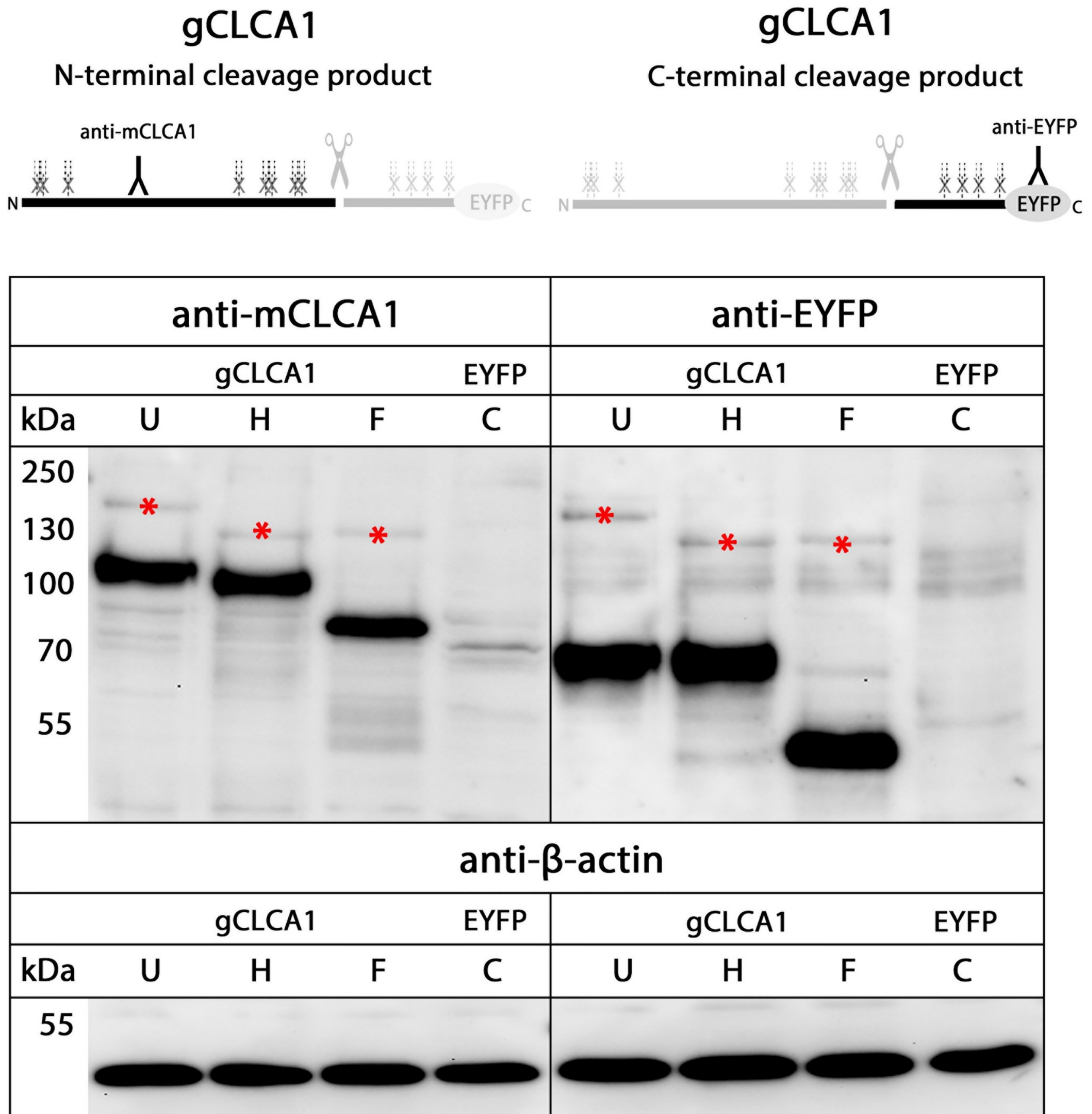


Fig 7. Glycosylated gCLCA1 is cleaved in the medial Golgi. Immunoblot of cell lysate of HEK293 cells transiently transfected with the *gCLCA1Nmabc1* plasmid (gCLCA1) and a control *EYFP*-mock plasmid (EYFP). The gCLCA1 precursor (*) was mannose-rich glycosylated, reduced in size by endo H and PNGase F treatment. In contrast, both cleavage products were solely sensitive to PNGase F as shown by a reduction in size of the N-terminal and C-terminal cleavage product to ~78 kDa or ~52 kDa, respectively. The cleavage event seems to occur in the medial Golgi. U = Untreated, C = untreated mock-control, H = endo H treated, F = PNGase F treated cell lysate. To control for equal total protein loading, the samples were identically immunoblotted with primary anti-beta-actin antibodies. Representative images of three independent experiments.

https://doi.org/10.1371/journal.pone.0266937.g007

The gCLCA1 protein is expressed in enterocytes throughout the intestine

The expression pattern of gCLCA1 was identified on the mRNA and protein levels. *gCLCA1* mRNA was detected in all segments of the alimentary tract (pharynx, esophagus, duodenum, jejunum, ileum, cecum, rectum, coprodeum), the Bursa of Fabricius, the eye, and the lung (S1 File). These tissues were further analyzed via immunofluorescence and immunohistochemistry to identify the cell types that express gCLCA1. The protein was exclusively found at the brush border of intestinal epithelial cells (Fig 8A), which also express villin, a structural marker of enterocytes (Fig 8B, [85]). However, mucin-producing intestinal goblet cells lacked any gCLCA1 expression (Fig 8A). In all segments of the intestinal tract, gCLCA1 protein was detected in enterocytes (Fig 9). Additionally, it was detected along the bursal epithelium (S7 File). No specific gCLCA1 protein signals were detected in any other tissue analyzed.

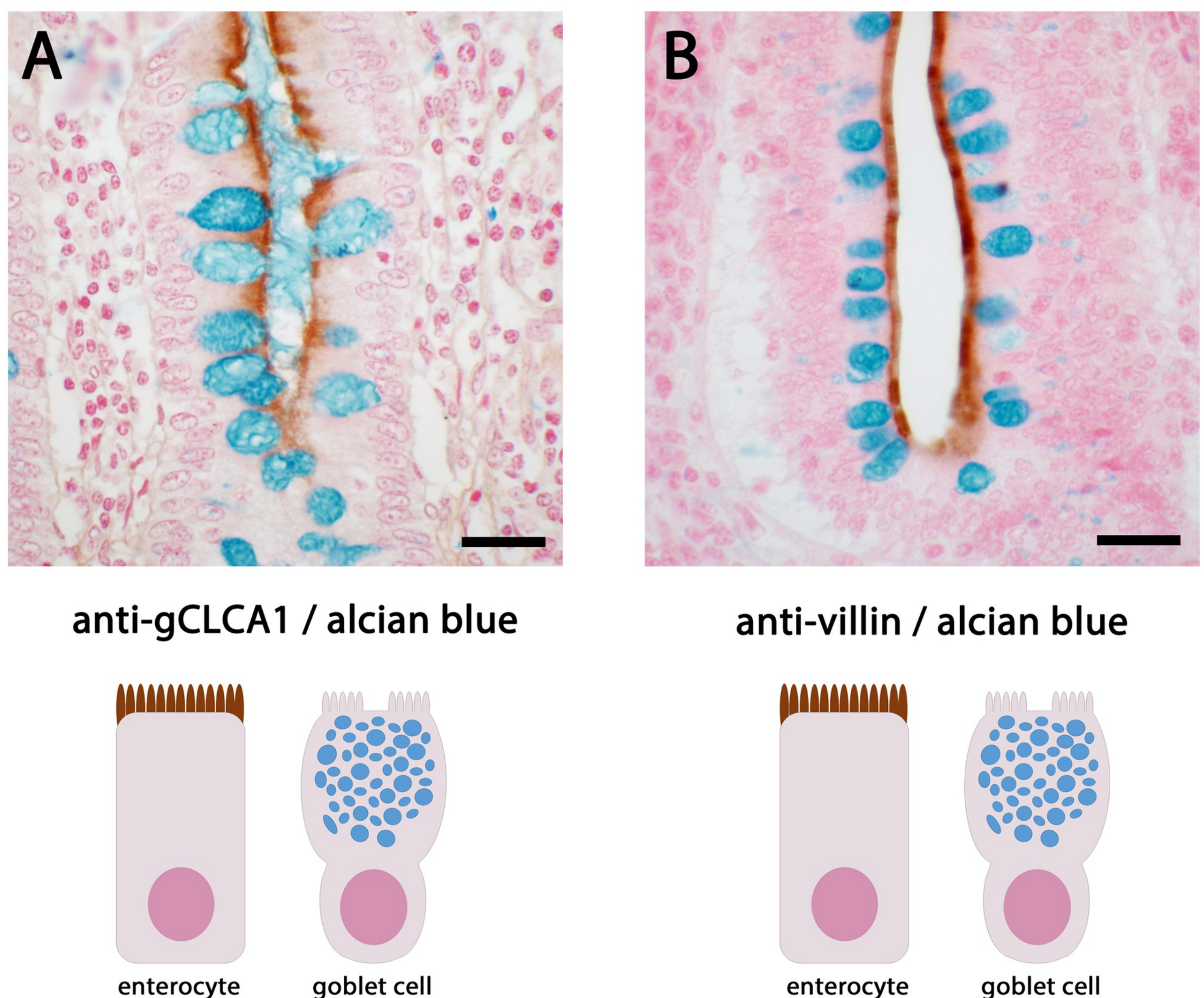


Fig 8. gCLCA1 protein is expressed in the brush border membrane of enterocytes, but not in goblet cells. (A) The gCLCA1 protein (brown) was detected at the apical surface of enterocytes of colonic crypts by immunohistochemistry using the gC1-C1 (anti-gCLCA1 C-terminal) antibody. A counterstain with alcian blue highlights goblet cells (dark blue) that did not show any gCLCA1 protein expression. (B) The cellular expression pattern was similar to villin, a marker for the brush border of enterocytes (brown). Color was developed using DAB as substrate (brown). Bars indicate 20 μm.

https://doi.org/10.1371/journal.pone.0266937.g008

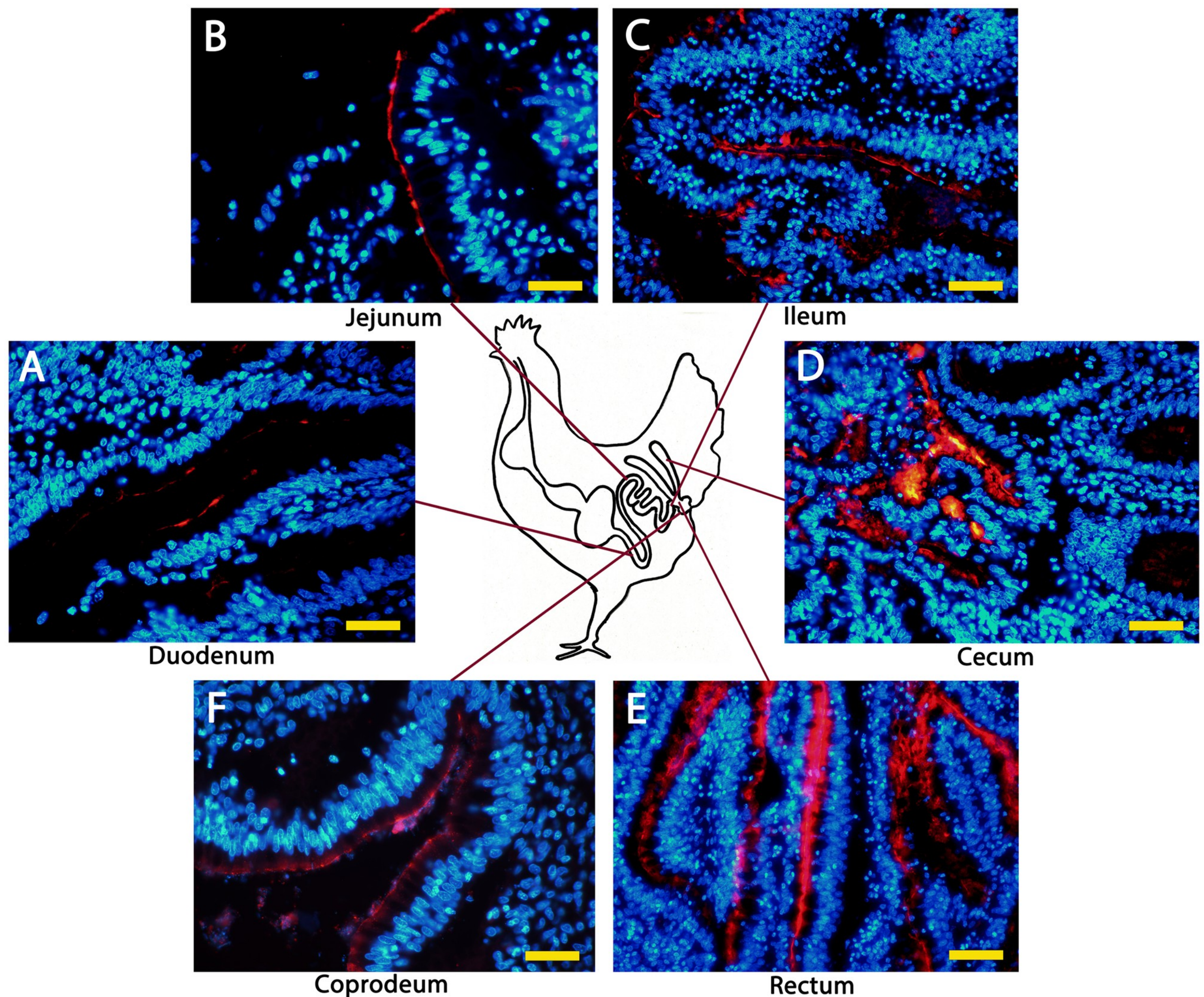


Fig 9. gCLCA1 protein is expressed along the chicken intestine. Anti-gCLCA1 C-terminal antibody was used to detect gCLCA1 along the epithelium in duodenum (A), jejunum (B), ileum (C), cecum (D), rectum (E) and coprodeum (F). Alexa fluor 568-conjugated secondary antibodies and DAPI counterstain (blue). Bars indicate 40 / 100 μm.

https://doi.org/10.1371/journal.pone.0266937.g009

Compared to the expression pattern of CLCA proteins in mammals, the expression of *gCLCA1* showed congruence with mammalian CLCA4 proteins [46,48,62].

The phylogenetic signal of full-length gCLCA1 scatters on the domain level

Phylogenetic analyses of five protein domains reveal a clustering of the CLCA proteins in these domains (Fig 10). Designated mammalian and avian CLCA2 form a outgroup from the avian CLCA1 and mammalian CLCA1, 3, 4. Any of the sequences cluster with the same group throughout the protein. The hierarchy within the mammalian clusters is not fully consistent, due to the known evolutionary dynamics of the rapid formation of major mammalian branches such as primates, rodents, ungulates, or carnivores within a short period of time and the independent expansion at different paces later on. Similar processes seem to have taken place in avian diversification [86] and accordingly, the relationship between the analyzed avian species remains elusive. Only close relationships such as between chicken and quail and, with

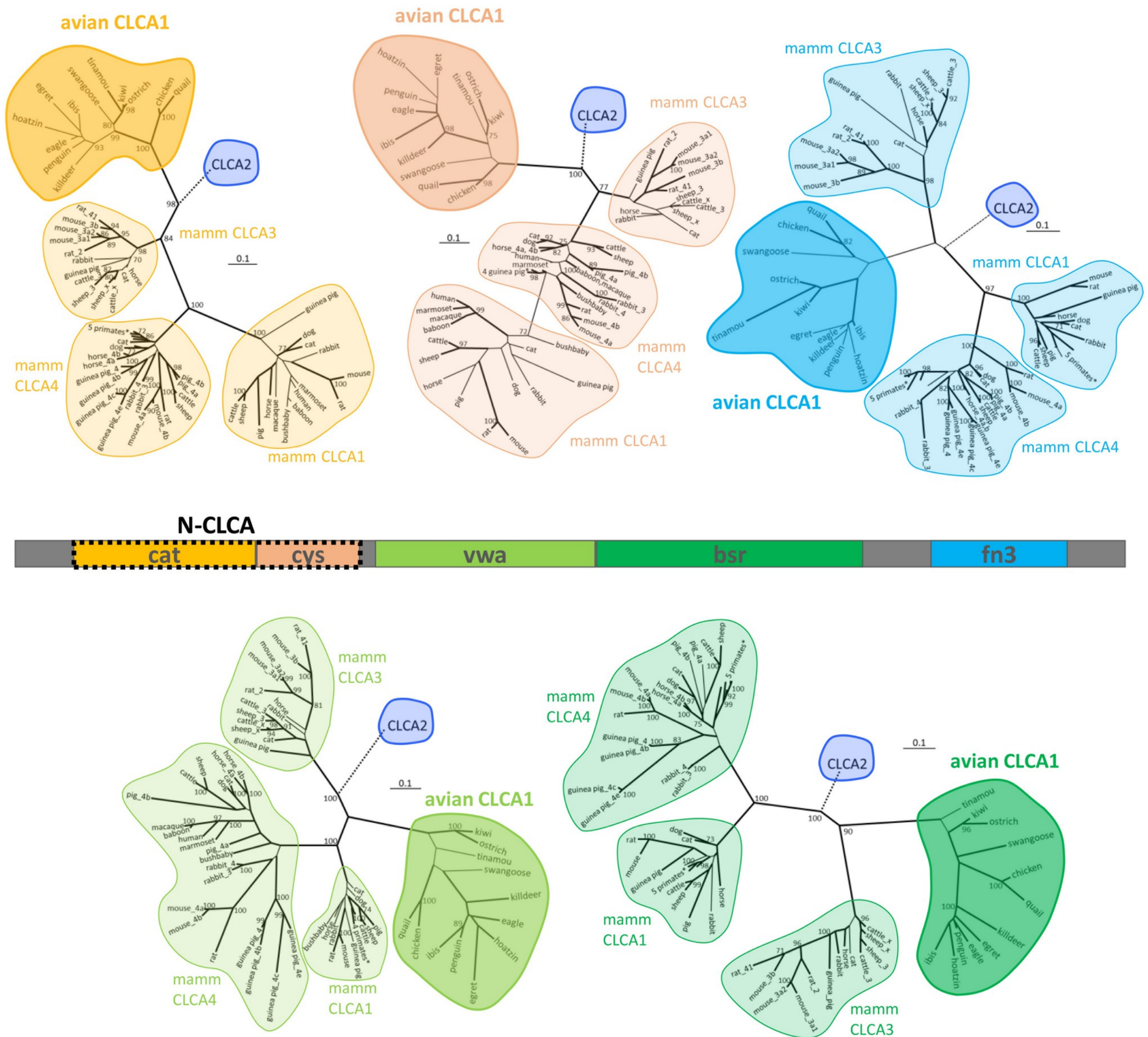


Fig 10. Phylogenetic relationship of CLCA proteins. Maximum likelihood (ml), most parsimony (mp) and 100 neighbor joining (nj) trees were generated for the five distinct CLCA domains and merged. Tree was based on the ml tree, with branches in thick, when occurring in the mp tree and the branch nodes indicated when occurring in more than 70 nj trees. CLCA2 orthologues were reduced to an outgroup. Other clusters were highlighted. To avoid confusion due to limited space, the five primate species analyzed (human, macaque, baboon, marmoset, bushbaby) are occasionally not separately defined (*). In the case of CLCA duplicates in mammalian species, the sequences are designated by the abbreviation given in GenBank (i.e. "mouse_3a2" for murine Clca3a2) or by "x" if the available sequence has been designated as "CLCA-like" or "unknown".

https://doi.org/10.1371/journal.pone.0266937.g010

some restrictions, kiwi, ostrich and tinamou, are consistently reflected throughout the CLCA protein. Adding to the separation of the analyzed clusters from the CLCA2 family, mammalian CLCA1 and CLCA4 form a major branch, independent from avian CLCA1 and mammalian CLCA3. The analysis of the five domains shows variable genetic distances among avian CLCA1 and mammalian CLCA3 and CLCA1/4. The relationship between CLCA2, avian CLCA1, mammalian CLCA3 and CLCA1/4 is therefore inconsistent throughout the full-length protein.

Discussion

In contrast to the mammalian *CLCA* gene family, which consists of up to eight members, there are only two *CLCA* homologues in chicken [51], suggesting that the mammalian and the avian CLCA loci have been subjected to different evolutionary pressures. *CLCA2* is conserved in mammalian and chicken loci [51], whereas the galline CLCA1 displays low genetic distance to all other three mammalian CLCA proteins. Here, we characterized gCLCA1 as a putative prototype of the avian class in terms of protein structure, biochemical properties, and expression pattern and compared the results with the nature of the mammalian CLCA 1, 3 and 4 proteins. The gCLCA1 protein shares features of all CLCA family members such as a signal peptide, a CLCA domain with a HExxH motif, a vWA domain, and an autocatalytic cleavage site (Fig 2, Table 1). The phylogenetic analysis revealed separated CLCA clusters, an evident separation of the avian and mammalian CLCA2 from the other clusters and a consistent relationship between the mammalian CLCA1 and CLCA4 clusters. The position of avian CLCA1 in the phylogenetic tree was difficult to interpret, confirming the previously postulated idea of high dynamics of mammalian CLCA clusters 1, 3, and 4 [51]. It appears that avian CLCA1 and its mammalian relatives may have developed independently from a common ancestor, the avian CLCA1 thus sharing select properties of mammalian CLCA 1, 3, and 4.

Table 1. Comparison of gCLCA1 with the mammalian CLCA 1, 3, and 4.

Trait	*gCLCA1*	*Mammalian CLCA 1*	*Mammalian CLCA 3*	*Mammalian CLCA 4*
Number of exons	*14*	*14*	*14*	*14*
Duplication of genes	-	-	+	+
Pseudogenization	-	-	+	(+)
Signal peptide	+	+	+	+
N-CLCA domain	+	+	+	+
HExxH motif	+	+	+	+
vWA domain	+	+	+	+
Canonical cleavage site	+	+	+	+
TM domain	+	-	-	+
N-terminal subunit secreted in vitro	+	+	+	+
C-terminal subunit secreted in vitro	-	+	+	-
Posttranslational cleavage	+	+	+	+
EQ mutation abrogates cleavage	+	+	*n.t.*	- (*)
N-glycosylation sites	+++	+	++	+++
Protein expressing cell type	*Enterocyte*	*Mucin-producing cells of intestinal, airway and female reproductive tract*	• *Ciliated epithelial airway cell* • *mucus producing airway submucosal cell* • *keratinocyte* • *endothelial cell* • *smooth muscle cell*	• *Enterocyte* • *ciliated epithelial airway cell*

* only partially inhibited.

n.t.: Not tested.

N: Amino terminal.

C: Carboxy terminal.

vWA: Von Willebrand factor type A.

TM: Transmembrane.

https://doi.org/10.1371/journal.pone.0266937.t001

Regarding the cellular expression pattern, gCLCA1 shares more properties with the CLCA4 proteins than with CLCA proteins 3 and 1 (Table 1). gCLCA1 protein was detected at the apical surface of enterocytes along the alimentary canal, and this expression pattern mirrors that of mammalian CLCA4 proteins [46,48,62]. In contrast, mucin-producing goblet cells, which abundantly express CLCA1 in all mammalian species investigated [2,3,49,50,53], did not appear to express gCLCA1. Recently, it has been shown that murine CLCA1 has mucus-processing properties [13] and controls mucus expansion in the colon [12]; the presence of a predicted equivalent protease motif in the secreted N-terminal cleavage product of gCLCA1 suggests that this protein may participate in mucus homeostasis in the chicken intestine. In addition, we found gCLCA1 protein in the epithelial lining of the cloaca and the Bursa of Fabricius, the latter is an immunological organ unique to birds. *gCLCA1* mRNA was present in other organs, such as pharynx, esophagus, eye, and lung, but this was not accompanied by detectable gCLCA1 protein levels.

The presence of a transmembrane domain in the C-terminal subunit, and the secretion of only the N-terminal subunit are other traits shared by gCLCA1 and mammalian CLCA4 proteins. In contrast, mammalian CLCA1 and CLCA3 are soluble proteins and secreted in their entirety [3,64–66]. The function of the different gCLCA1 protein domains as well as the relevance of its glycosylation remain to be investigated.

gCLCA1 constructs with a modified HExxH motif lacked posttranslational autocatalytic cleavage abilities, as reported for mutated human, murine and porcine CLCA1 [77,78]; in contrast, equivalent mutations in the HExxH domain of CLCA4 impair autocatalysis only partially [80]. Thus, whereas gCLCA1 appears to share tissue distribution and cellular expression patterns with mammalian CLCA4, regarding autocatalytic properties it would be functionally closer to mammalian CLCA proteins of cluster 1.

Which characteristics might a common ancestor of gCLCA1 and the mammalian CLCA 1, 3 and 4 proteins have had?

The protein and expressional characteristics of gCLCA1 and the comparison with its mammalian relatives allow us to speculate about the nature of their evolutionary ancestor. The commonly identified traits of gCLCA1 and the mammalian CLCA4 proteins, such as "proteolytic cleavage", "anchoring of the C-terminal subunit in the plasma membrane via a transmembrane (TM) domain", "secretion of the N-terminal subunit", glycosylation and enterocyte-exclusive expression (Table 1) might be considered as a molecular symplesiomorphy and thus be shared also by a common ancestor.

Such a hypothesis might gain some support from more distantly related species. *In silico* analysis of the 940 aa *chloride channel accessory 4 gene 1* (CLCA4.1, NP_001267595.1), a CLCA homologue from *Xenopus laevis*, suggested heavy glycosylation and a strongly hydrophobic C-terminus consistent with a transmembrane domain. RT-qPCR analysis had identified the gut as an expressing tissue, however, the expressing cell type of this CLCA member is still unknown [87]. Our comparison of the aa sequence of Xenopus CLCA4.1 (xCLCA4.1) with gCLCA1 and proteins of the mammalian CLCA 1, 3, and 4 identified a proteolytically active HExxH site as well as a canonical cleavage site at position 696 (RSR-ALY) in the amphibian CLCA member, which suggests a putative posttranslational cleavage of the protein. Thus, a common CLCA ancestor might have been a glycosylated and cleavable transmembrane protein expressed in the intestine.

The three mammalian *CLCA* 1, 3, and 4, seem to have developed from duplication events that did not happen in chicken. Gene duplication, as a key driver of evolution [88], is documented in nine vertebrate species for at least 37,000 genes [89]. Either, as the most common

case [90], the duplication is completely lost or the duplicated gene becomes inactivated, resulting in pseudogenization. Pseudogenization is known for the *CLCA3* gene in certain mammals such as dry-nose primates, pigs, and sheep [51]. On the other hand, the duplication can be maintained and adapted, resulting in the gain of a new function (neofunctionalization) or the original function is now shared among both, the original and the duplicated gene (subfunctionalization) [90]. Additional species-specific duplication events of clusters 3 and 4 have been described in mammals. Three and two intact *CLCA3* genes are present in the murine or bovine CLCA locus, respectively [51]. Duplication of the *CLCA4* gene has also been identified in the pig and the mouse [51]. The duplication of functional *CLCA* genes in mammals seems to amplify the expression pattern of this gene family. In contrast to the expression of the galline CLCA1 and mammalian CLCA4 proteins in enterocytes, CLCA1 proteins are expressed by mucin-producing cells such as goblet cells of the intestinal and respiratory tract [2,3,49,53]. The expression pattern of the duplicated CLCA3 members is broader than of other clusters [51] and is not restricted to epithelial cells. CLCA3 members are also expressed in cells derived from the mesoderm such as endothelial or smooth muscle cells [42]. In addition to the diversity of the cellular expression pattern of mammalian CLCA proteins, structural variations in terms of the lack of a transmembrane domain are evident in the CLCA proteins 1 and 3 compared to 4 and the gCLCA1. The structural differences of the CLCA proteins and the diverse expression pattern may reflect that the *CLCA* gene duplication events in mammals resulted in a sub- or even neofunctionalization of the genes. The chicken lacks the diversity of mammalian *CLCA* genes and the fully secreted CLCA 1/3 homologues seem to be dispensable for the chicken. In contrast, the expression of a glycosylated, a single TM domain containing CLCA protein like gCLCA1 or mammalian CLCA4 in the intestine suggests a taxon spanning, integral trait of CLCA homologues in this anatomical section.

Supporting information

S1 File. RT-qPCR of *gCLCA1* and *PGK1*.
(DOCX)

S2 File. In-detail protocols.
(DOCX)

S3 File. Testing of anti-gCLCA1 primary antibodies.
(DOCX)

S4 File. SNPs of the *gCLCA1* clone used in this study.
(DOCX)

S5 File. Detection of the N-terminal cleavage products of gCLCA1 in the cell supernatant using liquid chromatography-tandem mass spectrometry (LC-MS/MS).
(DOCX)

S6 File. EYFP- and HEK293 autofluorescence.
(DOCX)

S7 File. gCLCA1 protein expression at the apical brush border of bursal surface epithelium (red).
(DOCX)

S1 Raw images.
(PDF)

Acknowledgments

The excellent technical support of Annett Kannegießer, Michaela Scholz, Charlene Lamprecht, Nicole Huth, Angela Linke, Cornelia Ziegler, Jana Enders, Sophie Merz, Laura A. Eisenhauer, and Mandana J. Poetsch is greatly appreciated. We would like to acknowledge the assistance of the Core Facility BioSupraMol supported by the German Research Foundation. This work is part of the Ph.D. thesis of F.B.

Author Contributions

Conceptualization: Florian Bartenschlager, Nikolai Klymiuk, Achim D. Gruber, Lars Mundhenk.

Data curation: Florian Bartenschlager, Nikolai Klymiuk, Christoph Weise, Benno Kuropka, Lars Mundhenk.

Formal analysis: Florian Bartenschlager, Nikolai Klymiuk, Christoph Weise, Benno Kuropka, Lars Mundhenk.

Investigation: Florian Bartenschlager, Nikolai Klymiuk.

Methodology: Florian Bartenschlager, Nikolai Klymiuk, Christoph Weise, Benno Kuropka, Lars Mundhenk.

Project administration: Florian Bartenschlager, Achim D. Gruber, Lars Mundhenk.

Resources: Achim D. Gruber.

Supervision: Achim D. Gruber, Lars Mundhenk.

Validation: Florian Bartenschlager.

Visualization: Florian Bartenschlager, Nikolai Klymiuk.

Writing – original draft: Florian Bartenschlager, Nikolai Klymiuk, Benno Kuropka, Lars Mundhenk.

Writing – review & editing: Christoph Weise, Achim D. Gruber.

References

1. Nakanishi A, Morita S, Iwashita H, Sagiya Y, Ashida Y, Shirafuji H, et al. Role of gob-5 in mucus overproduction and airway hyperresponsiveness in asthma. Proceedings of the National Academy of Sciences. 2001; 98(9):5175–80. https://doi.org/10.1073/pnas.081510898 PMID: 11296262
2. Hoshino M, Morita S, Iwashita H, Sagiya Y, Nagi T, Nakanishi A, et al. Increased expression of the human Ca2+-activated Cl− channel 1 (CaCC1) gene in the asthmatic airway. American journal of respiratory and critical care medicine. 2002; 165(8):1132–6. https://doi.org/10.1164/ajrccm.165.8.2107068 PMID: 11956057
3. Range F, Mundhenk L, Gruber A. A soluble secreted glycoprotein (eCLCA1) is overexpressed due to goblet cell hyperplasia and metaplasia in horses with recurrent airway obstruction. Veterinary pathology. 2007; 44(6):901–11. https://doi.org/10.1354/vp.44-6-901 PMID: 18039903
4. Ritzka M, Stanke F, Jansen S, Gruber AD, Pusch L, Woelfl S, et al. The CLCA gene locus as a modulator of the gastrointestinal basic defect in cystic fibrosis. Human genetics. 2004; 115(6):483–91. https://doi.org/10.1007/s00439-004-1190-y PMID: 15490240
5. Hauber HP, Manoukian JJ, Nguyen LH, Sobol SE, Levitt RC, Holroyd KJ, et al. Increased expression of interleukin-9, interleukin-9 receptor, and the calcium-activated chloride channel hCLCA1 in the upper airways of patients with cystic fibrosis. The Laryngoscope. 2003; 113(6):1037–42. https://doi.org/10.1097/00005537-200306000-00022 PMID: 12782818
6. Young FD, Newbigging S, Choi C, Keet M, Kent G, Rozmahel RF. Amelioration of cystic fibrosis intestinal mucous disease in mice by restoration of mCLCA3. Gastroenterology. 2007; 133(6):1928–37. e3. https://doi.org/10.1053/j.gastro.2007.10.007 PMID: 18054564

7. Walia V, Ding M, Kumar S, Nie D, Premkumar LS, Elble RC. hCLCA2 is a p53-inducible inhibitor of breast cancer cell proliferation. Cancer research. 2009; 69(16):6624–32. https://doi.org/10.1158/0008-5472.CAN-08-4101 PMID: 19654313

8. Walia V, Yu Y, Cao D, Sun M, McLean JR, Hollier BG, et al. Loss of breast epithelial marker hCLCA2 promotes epithelial-to-mesenchymal transition and indicates higher risk of metastasis. Oncogene. 2012; 31(17):2237–46. https://doi.org/10.1038/onc.2011.392 PMID: 21909135

9. Yu Y, Walia V, Elble RC. Loss of CLCA4 promotes epithelial-to-mesenchymal transition in breast cancer cells. PloS one. 2013; 8(12):e83943. https://doi.org/10.1371/journal.pone.0083943 PMID: 24386311

10. Hou T, Zhou L, Wang L, Kazobinka G, Zhang X, Chen Z. CLCA4 inhibits bladder cancer cell proliferation, migration, and invasion by suppressing the PI3K/AKT pathway. Oncotarget. 2017; 8(54):93001. https://doi.org/10.18632/oncotarget.21724 PMID: 29190973

11. Chen H, Liu Y, Jiang C-J, Chen Y-M, Li H, Liu Q-A. Calcium-activated chloride channel A4 (CLCA4) plays inhibitory roles in invasion and migration through suppressing epithelial-mesenchymal transition via PI3K/AKT signaling in colorectal cancer. Medical science monitor: international medical journal of experimental and clinical research. 2019; 25:4176. https://doi.org/10.12659/MSM.914195 PMID: 31164625

12. Nyström EE, Birchenough GM, van der Post S, Arike L, Gruber AD, Hansson GC, et al. Calcium-activated chloride channel regulator 1 (CLCA1) controls mucus expansion in colon by proteolytic activity. EBioMedicine. 2018; 33:134–43. https://doi.org/10.1016/j.ebiom.2018.05.031 PMID: 29885864

13. Nyström EE, Arike L, Ehrencrona E, Hansson GC, Johansson ME. Calcium-activated chloride channel regulator 1 (CLCA1) forms non-covalent oligomers in colonic mucus and has mucin 2–processing properties. Journal of Biological Chemistry. 2019; 294(45):17075–89. https://doi.org/10.1074/jbc.RA119.009940 PMID: 31570526

14. Ching JC, Lobanova L, Loewen ME. Secreted hCLCA1 is a signaling molecule that activates airway macrophages. PloS one. 2013; 8(12):e83130. https://doi.org/10.1371/journal.pone.0083130 PMID: 24349445

15. Erickson NA, Dietert K, Enders J, Glauben R, Nouailles G, Gruber AD, et al. Soluble mucus component CLCA1 modulates expression of leukotactic cytokines and BPIFA1 in murine alveolar macrophages but not in bone marrow-derived macrophages. Histochemistry and cell biology. 2018; 149(6):619–33. https://doi.org/10.1007/s00418-018-1664-y PMID: 29610986

16. Keith BA, Ching JC, Loewen ME. Von Willebrand Factor Type A domain of hCLCA1 is sufficient for U-937 macrophage activation. Biochemistry and biophysics reports. 2019; 18:100630. https://doi.org/10.1016/j.bbrep.2019.100630 PMID: 30984882

17. Sala-Rabanal M, Yurtsever Z, Nichols CG, Brett TJ. Secreted CLCA1 modulates TMEM16A to activate Ca2+-dependent chloride currents in human cells. elife. 2015; 4:e05875.

18. Sharma A, Ramena G, Yin Y, Premkumar L, Elble RC. CLCA2 is a positive regulator of store-operated calcium entry and TMEM16A. PLoS One. 2018; 13(5):e0196512. https://doi.org/10.1371/journal.pone.0196512 PMID: 29758025

19. Morton J, Alevy Y, Peltz G, Holtzman M. Functional genomics of asthma: role of CLCA3 in goblet cell metaplasia (GCM) but not airway hyperreactivity (AHR). Journal of Allergy and Clinical Immunology. 2004; 113(2):S205.

20. Patel AC, Morton JD, Kim EY, Alevy Y, Swanson S, Tucker J, et al. Genetic segregation of airway disease traits despite redundancy of calcium-activated chloride channel family members. Physiological genomics. 2006; 25(3):502–13. https://doi.org/10.1152/physiolgenomics.00321.2005 PMID: 16569774

21. Patel A, Battaile J, Alevy Y, Patterson G, Swanson S. Homologous mouse Clca3, 5, 6 and human hCLCA1, 2, 4 gene clusters control mucous cell metaplasia. Am J Respir Crit Care Med. 2007; 175: A499.

22. Elble RC, Pauli BU. Tumor suppression by a proapoptotic calcium-activated chloride channel in mammary epithelium. Journal of Biological Chemistry. 2001; 276(44):40510–7. https://doi.org/10.1074/jbc.M104821200 PMID: 11483609

23. Beckley JR, Pauli BU, Elble RC. Re-expression of detachment-inducible chloride channel mCLCA5 suppresses growth of metastatic breast cancer cells. Journal of Biological Chemistry. 2004; 279(40):41634–41. https://doi.org/10.1074/jbc.M408334200 PMID: 15292178

24. Tanikawa C, Nakagawa H, Furukawa Y, Nakamura Y, Matsuda K. CLCA2 as a p53-inducible senescence mediator. Neoplasia. 2012; 14(2):141–9. https://doi.org/10.1593/neo.111700 PMID: 22431922

25. Rovillain E, Mansfield L, Caetano C, Alvarez-Fernandez M, Caballero OL, Medema RH, et al. Activation of nuclear factor-kappa B signalling promotes cellular senescence. Oncogene. 2011; 30(20):2356–66. https://doi.org/10.1038/onc.2010.611 PMID: 21242976

26. Seltmann K, Meyer M, Sulcova J, Kockmann T, Wehkamp U, Weidinger S, et al. Humidity-regulated CLCA2 protects the epidermis from hyperosmotic stress. Science translational medicine. 2018; 10 (440). https://doi.org/10.1126/scitranslmed.aao4650 PMID: 29743348

27. Hamalainen L, Bart G, Takabe P, Rauhala L, Deen A, Pasonen-Seppanen S, et al. The calcium-activated chloride channel-associated protein rCLCA2 is expressed throughout rat epidermis, facilitates apoptosis and is downmodulated by UVB. Histochem Cell Biol. 2021. https://doi.org/10.1007/s00418-021-01962-5 PMID: 33486586

28. Abdel-Ghany M, Cheng H-C, Elble RC, Pauli BU. The breast cancer β4 integrin and endothelial human CLCA2 mediate lung metastasis. Journal of Biological Chemistry. 2001; 276(27):25438–46.

29. Abdel-Ghany M, Cheng H-C, Elble RC, Lin H, DiBiasio J, Pauli BU. The interacting binding domains of the β4 integrin and calcium-activated chloride channels (CLCAs) in metastasis. Journal of biological chemistry. 2003; 278(49):49406–16.

30. Gruber AD, Elble RC, Ji H-L, Schreur KD, Fuller CM, Pauli BU. Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca2+-activated Cl− channel proteins. Genomics. 1998; 54(2):200–14. https://doi.org/10.1006/geno.1998.5562 PMID: 9828122

31. Gruber AD, Pauli BU. Tumorigenicity of human breast cancer is associated with loss of the Ca2+-activated chloride channel CLCA2. Cancer research. 1999; 59(21):5488–91. PMID: 10554024

32. Gruber AD, Schreur KD, Ji H-L, Fuller CM, Pauli BU. Molecular cloning and transmembrane structure of hCLCA2 from human lung, trachea, and mammary gland. American Journal of Physiology-Cell Physiology. 1999; 276(6):C1261–C70.

33. Agnel M, Vermat T, Culouscou J-M. Identification of three novel members of the calcium-dependent chloride channel (CaCC) family predominantly expressed in the digestive tract and trachea. FEBS letters. 1999; 455(3):295–301. https://doi.org/10.1016/s0014-5793(99)00891-1 PMID: 10437792

34. Bustin SA, Li S-R, Dorudi S. Expression of the Ca2+-activated chloride channel genes CLCA1 and CLCA2 is downregulated in human colorectal cancer. DNA and cell biology. 2001; 20(6):331–8. https://doi.org/10.1089/10445490152122442 PMID: 11445004

35. Kamada F, Suzuki Y, Shao C, Tamari M, Hasegawa K, Hirota T, et al. Association of the hCLCA1 gene with childhood and adult asthma. Genes & Immunity. 2004; 5(7):540–7. https://doi.org/10.1038/sj.gene.6364124 PMID: 15318163

36. Van Der Doef H, Slieker M, Staab D, Alizadeh B, Seia M, Colombo C, et al. Association of the CLCA1 p. S357N variant with meconium ileus in European patients with cystic fibrosis. Journal of pediatric gastroenterology and nutrition. 2010; 50(3):347–9. https://doi.org/10.1097/MPG.0b013e3181afce6c PMID: 20179644

37. Pan X, Wang Q, Xu C, Yan L, Pang S, Gan J. Prognostic value of chloride channel accessory mRNA expression in colon cancer. Oncology letters. 2019; 18(3):2967–76. https://doi.org/10.3892/ol.2019.10615 PMID: 31404307

38. Gandhi R, Elble R, Gruber A, Schreur K, Ji H-L, Fuller C, et al. Molecular and functional characterization of a calcium-sensitive chloride channel from mouse lung. Journal of Biological Chemistry. 1998; 273 (48):32096–101. https://doi.org/10.1074/jbc.273.48.32096 PMID: 9822685

39. Romio L, Musante L, Cinti R, Seri M, Moran O, Zegarra-Moran O, et al. Characterization of a murine gene homologous to the bovine CaCC chloride channel. Gene. 1999; 228(1–2):181–8. https://doi.org/10.1016/s0378-1119(98)00620-9 PMID: 10072771

40. Lee D, Ha S, Kho Y, Kim J, Cho K, Baik M, et al. Induction of mouse Ca2+-sensitive chloride channel 2 gene during involution of mammary gland. Biochemical and biophysical research communications. 1999; 264(3):933–7. https://doi.org/10.1006/bbrc.1999.1583 PMID: 10544033

41. Komiya T, Tanigawa Y, Hirohashi S. Cloning and identification of the gene gob-5, which is expressed in intestinal goblet cells in mice. Biochemical and biophysical research communications. 1999; 255 (2):347–51. https://doi.org/10.1006/bbrc.1999.0168 PMID: 10049711

42. Elble RC, Ji G, Nehrke K, DeBiasio J, Kingsley PD, Kotlikoff MI, et al. Molecular and functional characterization of a murine calcium-activated chloride channel expressed in smooth muscle. Journal of Biological Chemistry. 2002; 277(21):18586–91. https://doi.org/10.1074/jbc.M200829200 PMID: 11896056

43. Evans SR, Thoreson WB, Beck CL. Molecular and functional analyses of two new calcium-activated chloride channel family members from mouse eye and intestine. Journal of Biological Chemistry. 2004; 279(40):41792–800. https://doi.org/10.1074/jbc.M408354200 PMID: 15284223

44. Al-Jumaily M, Kozlenkov A, Mechaly I, Fichard A, Matha V, Scamps F, et al. Expression of three distinct families of calcium-activated chloride channel genes in the mouse dorsal root ganglion. Neuroscience bulletin. 2007; 23(5):293–9. https://doi.org/10.1007/s12264-007-0044-8 PMID: 17952139

45. Gaspar K, Racette K, Gordon J, Loewen M, Forsyth G. Cloning a chloride conductance mediator from the apical membrane of porcine ileal enterocytes. Physiological genomics. 2000; 3(2):101–11. https://doi.org/10.1152/physiolgenomics.2000.3.2.101 PMID: 11015605
46. Plog S, Grötzsch T, Klymiuk N, Kobalz U, Gruber AD, Mundhenk L. The porcine chloride channel calcium-activated family member pCLCA4a mirrors lung expression of the human hCLCA4. Journal of Histochemistry & Cytochemistry. 2012; 60(1):45–56. https://doi.org/10.1369/0022155411426455 PMID: 22205680
47. Plog S, Mundhenk L, Langbein L, Gruber AD. Synthesis of porcine pCLCA2 protein during late differentiation of keratinocytes of epidermis and hair follicle inner root sheath. Cell and tissue research. 2012; 350(3):445–53. https://doi.org/10.1007/s00441-012-1482-9 PMID: 22968961
48. Plog S, Klymiuk N, Binder S, Van Hook MJ, Thoreson WB, Gruber AD, et al. Naturally Occurring Deletion Mutants of the Pig-Specific, Intestinal Crypt Epithelial Cell Protein CLCA4b without Apparent Phenotype. PloS one. 2015; 10(10):e0140050. https://doi.org/10.1371/journal.pone.0140050 PMID: 26474299
49. Plog S, Mundhenk L, Klymiuk N, Gruber AD. Genomic, tissue expression, and protein characterization of pCLCA1, a putative modulator of cystic fibrosis in the pig. Journal of Histochemistry & Cytochemistry. 2009; 57(12):1169–81.
50. Anton F, Leverkoehne I, Mundhenk L, Thoreson WB, Gruber AD. Overexpression of eCLCA1 in small airways of horses with recurrent airway obstruction. Journal of Histochemistry & Cytochemistry. 2005; 53(8):1011–21. https://doi.org/10.1369/jhc.4A6599.2005 PMID: 15879574
51. Mundhenk L, Erickson NA, Klymiuk N, Gruber AD. Interspecies diversity of chloride channel regulators, calcium-activated 3 genes. PloS one. 2018; 13(1):e0191512. https://doi.org/10.1371/journal.pone.0191512 PMID: 29346439
52. Erickson N, Gruber A, Mundhenk L. The family of chloride channel regulator, calcium-activated proteins in the feline respiratory tract: a comparative perspective on airway Diseases in man and animal models. Journal of comparative pathology. 2020; 174:39–53. https://doi.org/10.1016/j.jcpa.2019.10.193 PMID: 31955802
53. Leverkoehne I, Gruber AD. The murine mCLCA3 (alias gob-5) protein is located in the mucin granule membranes of intestinal, respiratory, and uterine goblet cells. Journal of Histochemistry & Cytochemistry. 2002; 50(6):829–38.
54. KOEGEL H, ALZHEIMER C. Expression and biological significance of Ca2+-activated ion channels in human keratinocytes. The FASEB Journal. 2001; 15(1):145–54. https://doi.org/10.1096/fj.00-0055com PMID: 11149902
55. Connon CJ, Yamasaki K, Kawasaki S, Quantock AJ, Koizumi N, Kinoshita S. Calcium-activated chloride channel-2 in human epithelia. Journal of Histochemistry & Cytochemistry. 2004; 52(3):415–8. https://doi.org/10.1177/002215540405200313 PMID: 14966209
56. Braun J, Bothe MK, Mundhenk L, Beck CL, Gruber AD. Murine mCLCA5 is expressed in granular layer keratinocytes of stratified epithelia. Histochem Cell Biol. 2010; 133(3):285–99. https://doi.org/10.1007/s00418-009-0667-0 PMID: 20012443
57. Gruber AD, Gandhi R, Pauli BU. The murine calcium-sensitive chloride channel (mCaCC) is widely expressed in secretory epithelia and in other select tissues. Histochem Cell Biol. 1998; 110(1):43–9. https://doi.org/10.1007/s004180050263 PMID: 9681688
58. Abdel-Ghany M, Cheng H-C, Elble RC, Pauli BU. Focal adhesion kinase activated by β4 integrin ligation to mCLCA1 mediates early metastatic growth. Journal of Biological Chemistry. 2002; 277(37):34391–400. https://doi.org/10.1074/jbc.M205307200 PMID: 12110680
59. Furuya M, Kirschbaum SB, Paulovich A, Pauli BU, Zhang H, Alexander JS, et al. Lymphatic endothelial murine chloride channel calcium-activated 1 is a ligand for leukocyte LFA-1 and Mac-1. The Journal of Immunology. 2010; 185(10):5769–77. https://doi.org/10.4049/jimmunol.1002226 PMID: 20937843
60. Roussa E, Wittschen P, Wolff NA, Torchalski B, Gruber AD, Thévenod F. Cellular distribution and subcellular localization of mCLCA1/2 in murine gastrointestinal epithelia. Journal of Histochemistry & Cytochemistry. 2010; 58(7):653–68. https://doi.org/10.1369/jhc.2010.955211 PMID: 20385786
61. Leverkoehne I, Horstmeier BA, von Samson-Himmelstjerna G, Scholte BJ, Gruber AD. Real-time RT-PCR quantitation of mCLCA1 and mCLCA2 reveals differentially regulated expression in pre- and postnatal murine tissues. Histochem Cell Biol. 2002; 118(1):11–7. https://doi.org/10.1007/s00418-002-0420-4 PMID: 12122442
62. Bothe MK, Braun J, Mundhenk L, Gruber AD. Murine mCLCA6 is an integral apical membrane protein of non-goblet cell enterocytes and co-localizes with the cystic fibrosis transmembrane conductance regulator. Journal of Histochemistry & Cytochemistry. 2008; 56(5):495–509. https://doi.org/10.1369/jhc.2008.950592 PMID: 18285349

63. Mall M, Gonska T, Thomas J, Schreiber R, Seydewitz HH, Kuehr J, et al. Modulation of Ca2+-activated Cl- secretion by basolateral K+ channels in human normal and cystic fibrosis airway epithelia. Pediatr Res. 2003; 53(4):608–18. https://doi.org/10.1203/01.PDR.0000057204.51420.DC PMID: 12612194

64. Patel AC, Brett TJ, Holtzman MJ. The role of CLCA proteins in inflammatory airway disease. Annu Rev Physiol. 2009; 71:425–49. https://doi.org/10.1146/annurev.physiol.010908.163253 PMID: 18954282

65. Gibson A, Lewis AP, Affleck K, Aitken AJ, Meldrum E, Thompson N. hCLCA1 and mCLCA3 are secreted non-integral membrane proteins and therefore are not ion channels. Journal of Biological Chemistry. 2005; 280(29):27205–12.

66. Mundhenk L, Alfalah M, Elble RC, Pauli BU, Naim HY, Gruber AD. Both cleavage products of the mCLCA3 protein are secreted soluble proteins. Journal of Biological Chemistry. 2006; 281(40):30072–80. https://doi.org/10.1074/jbc.M606489200 PMID: 16895902

67. Elble RC, Walia V, Cheng H-c, Connon CJ, Mundhenk L, Gruber AD, et al. The putative chloride channel hCLCA2 has a single C-terminal transmembrane segment. Journal of Biological Chemistry. 2006; 281(40):29448–54.

68. Lenart A, Dudkiewicz M, Grynberg M, Pawłowski K. CLCAs-a family of metalloproteases of intriguing phylogenetic distribution and with cases of substituted catalytic sites. PLoS One. 2013; 8(5):e62272. https://doi.org/10.1371/journal.pone.0062272 PMID: 23671590

69. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: the conserved domain database in 2020. Nucleic acids research. 2020; 48(D1):D265–D8. https://doi.org/10.1093/nar/gkz991 PMID: 31777944

70. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. Nucleic acids research. 2018; 46(W1):W200–W4. https://doi.org/10.1093/nar/gky448 PMID: 29905871

71. Gough J. The SUPERFAMILY database in structural genomics. Acta Crystallographica Section D: Biological Crystallography. 2002; 58(11):1897–900. https://doi.org/10.1107/s0907444902015160 PMID: 12393919

72. Pandurangan AP, Stahlhacke J, Oates ME, Smithers B, Gough J. The SUPERFAMILY 2.0 database: a significant proteome update and a new webserver. Nucleic acids research. 2019; 47(D1):D490–D4. https://doi.org/10.1093/nar/gky1130 PMID: 30445555

73. Käll L, Krogh A, Sonnhammer EL. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. Nucleic acids research. 2007; 35(suppl_2):W429–W32. https://doi.org/10.1093/nar/gkm256 PMID: 17483518

74. Hirokawa T, Boon-Chieng S, Mitaku S. SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics (Oxford, England). 1998; 14(4):378–9. https://doi.org/10.1093/bioinformatics/14.4.378 PMID: 9632836

75. Dobson L, Reményi I, Tusnády GE. CCTOP: a Consensus Constrained TOPology prediction web server. Nucleic acids research. 2015; 43(W1):W408–W12. https://doi.org/10.1093/nar/gkv451 PMID: 25943549

76. Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nature biotechnology. 2019; 37(4):420–3. https://doi.org/10.1038/s41587-019-0036-z PMID: 30778233

77. Pawłowski K, Lepistö M, Meinander N, Sivars U, Varga M, Wieslander E. Novel conserved hydrolase domain in the CLCA family of alleged calcium-activated chloride channels. Proteins: Structure, Function, and Bioinformatics. 2006; 63(3):424–39. https://doi.org/10.1002/prot.20887 PMID: 16470849

78. Bothe MK, Mundhenk L, Kaup M, Weise C, Gruber AD. The murine goblet cell protein mCLCA3 is a zinc-dependent metalloprotease with autoproteolytic activity. Molecules and cells. 2011; 32(6):535–41. https://doi.org/10.1007/s10059-011-0158-8 PMID: 22080371

79. Olias P, Adam I, Meyer A, Scharff C, Gruber AD. Reference genes for quantitative gene expression studies in multiple avian species. PloS one. 2014; 9(6):e99678. https://doi.org/10.1371/journal.pone.0099678 PMID: 24926893

80. Bothe MK, Mundhenk L, Beck CL, Kaup M, Gruber AD. Impaired autoproteolytic cleavage of mCLCA6, a murine integral membrane protein expressed in enterocytes, leads to cleavage at the plasma membrane instead of the endoplasmic reticulum. Molecules and cells. 2012; 33(3):251–7. https://doi.org/10.1007/s10059-012-2217-1 PMID: 22350745

81. Plog S, Mundhenk L, Bothe MK, Klymiuk N, Gruber AD. Tissue and cellular expression patterns of porcine CFTR: similarities to and differences from human CFTR. Journal of Histochemistry & Cytochemistry. 2010; 58(9):785–97.

82. Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K, et al. Database resources of the national center for biotechnology information. Nucleic acids research. 2019; 47(Database issue):D23. https://doi.org/10.1093/nar/gky1069 PMID: 30395293

83. Yurtsever Z, Sala-Rabanal M, Randolph DT, Scheaffer SM, Roswit WT, Alevy YG, et al. Self-cleavage of human CLCA1 protein by a novel internal metalloprotease domain controls calcium-activated chloride channel activation. Journal of Biological Chemistry. 2012; 287(50):42138–49. https://doi.org/10.1074/jbc.M112.410282 PMID: 23112050

84. Pult F, Fallah G, Braam U, Detro-Dassen S, Niculescu C, Laube B, et al. Robust post-translocational N-glycosylation at the extreme C-terminus of membrane and secreted proteins in Xenopus laevis oocytes and HEK293 cells. Glycobiology. 2011; 21(9):1147–60. https://doi.org/10.1093/glycob/cwr013 PMID: 21303814

85. Coudrier E, Arpin M, Dudouet B, Finidori J, Garcia A, Huet C, et al. Villin as a structural marker to study the assembly of the brush border. Protoplasma. 1988; 145(2–3):99–105.

86. Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO. The global diversity of birds in space and time. Nature. 2012; 491(7424):444–8. https://doi.org/10.1038/nature11631 PMID: 23123857

87. Lee RM, Ryu RH, Jeong SW, Oh SJ, Huang H, Han JS, et al. Isolation and Expression Profile of the Ca-Activated Chloride Channel-like Membrane Protein 6 Gene in Xenopus laevis. Laboratory animal research. 2011; 27(2):109–16. https://doi.org/10.5625/lar.2011.27.2.109 PMID: 21826170

88. Taylor JS, Raes J. Duplication and divergence: the evolution of new genes and old ideas. Annu Rev Genet. 2004; 38:615–43. https://doi.org/10.1146/annurev.genet.38.072902.092831 PMID: 15568988

89. Ouedraogo M, Bettembourg C, Bretaudeau A, Sallou O, Diot C, Demeure O, et al. The duplicated genes database: identification and functional annotation of co-localised duplicated genes across genomes. PloS one. 2012; 7(11):e50653. https://doi.org/10.1371/journal.pone.0050653 PMID: 23209799

90. Conant GC, Wolfe KH. Turning a hobby into a job: how duplicated genes find new functions. Nature Reviews Genetics. 2008; 9(12):938–50. https://doi.org/10.1038/nrg2482 PMID: 19015656

1.7 Genomic, biochemical and expressional properties reveal strong conservation of the *CLCA2* gene in birds and mammals

Authors: Florian Bartenschlager, Nikolai Klymiuk, Achim D Gruber, Lars Mundhenk

Year: 2022

Journal: PeerJ

DOI: 10.7717/peerj.14202

Bibliographic Source: Bartenschlager, F., Klymiuk, N., Gruber A.D., Mundhenk L. (2022): Genomic, biochemical and expressional properties reveal strong conservation of the CLCA2 gene in birds and mammals. PeerJ 10:e14202.

No permission by *PeerJ* is required for inclusion in this thesis as all content published by *PeerJ* is distributed under the terms of the Creative Commons Attribution License. This permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Declaration of own portion of work in this research publication:

Contributions of F. Bartenschlager:

1. Development and conceptualization of the study design with planning, preparation, and project administration including experimental planning, preparation, performance, analysis, evaluation, validation and data curation of *in silico* gene, genome and protein analyses as well as cloning, RT-qPCR, cell culture, western blot, immunocytochemistry, immunohistochemistry and immunofluorescence experiments.
2. Visualization of experimental results and subsequent writing of original draft including supplementary files.

Contributions of the other authors:

All co-authors participated in the development of study design, evaluation of experimental results and the setup and review of the manuscript.

Declaration on ethics: In accordance with the 3R principle for the reduction of animal experiments, all tissues used in this study were obtained from the veterinary diagnostic pathology tissue archive of the Department of Veterinary Pathology, Freie Universität Berlin, Germany and originated either from the veterinary clinical diagnostic service unit or previous experimental studies. Previous experimental studies were approved by the local authorities

and published in peer reviewed journals with international reputation and high ethical standards.

The publisher´s version can be openly accessed under the following link:

https://doi.org/10.7717/peerj.14202

Genomic, biochemical and expressional properties reveal strong conservation of the *CLCA2* gene in birds and mammals

Florian Bartenschlager[1], Nikolai Klymiuk[2,3], Achim D. Gruber[1] and Lars Mundhenk[1]

[1] Department of Veterinary Pathology, Faculty of Veterinary Medicine, Freie Universität Berlin, Berlin, Germany
[2] Large Animal Models in Cardiovascular Research, Internal Medical Department I, Technische Universität München, Munich, Germany
[3] Center for Innovative Medical Models, Ludwig-Maximilians-Universität München, Munich, Germany

Submitted 1 August 2022
Accepted 19 September 2022
Published 8 November 2022

Corresponding author
Florian Bartenschlager,
florian.bartenschlager@fu-berlin.de

Academic editor
David Ferrier

Additional Information and Declarations can be found on page 19

DOI **10.7717/peerj.14202**

Copyright
2022 Bartenschlager et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

ABSTRACT

Recent studies have revealed the dynamic and complex evolution of *CLCA1* gene homologues in and between mammals and birds with a particularly high diversity in mammals. In contrast, *CLCA2* has only been found as a single copy gene in mammals, to date. Furthermore, *CLCA2* has only been investigated in few mammalian species but not in birds. Here, we established core genomic, protein biochemical and expressional properties of *CLCA2* in several bird species and compared them with mammalian *CLCA2*. Chicken, turkey, quail and ostrich *CLCA2* were compared to their mammalian orthologues using *in silico*, biochemical and expressional analyses. *CLCA2* was found highly conserved not only at the level of genomic and exon architecture but also in terms of the canonical CLCA2 protein domain organization. The putatively prototypical galline *CLCA2 (gCLCA2)* was cloned and immunoblotting as well as immunofluorescence analyses of heterologously expressed *gCLCA2* revealed protein cleavage, glycosylation patterns and anchoring in the plasma membrane similar to those of most mammalian CLCA2 orthologues. Immunohistochemistry found highly conserved CLCA2 expression in epidermal keratinocytes in all birds and mammals investigated. Our results suggest a highly conserved and likely evolutionarily indispensable role of CLCA2 in keratinocyte function. Its high degree of conservation on the genomic, biochemical and expressional levels stands in contrast to the dynamic structural complexities and proposed functional diversifications between mammalian and avian *CLCA1* homologues, insinuating a significant degree of negative selection of *CLCA2* orthologues among birds and mammals. Finally, and again in contrast to *CLCA1*, the high conservation of *CLCA2* makes it a strong candidate for studying basic properties of the functionally still widely unresolved *CLCA* gene family.

Subjects Biochemistry, Evolutionary Studies, Genetics, Genomics, Molecular Biology
Keywords Evolution , CLCA, Avian, Mammal, Chicken, Ostrich, Turkey, Quail, Keratinocyte, Skin

INTRODUCTION

Chloride channel regulators, calcium activated (*CLCA*) constitute a family of genes that has been correlated to various disease conditions, including chronic inflammatory airway

How to cite this article Bartenschlager F, Klymiuk N, Gruber AD, Mundhenk L. 2022. Genomic, biochemical and expressional properties reveal strong conservation of the *CLCA2* gene in birds and mammals. *PeerJ* **10**:e14202 http://doi.org/10.7717/peerj.14202

diseases (*Hoshino et al., 2002*; *Nakanishi et al., 2001*; *Patel, Brett & Holtzman, 2009*; *Range, Mundhenk & Gruber, 2007*), cystic fibrosis (*Hauber et al., 2003*; *Ritzka et al., 2004*; *Young et al., 2007*) and cancer (*Chen et al., 2019*; *Hou et al., 2017*; *Walia et al., 2009*; *Walia et al., 2012*; *Yu, Walia & Elble, 2013*) and shows striking evolutionary dynamics (*Bartenschlager et al., 2022*; *Mundhenk et al., 2018*). In general, CLCAs comprise a prototypical protein domain architecture of a CLCA N-terminal (N-CLCA), a von Willebrand factor type A (vWA), a beta sheet rich (bsr) as well as a carboxy (C)-terminal fibronectin type III domain (fn3) that is separated from N-CLCA, vWA and bsr by proteolytical cleavage (*Patel, Brett & Holtzman, 2009*). Some CLCAs contain a transmembrane (TM) domain that anchors the C-terminal cleavage product in the plasma membrane (*Bartenschlager et al., 2022*; *Braun et al., 2010*; *Elble et al., 2006*; *Patel, Brett & Holtzman, 2009*; *Plog et al., 2012a*). CLCAs lacking the TM appear completely secreted (*Gibson et al., 2005*; *Mundhenk et al., 2006*; *Patel, Brett & Holtzman, 2009*; *Plog et al., 2009*).

On the genomic level, all *CLCA* genes in a given species are located within in a single locus, which is consistently flanked by the *Outer dense fibre of sperm tails 2-like* (*ODF2L*) and *SH3-domain GRB2-like endophilin B1* (*SH3GLB1*) genes (*Bartenschlager et al., 2022*; *Mundhenk et al., 2018*; *Plog et al., 2009*). Although *CLCA* genes have been identified in numerous avian, reptile, amphibian or fish species in the recent past, the main research has been focused on *CLCA* homologues in mammalian species so far (*Cunningham et al., 2021*). From the phylogenetic perspective, mammalian *CLCA* genes have been assigned to four clearly distinct clusters (*Mundhenk et al., 2018*; *Plog et al., 2009*). While clusters 3 and 4 exhibit a complex arrangement of multiple genes that seemingly arose from several independent duplication and inactivation events during the emergence of mammalian species (*Bartenschlager et al., 2022*; *Mundhenk et al., 2018*), both *CLCA* clusters 1 and 2 comprise only one intact single gene in each mammalian species. Recently, we described a *CLCA* locus in the genome of chicken (*Gallus gallus*) which is similarly flanked by the galline *ODF2L* and the *SH3GLB1* genes (*Bartenschlager et al., 2022*; *Mundhenk et al., 2018*). In contrast to the four mammalian *CLCA* (ma*CLCA)* clusters, however, chickens possess only two *CLCA* homologous genes, the galline CLCA1 (*gCLCA1*) and *gCLCA2*. From an evolutionary perspective, *gCLCA1* groups with mammalian clusters *CLCA1/3/4* and might therefore be the avian representative of the monophyletic ancestor of this group (*Bartenschlager et al., 2022*). In contrast to the genetically and functionally divergent *CLCA1/3/4* cluster, g*CLCA2* seems to be much more conserved with *gCLCA2* having a close genetic relationship to ma*CLCA2* (*Bartenschlager et al., 2022*; *Mundhenk et al., 2018*).

Concordantly with the high degree of conservation, maCLCA2 is consistently expressed in keratinocytes of stratified epithelia such as those of skin or esophagus and in certain glandular epithelia in all mammals investigated so far (*Braun et al., 2010*; *Connon et al., 2006*; *Connon et al., 2004*; *Erickson, Gruber & Mundhenk, 2020*; *Hamalainen et al., 2021*; *Plog et al., 2012b*; *Seltmann et al., 2018*; *Walia et al., 2012*). It has been proposed that the CLCA2 protein is involved in epithelial differentiation, growth arrest and maturation of keratinocytes (*Connon et al., 2004*; *Koegel & Alzheimer, 2001*; *Ramena et al., 2016*). Some species-specific differences regarding the expression, however, have been revealed in the mammalian respiratory tract, where human, porcine, and murine, but no feline CLCA2

has been detected in airway submucosal glands (*Dietert et al., 2015*; *Erickson et al., 2018*). Further, additional protein expression has been found in a specific subset of respiratory epithelial cells of the bronchial bifurcation in mice (*Dietert et al., 2015*).

Aiming at clarifying the relevance of non-mammalian *CLCA2*, we characterized the genomic organization and tissue und cellular expression patterns of *CLCA2* in four bird species, including chicken (*Gallus gallus*), turkey (*Meleagris gallopavo*), quail (*Coturnix species (sp.)*), and ostrich (*Struthio camelus*). The genomic organization as well as the protein domain architecture are described for the galline, quail and ostrich orthologues. We further describe the biochemical properties of CLCA2 in chicken as a presumed avian prototype. By comparison with its porcine, feline and murine CLCA2 homologues, conserved architectural elements, specific traits of biochemical processing and tissue expression patterns were identified, which will serve as the basis for functional investigations and structure-function-correlations in the future.

MATERIAL AND METHODS

In silico sequence analysis of gCLCA2 and generation of antibodies

Detailed gene positions, sizes, gene and amino acid (aa) sequences from chicken, quail, ostrich, pig, cat, and mouse *CLCA* loci were extracted from the NCBI (https://www.ncbi.nlm.nih.gov/) and Ensembl (http://www.ensembl.org/index.html) databases as described by *Plog et al. (2009)*. NCBI or Ensembl identifiers for CLCA2 sequences used are listed in Fig. S1. Exon-intron boundaries were established using WebScipio (*Hatje, Hammesfahr & Kollmar, 2013*) and aligned by GenePainter (*Hammesfahr et al., 2013*). Predicted protein domains were identified by the NCBI Conserved Domain Database (*Lu et al., 2020*), EMBL-EBI HMMER web server (*Potter et al., 2018*), Phobius webserver (*Käll, Krogh & Sonnhammer, 2007*), SOSUI (*Hirokawa, Boon-Chieng & Mitaku, 1998*), SignalP 3.0 (*Bendtsen et al., 2004*) algorithms and manual alignments. Asparagine (N)-linked glycosylation sites were predicted using the NetNGlyc webserver 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/). Turkey CLCA2 was not incorporated in the *in silico* analysis due to the low quality of the full-length gene and aa sequences stored in the NCBI and Ensemble databases (XP_031410715.1, XM_031554855.1, ENSMGAT00000009704.2). Phylogenetic relationship based on protein sequences of galline, quail, ostrich, feline, porcine and murine CLCA2 sequences was inferred by using the Maximum Likelihood method and JTT matrix-based model conducted in the MEGA X software package with 100 bootstrap replicates (*Tamura, Stecher & Kumar, 2021*) (S6).

Anti-gCLCA2 antibodies were generated similar to anti-porcine CLCA1 antibodies (*Plog et al., 2009*). In brief, an oligopeptide corresponding to aa 875 to 888 (WTAPGDDFD-KGQAA) in the C-terminal region of gCLCA2 was synthesized and conjugated with *Limulus polyphemus* hemocyanin (LPH). The LPH- conjugated peptide was used for immunization of two rabbits. Specific IgG-antibodies were isolated from the antisera using a cyanogen bromide immunization-peptide coupled sepharose column and named gC2.

Animals and tissues

In accordance with the 3R principle for the reduction of animal experiments, all tissues used in this study were obtained from the veterinary diagnostic pathology tissue archive of the Department of Veterinary Pathology, Freie Universität Berlin, Germany (VetPathFU) and originated either from the veterinary clinical diagnostic service unit or previous experimental studies. No animal was bred, raised, kept or euthanized specifically for this study. For chickens, 45 freshly frozen or formalin fixed and paraffin embedded (FFPE) tissues (Table S2) from ten-week old female individuals (*Gallus gallus domesticus*, Hampshire x White Leghorn, $n = 3$) and the gonads of age-matched male chickens (Hampshire x White Leghorn, $n = 3$) were used from (*Bartenschlager et al., 2022*). In brief, the tissues were by-products from slaughtered animals intended for human consumption. The animals had been bred, housed, and slaughtered in the Albrecht Daniel Thaer-Institute of Agricultural and Horticultural Sciences of the Humboldt-Universität zu Berlin, Germany under the permission of the State Office of Health and Social Affairs (approval number IC 114-ZH70). Weight at harvest was 1–1.2 kg (females) and 1.3–1.5 kg (males). The animals were raised in groups of 25, with infrared heat lamps offered until week five. They were fed with fledgling rearing feed until week eight and young hen feed afterwards. Miscanthus litter was used for housing enrichment. Harvesting was conducted according to national guidelines, which includes anesthesia by head blow and rapid exsanguination *via* jugular veins and carotid arteries. Females were harvested in the morning and males in the morning of the following day. FFPE tissues for immunohistochemical analyses, including esophagus and skin from shanks, abdomen and foot from one female and two male ostriches (*Struthio camelus*), three male quails (*Coturnix sp.*), and skin from three turkeys (*Meleagris gallopavo*) and skin of cats (*Felis catus*) were provided by the veterinary clinical postmortem diagnostic service unit of VetPathFU with no association to animal experiments. The tissues were obtained from the routine diagnostic spectrum to determine the cause of death of animals kept by private owners and free of histopathological changes. The permission to further use these tissues for research purposes was given by signature on the necropsy submission form by the owners. Additionally, FFPE skin samples from each of three mice and pigs from previous experimental studies (*Braun et al., 2010*; *Plog et al., 2012a* approval numbers T 0104/06 and G 0323/06, respectively) were obtained from the archive of VetPathFU and used for immunohistochemical analyses. In brief, 10-weeks-old female C57BL/6J mice were kept in cages enriched with nesting material. All animals had unlimited access to standard pelleted food and tap water. The room temperature was at 22 ± 2 °C and the relative humidity at 45–65%. A 12-h light/dark cycle was maintained. For experimental procedures, all mice were sacrificed by cervical dislocation in accordance with the national guidelines. Furthermore, the piglets (Euroc × Piétrain) were 18 days old, male, castrated and kept for four weeks in flatdeck compartments in groups of six piglets enriched with playthings. All animals had unlimited access to mash food and tap water. The room temperature was 28 °C at stabling and was gradually decreased to 22 °C within 10 days with air humidity at approx. 65%. The light programme consisted of a 16 h light and 8 h dark phase. With 45 days of age, all piglets were anesthezized with ketamine hydrochloride (Ursotamin®, 10%; Serumwerk Bernburg AG, Germany) and azaperone

(Stresnil®, Jansen-Cilag, Neuss, Germany) and euthanized using tetracaine hydrochloride, mebezonium iodide and embutramide (T61®, Intervet, Germany). All efforts were made to minimize animal discomfort and suffering.

Molecular cloning and sequencing of *gCLCA2*

The *gCLCA2* open reading frame (ORF) was cloned as described with minor modifications (*Bartenschlager et al., 2022*). In brief, the *gCLCA2* ORF was amplified from a batch of tissues including pharynx, crop, proctodeum and footpad from animal #2 (Table S2). The gCLCA2 ORF was tagged with the enhanced yellow fluorescent protein (*EYFP*) at the C-terminus by cloning it into the p*EYFP*-N1 vector (Clontech, Mountain View, California, USA). The resulting plasmid (*gCLCA2#2*) was sequenced using the primer walking method (Data S5). Three plasmids from independent experiments yielded identical results.

RT-qPCR tissue localization of *gCLCA2* mRNA

mRNA expression was analyzed using RT-qPCR as described (*Bartenschlager et al., 2022*). In brief, total RNA was isolated from galline tissues (Table S2), reverse transcribed, and the cDNA diluted to a final concentration of 1 ng/μl. Specific exon 13/14-boundaries spanning primers (upstream: 5′-CCAGGCTAACAGGACTACC-3′; downstream: 5′-GAAACCTCCTCTTCTGACCTGAAC-3′) were used to detect *gCLCA2* or the reference gene phosphoglycerate kinase (*PGK1*, upstream: 5′-AAAGTTCAGGATAAGATCCAGCTG-3′; downstream 5′-GCCATCAGGTCCTTGACAAT-3′; *Olias et al., 2014*) using a SYBR green qPCR assay (Thermo Fisher Scientific, Waltham, MA, USA). The *gCLCA2*-PCR product corresponds to the gCLCA2 protein region from aa760 to aa809 (QANRTTVPQTAMPWSHAMYIPGYVENGKLKMNPSRPPAI-ENNVQVRRGGF). *gCLCA2* mRNA was considered to be expressed when C_t-values of 35 or less were detected in at least two out of three tested animals.

Transient transfection of HEK293 cells

HEK293 cells (ATCC, Manassas, Virginia, USA) were transiently transfected as described with minor modifications (*Bartenschlager et al., 2022*). In brief, cells were grown in six-well plates in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% heat-inactivated fetal calf serum (FCS), 1% HEPES, and 1% penicillin/streptomycin. When reaching 80–90% confluence, the cells were transfected with 2 μg of a plasmid containing gCLCA2#2 or EYFP alone (mock) using 8 μl polyethylenimine (PEI) per well. 12 h post transfection, the cells were washed with phosphate buffered saline (PBS) and serum-free DMEM was added. 48 h after transfection, the cells of each well were lyzed using 500 μl radioimmunoprecipitation assay (RIPA) buffer supplemented with a protease inhibitor cocktail (complete Mini, EDTA-free, Roche Diagnostics, Rotkreuz, Switzerland). The protein concentration of supernatants and cell lysates were quantified using the bicinchoninic acid (Thermo Fisher Scientific, Waltham, Massachusetts, USA) method prior to freezing at −20 °C.

Endoglycosidase treatment

For glycosylation analysis, lysates from *gCLCA2*-transfected cells were deglycosylated by incubation with 25 U/ml endo H, 50 U/ml PNGase F or left untreated at 37 °C over night

according to the manufacturer's protocols (New England Biolabs, Ipswich, Massachusetts, USA).

Immunoblotting

Cell lysates and supernatants of *gCLCA2* transfected cells were analyzed using immunoblotting as described with minor modifications (*Bartenschlager et al., 2022*). In brief, samples of cell lysates or concentrated cell culture supernatant were reduced in 1,4-dithiothreitol (DTT) and separated using a 10% SDS-polyacrylamide gel electrophoresis. Proteins were transferred to a (PVDF)-membrane (https://www.linguee.de/englisch-deutsch/uebersetzung/polyvinylidene+fluoride.html) and blocked with 5% non-fat milk. Membranes were probed with antibody gC2 in a three-fold dilution series from 5 μg/ml to 0.05 μg/ml, or mouse monoclonal anti-YFP (cat. G163; ABM, Vancouver, Canada) diluted at 1:500, or mouse monoclonal anti-beta-actin (A5441, Sigma-Aldrich, St. Louis, Missouri, USA) diluted at 1:1,000. Membranes were incubated with horseradish peroxidase-conjugated goat anti-rabbit (115-035-068, Jackson Immuno Research Laboratories, Inc., West Grove, Pennsylvania, USA) or goat anti-mouse (111-035-144, Jackson Immuno Research Laboratories, Inc.) secondary antibodies and developed using enhanced chemiluminescence (Supersignal West Pico Plus, Thermo Fisher Scientific, Waltham, MA, USA). The gCLCA2 protein was only detected by this technique when using the anti-YFP antibody; however, it was undetectable when the gC2 antibodies were used.

Immunocytochemistry of transfected HEK293 cells

Immunocytochemistry was performed as described with minor modifications (*Bartenschlager et al., 2022*). In brief, HEK293 (ATCC) cells were grown on 8-well tissue chamber slides and transfected with *gCLCA2#2* or *EYFP*- mock plasmids. 48 h after transfection, the cells were briefly fixed in ice-cold methanol followed by a 4% paraformaldehyde fixation for 10 min. After permeabilization with 0.1% Triton X-100 in PBS and blocking with 10% goat normal serum (GS) and 0.05% Tween 20 in PBS, cells were probed with untreated or pre-absorbed antibody gC2 each used at 2 μg/ml or irrelevant affinity-purified rabbit polyclonal anti-porcine CFTR antibody (*Plog et al., 2010*) (S3). Alexa fluor 568 conjugated goat anti-rabbit (AB_143157, Invitrogen, Carlsbad, California, USA) were used as secondary antibodies followed by 4′, 6-diamidino-2-phenylindole (DAPI) nuclear counterstain. All *in vitro* experiments were repeated three times.

Tissue and cellular localization of gCLCA2 protein using immunohistochemistry and immunofluorescence

All galline tissues in which g*CLCA2* mRNA was detected at C_t-values below 35 were analyzed *via* immunofluorescence to identify gCLCA2 expressing cell types. Furthermore, skin and esophagus from chicken, ostriches, quails, as well as skin of turkeys, mice, pigs, and cats were analyzed *via* immunohistochemistry. Immunofluorescence and immunohistochemistry were performed as described with minor modifications (*Bartenschlager et al., 2022*). In brief, FFPE-tissues were cut, mounted on adhesive glass slides, and dewaxed. For immunohistochemistry, endogenous peroxidase was blocked by adding 0.5% H_2O_2 in methanol. For immunofluorescence analysis, tissue sections of chickens were permeabilized

with 0.1% Triton X-100 in PBS. Antigen was retrieved using 1 mg/ml recombinant protease from *Streptomyces griseus*. Slides were blocked with 10% Roti-ImmunoBlock and 20% GS in PBS for immunohistochemistry and 10% GS and 0.05% Tween 20 in PBS for immunofluorescence, both for 30 min. The slides were probed with the immunopurified gC2 or irrelevant affinity-purified rabbit polyclonal (anti-porcine CFTR, *Plog et al., 2010*) antibodies at 2 µg/ml. Additionally, mouse monoclonal anti-cytokeratin (AE1/AE3, M3515, Agilent Dako, Santa Clara, California, USA) antibodies were used at 1:400. For immunofluorescence, Alexa fluor 568-conjugated goat anti-rabbit (AB_143157; Invitrogen, Waltham, MA, USA) secondary antibodies were used diluted at 1:200, followed by DAPI nuclear counterstain. For immunohistochemistry with AE1/AE3 primary antibodies, 3,3′-diaminobenzidine (DAB) was added after incubation with goat anti-mouse biotinylated secondary antibodies (BA-9200; Vector Laboratories, Burlingame, California, USA) diluted at 1: 200 and an avidin-biotin complex. For immunohistochemistry with the gC2 primary antibody, DAB was added after incubation with the Histofine Simple Stain Mouse MAX PO anti-rabbit polymer kit (414341F; Nichirei Biosciences Inc., Tokyo, Japan). Potential cross reactivity of the gC2 antibody with pig, cat, mouse, turkey, quail and ostrich CLCA2 orthologues was tested by epitope sequence alignment using the NCBI Protein Blast (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins, Table S4).

RESULTS

Avian and mammalian *CLCA2* genes and their overall protein structures are conserved

Similar to mammals (*Bartenschlager et al., 2022*; *Mundhenk et al., 2018*), avian *CLCA2* (a*CLCA2*) are single-copy genes located directly adjacent to the *ODF2L* gene in all species analyzed here (Fig. 1A). While in mammals the region between *CLCA2* and *SH3GLB1* comprise the complex and divergent *CLCA1/3/4* locus, birds contain only one single *CLCA1* gene (Fig. 1A). The presence of only two *CLCA* homologues but also the shorter intronic and intergenic regions that correspond to the more compact organization of avian genomes (*Ellegren, 2005*) make the consensus avian *CLCA* gene locus much smaller than that in mammals (Fig. 1A, *Bartenschlager et al., 2022*). The a*CLCA2* genes comprise 14 exons that encode for a single, putatively functional protein in all bird species analyzed here (Figs. 1B, 2), identical to all mammals investigated to date (*Patel, Brett & Holtzman, 2009*). The nucleotide numbers vary in only four exons by three to nine nucleotides between avian and mammalian *CLCA2* genes (Fig. 1B), causing only slightly distinct protein lengths (Fig. 2). Overall, our findings establish a high degree of evolutionary conservation of the genomic *CLCA2* architecture across birds and mammals (Fig. 1B). For all *CLCA2* orthologous genes investigated here, the predicted proteins share the canonical CLCA protein domain architecture as described (*Patel, Brett & Holtzman, 2009*). *In silico* prediction suggests a signal peptide within the first 22 to 43 aa, indicating entry into the secretory pathway which is a highly conserved trait in all avian and mammalian CLCA2 sequences (Fig. 2). An N-CLCA domain (PFAM identifier: pfam08434) containing a proteolytic HExxH-motif and a cysteine-rich domain is prepended to a vWA domain

(PFAM identifier: pfam13519) and a bsr domain (Fig. 2, S1, *Patel, Brett & Holtzman, 2009*). In contrast to other CLCA proteins but consistent with maCLCA2, the vWA domain of CLCA2 of birds does not contain an intact metal ion dependent adhesion site (MIDAS, DxSxS-T4-D5). In concordance to maCLCA2, an fn3 (PFAM identifier: PF00041) and a TM domain are predicted in the C-terminal cleavage product for aCLCA2 (Fig. 2, Fig. S1). The presence of designated beta4-integrin binding motif (IBM, consensus sequence F(S/N)R(I/L/V)(S/T)S, *Abdel-Ghany et al., 2003*) appears less consistent. Human and porcine CLCA2 show such an IBM within the vWA domain (human aa480-485: FSRISS (*Patel, Brett & Holtzman, 2009*), pig aa480-485: FSRISS, Fig. S1) while it is absent from chicken, quail, ostrich, feline and murine CLCA2 (Fig. S1). Another IBM motif is lacking in the C-terminal cleavage product of birds and pigs whereas such a motif was found in human, feline and murine CLCA2 (human aa480-485: FSRISS (*Patel, Brett & Holtzman, 2009*), cat aa741-746: FSRVSS, Fig. S1, mouse aa740-745: FSRVSS (*Patel, Brett & Holtzman, 2009*), Fig. S1).

Noteworthy, aCLCA2 amino acid (aa) sequences are 921 to 930 aa long and therefore shorter than all of their mammalian homologues with 942 to 944 aa (Fig. 2). In addition, aCLCA2 do not contain a predicted glycosylation site after the predicted transmembrane domain (Fig. 2).

A phylogeny based on galline, quail, ostrich, feline, porcine and murine CLCA2 protein sequences revealed a monophyletic aCLCA2 group separate from maCLCA2 (Fig. S6).

The gCLCA2 protein shares many biochemical properties with mammalian CLCA2 proteins

Posttranslational cleavage

The cleavage of a precursor protein into a larger N- and a shorter C-terminal subunit belongs to the conserved properties of all maCLCA proteins (*Patel, Brett & Holtzman, 2009*). It is thought to be mediated by the zinc-binding HExxH motif in the N-CLCA domain, cleaving the protein at a canonical cleavage site (*Bothe et al., 2012*; *Lenart et al., 2013*; *Pawłowski et al., 2006*; *Yurtsever et al., 2012*). Consistently, canonical HExxH motifs and putative proteolytic cleavage sites were identified in aCLCA2 proteins (Fig. 2). The predicted cleavage was verified *in vitro* for gCLCA2 as a putative avian prototype by immunoblot analysis of lysates from heterologously transfected HEK293 cells. A band consistent with a precursor protein at approx. 145 kilodalton (kDa) and a band consistent with the C-terminal cleavage product of approx. 57 kDa were detected (Fig. 3), suggesting that posttranslational cleavage of CLCA2 also occurs in chicken.

N-glycosylation and cleavage in the medial Golgi

maCLCA2 proteins are multiple N-linked glycosylated (*Braun et al., 2010*; *Elble et al., 2006*; *Gruber et al., 1999*; *Plog et al., 2012b*). Consistently, our *in silico* analyses predicted five N-linked glycosylation sites for aCLCA2 (Fig. 2). To verify this prediction experimentally, lysates from *gCLCA2*- transfected HEK293 cells were treated with endoglycosidases endo H and PNGase F and immunoblotted to identify the kind and extent of glycosylation. The approx. 145 kDa precursor protein was sensitive to endo H and PNGase F treatments (Fig. 4), resulting in a size shift from approx. 145 kDa to approx. 130 kDa, suggestive

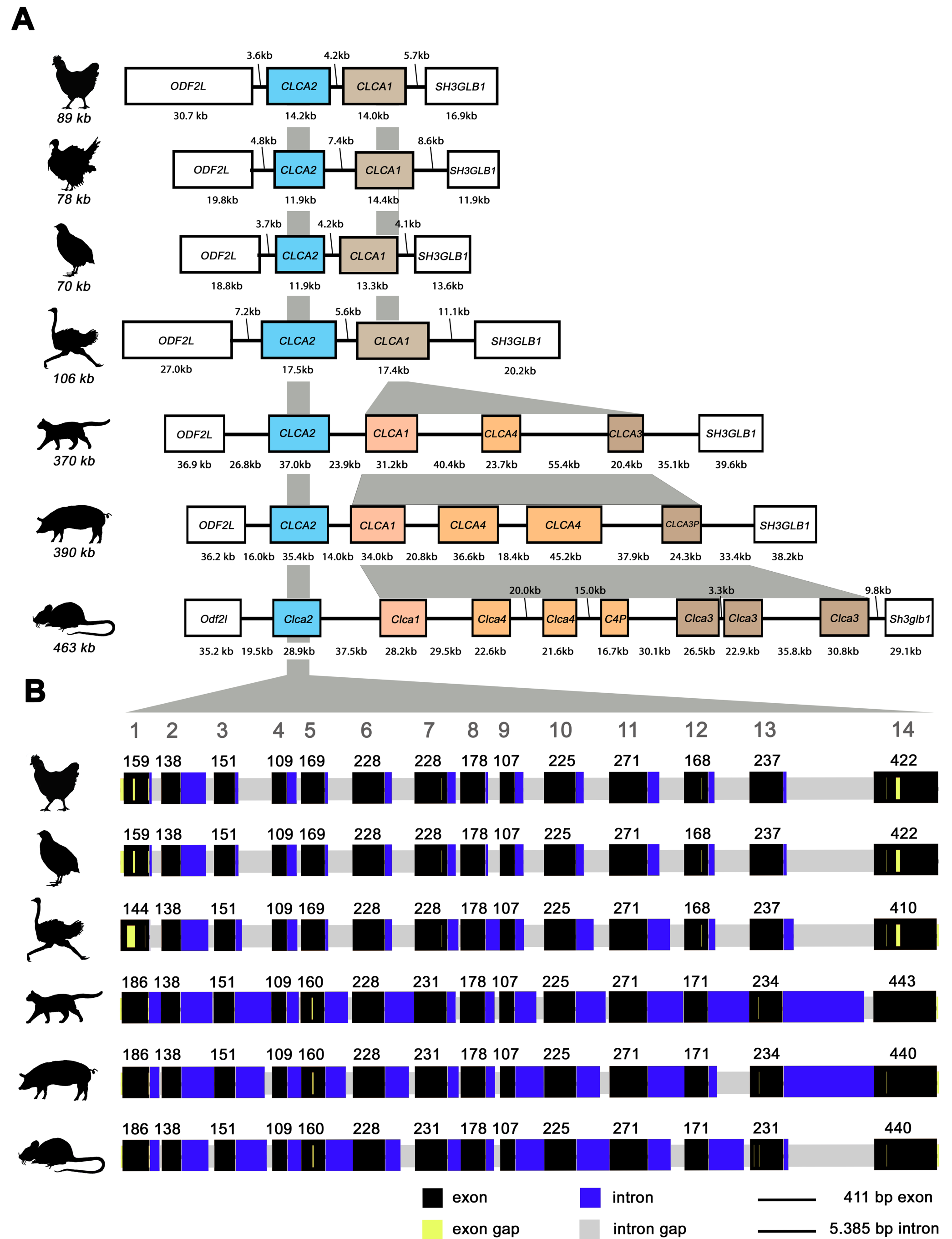


Figure 1 **The genomic architecture of avian and mammalian *CLCA2* genes is highly conserved while *CLCA1* homologues are diverse only in mammals but not in birds.** (A) Boxes indicate genes, black lines illustrate intergenic regions. Box colors highlight *CLCA* orthologues with white boxes indicating neighboring genes. Gray areas indicate regions of high sequence homology. The avian *CLCA* loci are scaled 2-fold larger for illustrative purposes. kb, kilobases. (continued on next page...)

Full-size DOI: 10.7717/peerj.14202/fig-1

Figure 1 (...continued)
(B) The exon architecture of avian and mammalian *CLCA2* is highly conserved. Identical to mammals, a*CLCA2* genes are encoded by 14 exons with concurring exon lengths in exons 2, 3, 4, 6, 8, 9, 10 and 11. Black boxes represent exons and blue boxes depict introns. Gaps in the alignment are highlighted in yellow for exons and in grey for introns. Large grey digits at the top indicate exon numbers, small black digits indicate base pairs per exon. bp: base pairs.

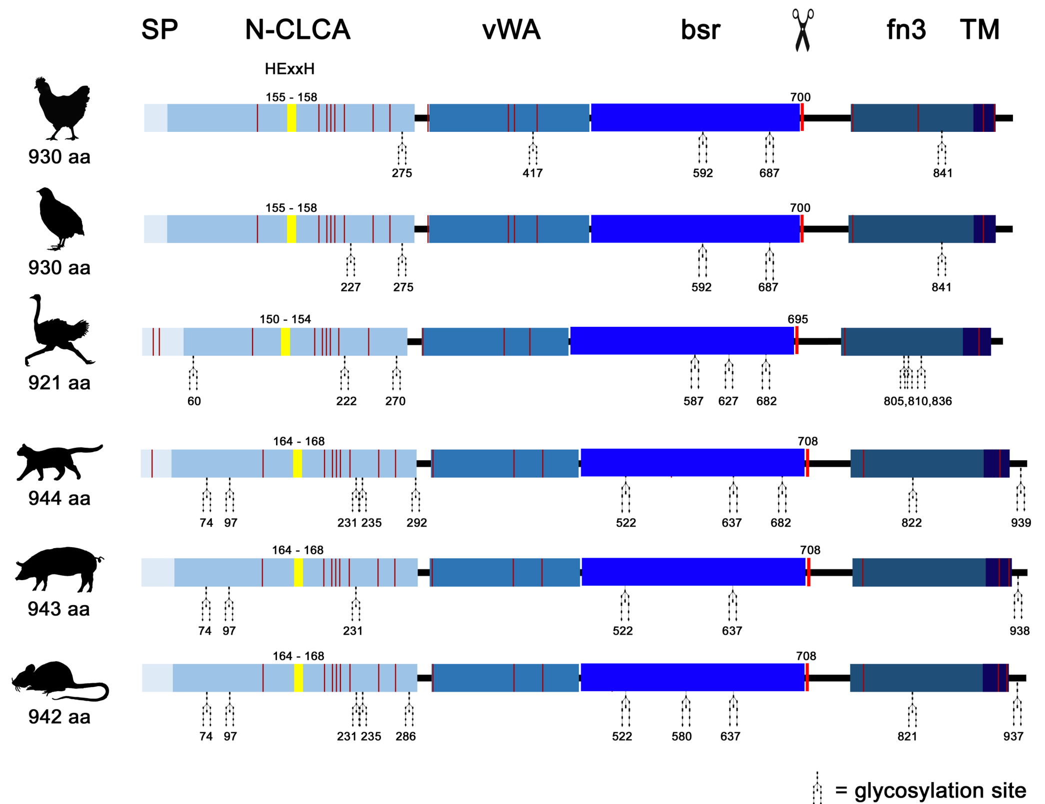


Figure 2 Avian and mammalian CLCA2 proteins share the canonical CLCA2 predicted protein architecture. Graphical alignment of CLCA2 orthologues. Protein domains are indicated by boxes in shades of blue and entitled by: SP, signal peptide; N-CLCA, N-CLCA domain; vWA, von Willebrand factor type A domain; bsr, beta sheet rich domain; fn3, fibronectin type III domain; TM, transmembrane domain. Yellow vertical boxes illustrate HExxH-motifs, red vertical boxes and the scissor sketch depict putative cleavage sites, vertical red lines constitute cysteines. Numbers indicate aa positions or lengths, respectively.
Full-size DOI: 10.7717/peerj.14202/fig-2

of an immature high mannose-type glycosylation pattern. In contrast, the C-terminal cleavage product was resistant to endo H treatment but sensitive to PNGase F treatment, as suggested by a reduction in size from approx. 57 kDa to approx. 54 kDa (Fig. 4). Our results indicate that virtually all predicted sites might be glycosylated, presuming a molecular weight of approx. 3 kDa per glycosylation (*Pult et al., 2011*). Furthermore, the complex high mannose-rich glycosylation pattern of the C-terminal cleavage products, in contrast to the immature glycosylated precursor protein, suggests cleavage of gCLCA2 early in the medial Golgi, similar to what has been observed for the murine CLCA2 (*Braun*

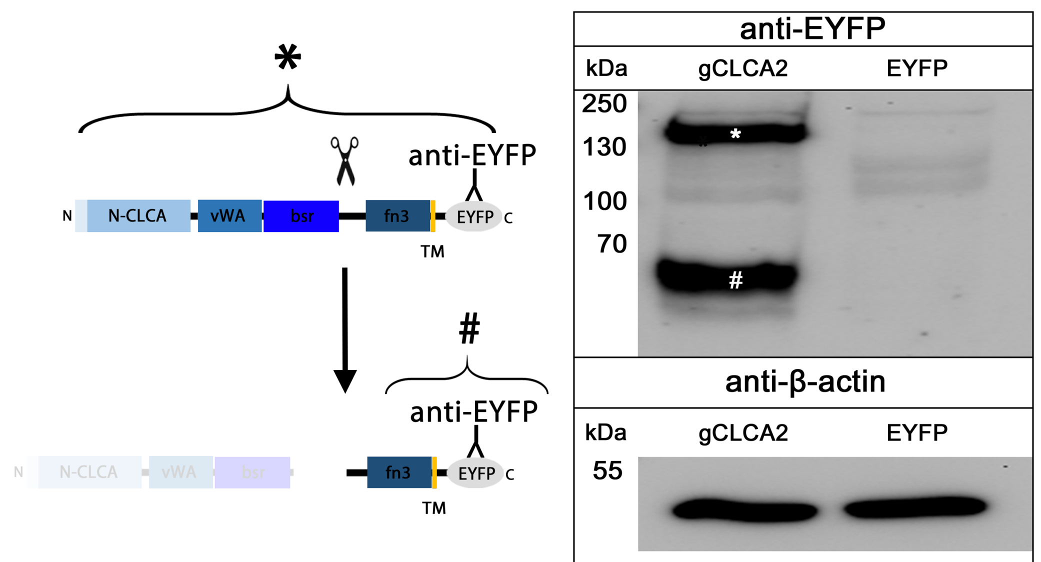


Figure 3 Posttranslational cleavage of the gCLCA2 precursor protein. Left panel: illustration of protein cleavage: N-CLCA: N-CLCA domain, vWA: von Willebrand factor type A domain, bsr, beta sheet rich domain; fn3, fibronectin type III domain; TM, transmembrane domain; EYFP, enhanced yellow fluorescent protein tag; scissor sketch, putative cleavage site; Y, antibody binding site. Right panel: immunoblot analysis of lysates from *gCLCA2* or *EYFP*-mock plasmid transfected HEK293 cells using an anti-EYFP antibody. An asterisk (*) indicates uncleaved precursor protein, a hash (#) indicates C-terminal cleavage product. To control for equal total protein loading, the samples were identically immunoblotted with primary anti-beta-actin antibodies (bottom panel). Representative image of three independent experiments are shown.
Full-size DOI: 10.7717/peerj.14202/fig-3

et al., 2010). Thus, glycosylation and cleavage in the medial Golgi also appear as conserved traits.

Anchoring in the plasma membrane

Identically to its maCLCA2 orthologues, a TM domain in the C-terminal subunit was predicted for aCLCA2, which anchors the protein to the plasma membrane (Fig. 2; *Braun et al., 2010*; *Elble et al., 2006*; *Gruber et al., 1999*; *Plog et al., 2012b*). Consistently, a prominent green autofluorescent EYFP signal was detected along the plasma membrane of HEK293 cells transfected with the *gCLCA2* construct that contained an EYFP tag downstream to the TM domain (Figs. 5A and Figs. 5B1). A coinciding signal (Fig. 5B3) was found by immunocytochemistry using the anti-gC2 antibody directed against the fn3 domain of the protein, which is located upstream to the C-terminal transmembrane domain (Fig. 5B2). This is in contrast to the diffuse green signal of cytosolic EYFP protein (Fig. 5C). Therefore, anchorage of the CLCA2 protein in the cell membrane *via* a TM domain seems also conserved with mammals.

aCLCA2 is expressed in stratified squamous epithelia of skin and mucous membranes

The tissue and cellular expression patterns of gCLCA2 were examined at the mRNA- and protein-levels. *gCLCA2* mRNA was detected by RT-qPCR in all tested locations of the skin

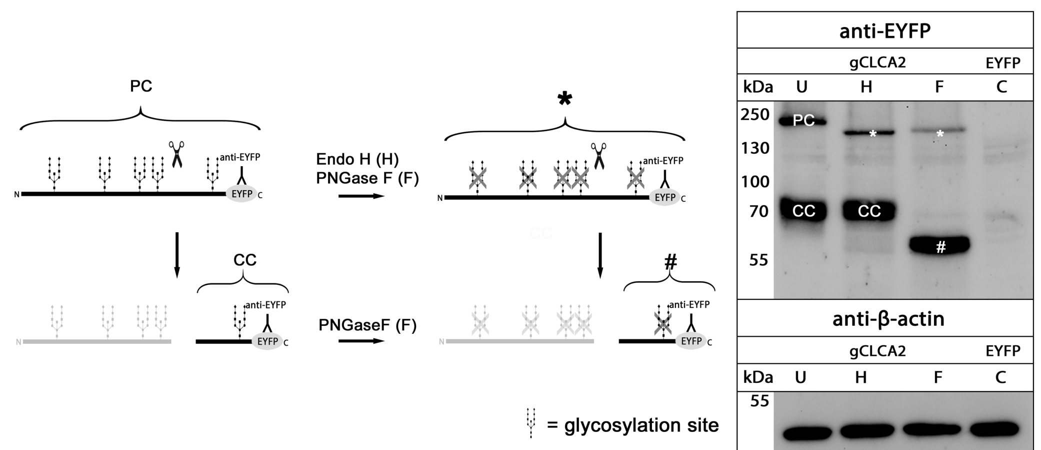


Figure 4 gCLCA2 is a multiple N-linked glycosylated protein and cleaved early in the medial golgi. Left panel: illustration of protein cleavage and deglycosylation: scissor sketch, cleavage; Y, antibody binding site; PC, precursor protein; CC, C-terminal cleavage product; asterisk, deglycosylated, uncleaved precursor protein; hash, deglycosylated C-terminal cleavage product. Right panel: immunoblot analysis of lysates from HEK293 cells transfected with *gCLCA2* or the *EYFP*-mock plasmids (C), using an anti-EYFP antibody. Lysates were treated with endoglycosidases endo H (H), PNGase F (F) or left untreated (U). To control for equal total protein loading, the samples were immunoblotted with anti-beta-actin antibodies (bottom panel). Representative image of three independent experiments.

Full-size DOI: 10.7717/peerj.14202/fig-4

(back, foot, wattle, ball of the foot, proctodeum) and skin appendages (feather follicle, beak, uropygial gland) as well as in organs with keratinizing mucosal membranes, such as the nose, pharynx, esophagus, crop, and proctodeum (Fig. 6, Table S2). Additionally, *gCLCA2* mRNA was found in the trachea, cecum, kidney, bursa of Fabricius, thyroid gland, sciatic nerve, eye and in the liver (Fig. 6, Table S2). gCLCA2 protein was exclusively detected immunohistochemically in keratinocytes of the skin, skin appendages and keratinizing mucosal membranes of the larynx, esophagus and crop (Fig. 7A, Fig. S3). gCLCA2 was localized in all layers of the stratified epithelium similar to the epithelial cell marker cytokeratins (Fig. 7B). At the subcellular level, signals consistently appeared as multiple, evenly distributed dots within the cytosol, with no specific signal enrichment detected at the plasma membrane. However, we failed to detect the CLCA2 protein in other tissues with notable *gCLCA2* mRNA presence (Fig. 6). Abundant expression in epidermal keratinocytes appears as a consistent hallmark of CLCA2 proteins, as verified in chicken, turkeys, quails, ostriches, cats, pigs and mice (Fig. 8).

DISCUSSION

Previous investigations on the *CLCA* gene family have revealed unusually complex evolutionary developments in some of its members (*Bartenschlager et al., 2022*; *Mundhenk et al., 2018*). The divergent cluster of mammalian *CLCA1/3/4* genes is characterized by multiple and independent duplication and inactivation events. This suggests flexible adaptation to environmental conditions and separation and specification of gene functions (*Mundhenk et al., 2018*; *Patel, Brett & Holtzman, 2009*; *Plog et al., 2015*). In sharp contrast,

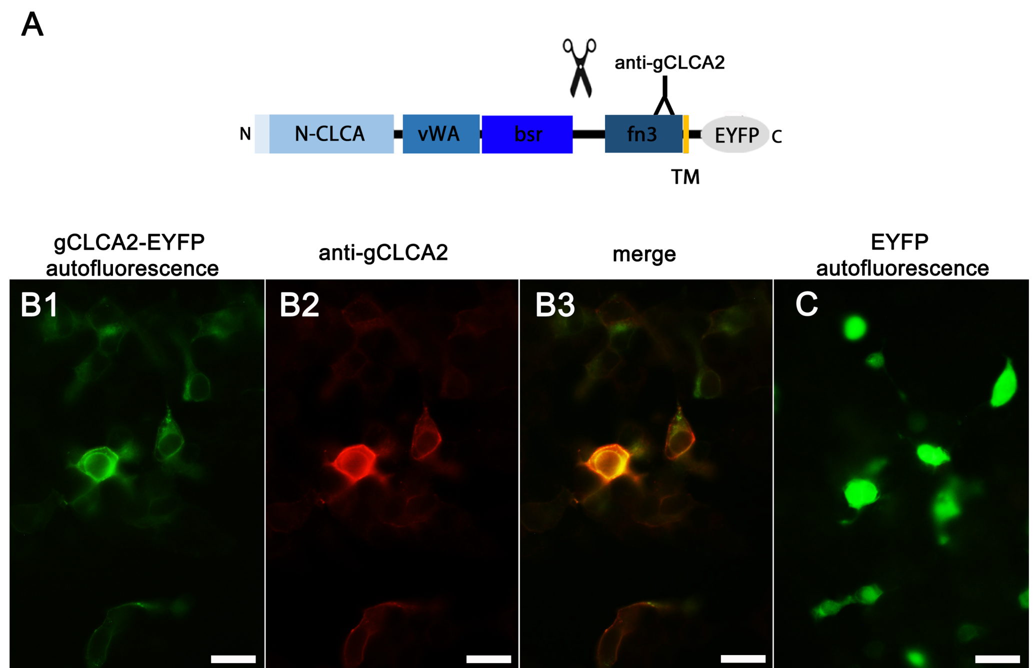


Figure 5 Heterologously expressed gCLCA2 is associated with the plasma membrane. (A) illustration of the anti-gCLCA2 binding site: N-CLCA, N-CLCA domain; vWA, von Willebrand factor type A domain; bsr, beta sheet rich domain; scissor sketch, putative cleavage site; fn3, fibronectin type III domain; TM, transmembrane domain; EYFP, enhanced yellow fluorescent protein tag; Y, antibody binding site. (B) auto- and immunofluorescent localization of the C-terminal cleavage product of gCLCA2 using antibody gC2 in HEK293 cells transfected with the *gCLCA2* plasmid. B1, green signal; EYFP, autofluorescence; B2, red signal anti-gCLCA2 immunofluorescence; B3, orange; merged signals of EYFP autofluorescence and anti-gCLCA2 immunofluorescence. Alexa fluor 568-conjugated secondary antibodies. (C) autofluorescence (green signals) of HEK293 cells transfected with the *EYFP* mock plasmid. Bars indicate 20 μm. Representative images of three independent experiments are shown.

Full-size DOI: 10.7717/peerj.14202/fig-5

the *CLCA2* gene appears as a consistent single intact gene in mammalian and avian species examined here (Fig. 1A). It thus appears that *CLCA2* is the most conserved gene of the family, providing an opportunity to gain a consistent and detailed insight into basic properties of *CLCA* genes. Here, our systematic comparisons of the *CLCA2* genomic and protein structures, biochemical properties and tissue as well as cellular expression levels confirmed the high degree of conservation within and between birds and mammals.

Similar to all previously studied *CLCA* genes, avian *CLCA2* (a*CLCA2)* genes are encoded by 14 exons with small in-frame deletions, compared to ma*CLCA2* (Fig. 1B), leading to slightly shorter protein sequences. Furthermore, aCLCA2 match the canonical protein architecture of CLCA that had been described for mammalian species (*Braun et al., 2010*; *Elble et al., 2006*; *Evans, Thoreson & Beck, 2004*; *Patel, Brett & Holtzman, 2009*; *Plog et al., 2012b*), namely the consecutive sequence of a signal peptide followed by the domains N-CLCA, vWA, bsr, fn3. These domains as well as the consistent proteolytic cleavage motif located between bsr and fn3 might constitute a not yet defined common functional property of intact CLCA gene family members. Obviously, however, the evolutionary turnover

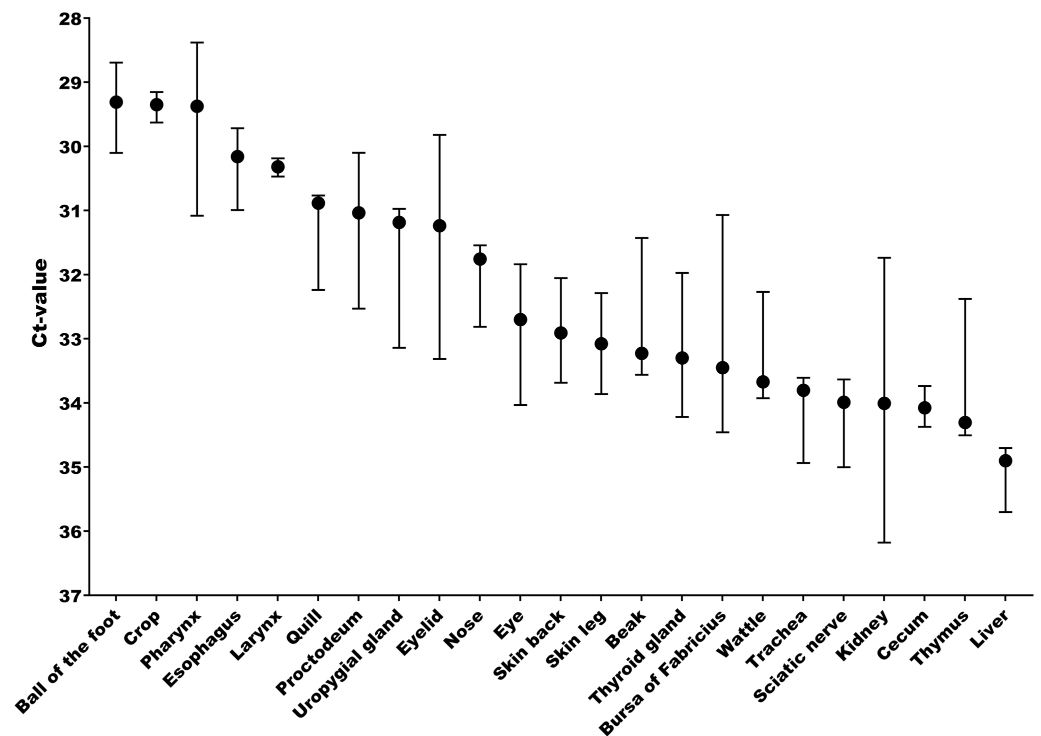


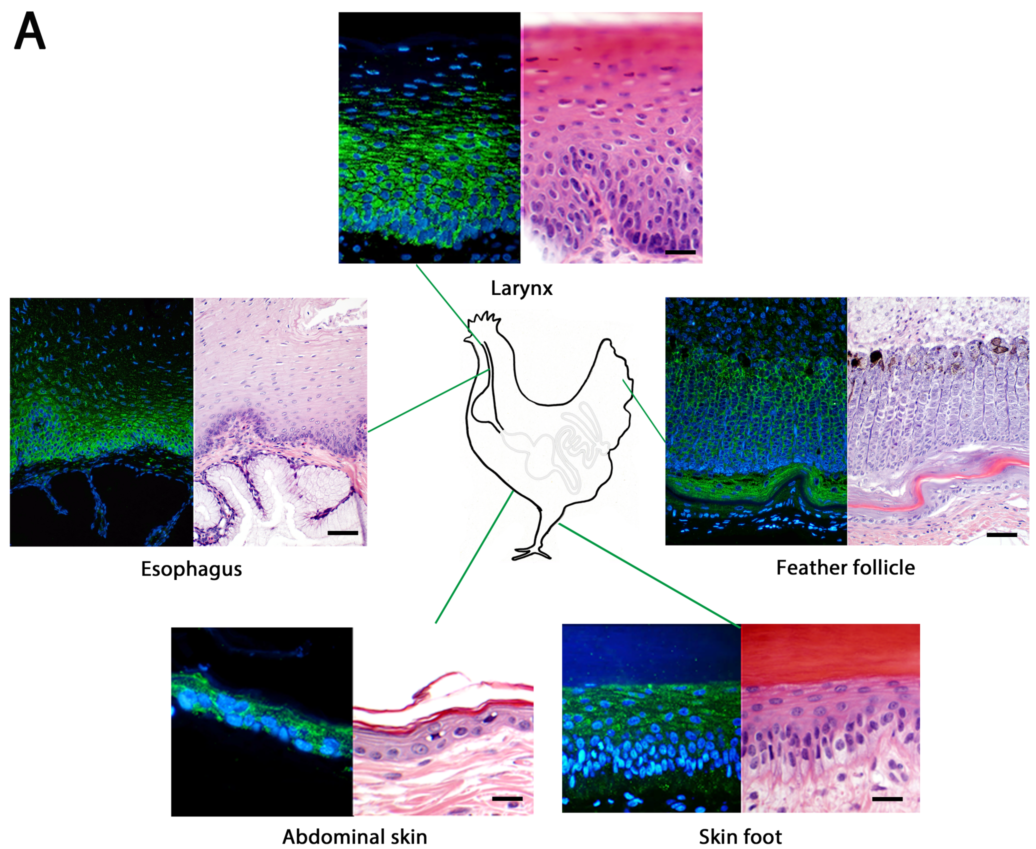
Figure 6 Abundant g*CLCA2* mRNA was detected by quantitative RT-PCR in tissues containing keratinizing epithelium. Black dots represent median Ct values, error bars indicate the range. $n = 3$ animals per group. Ct, cycle threshold.

Full-size DOI: 10.7717/peerj.14202/fig-6

has modified functional properties, as suggested by the lack of an otherwise consistent TM domain in maCLCA1 (*Bartenschlager et al., 2022*). Similarly, an intact MIDAS site that is present in the mammalian CLCA1/3/4 cluster and its avian CLCA1 orthologue (*Bartenschlager et al., 2022*) was found to be interrupted in CLCA2 proteins in mammals and birds. This conserved feature of CLCA2 might point to a similar sub-functionalization in mammals and birds.

Cleavage of gCLCA2 into a larger N-terminal part and a smaller membrane-anchored C-terminal tail is presumably caused by proteolytic cleavage between the bsr and fn3 domain in the medial Golgi. This is similar to what has been found in its murine orthologue (*Braun et al., 2010*) whereas the human orthologue is cleaved extracellularly after its insertion into the plasma membrane (*Elble et al., 2006*). Thus, cleavage in the medial Golgi appears to be the evolutionarily older, prototypical process, which might have been changed in certain lineages, including humans.

The CLCA2 protein was consistently found to be expressed in epidermal keratinocytes of various cutaneous and mucosal body coverings of chicken, turkey, quail and ostrich. This highly select cellular protein expression is similar to that of human, murine, feline, and porcine orthologues (*Braun et al., 2010*; *Connon et al., 2006*; *Connon et al., 2004*; *Erickson, Gruber & Mundhenk, 2020*; *Hamalainen et al., 2021*; *Plog et al., 2012b*; *Seltmann et al.,*



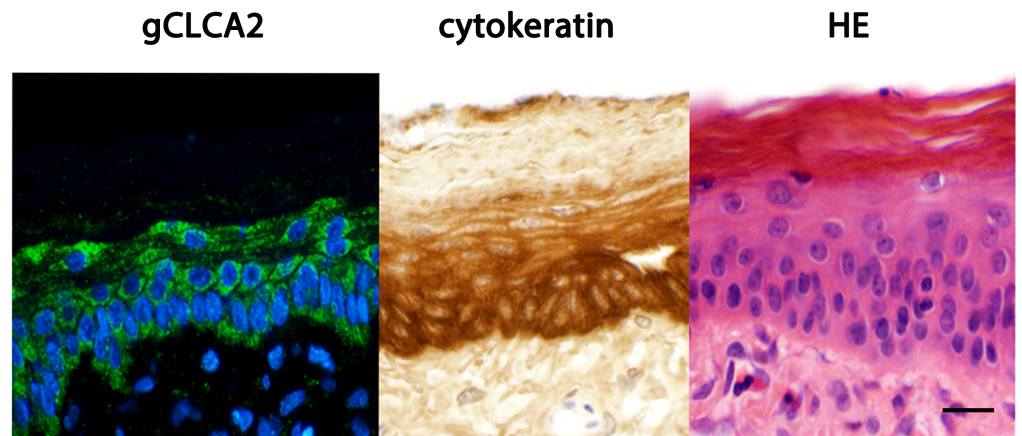


Figure 7 Immunofluorescence localization of the gCLCA2 protein in the cells of stratified/cornified epithelia of the skin, mucus membranes and feather follicles. (A) Immunofluorescence on galline tissues using the gC2 antibody (left panels), compared to hematoxylin and eosin (HE) stained tissue sections of FFPE-samples (right panels) for structural comparisons. Alexa fluor 488-conjugated secondary antibodies were used for visualization (green) and DAPI as nuclear (continued on next page…)

Full-size DOI: 10.7717/peerj.14202/fig-7

Figure 7 (...continued)
counterstain (blue). Bars: 10 μm. (B) the gCLCA2 protein is located in stratified epithelial cells characterized by the expression of cytokeratins. Left panel: Immunofluorescent localization of the gCLCA2 protein in stratified epithelium of wattle skin using antibody gC2 and Alexa fluor 488-conjugated secondary antibodies with DAPI nuclear counterstain (blue). Central panel: immunohistochemical localization of acidic and basic cytokeratins in a serial tissue section using the AE1/AE3 antibody cocktail, biotinylated anti-mouse IgG secondary antibodies conjugated with horseradish peroxidase and DAB as chromogen (brown) with hematoxylin counterstain (blue). Right panel: tissue morphology visualized using hematoxylin (blue) and eosin (pink) stain on a consecutive tissue section. Bars: 10 μm. Representative images of sections from three animals.

2018). Despite the obvious differences of avian and mammalian skin anatomy (*Akat et al., 2022*), CLCA2 expression in keratinizing epithelial cells seems to be highly conserved, supporting the notion that skin and skin appendages including hair in mammals and feathers in birds share a common ancestry (*Di-Poï & Milinkovitch, 2016*). The idea of the symplesiomorphic nature of CLCA2 protein abundance in hair (*Plog et al., 2012b*) and feather follicles might be extended to other sites as well. It had been reported that corneal epithelial cells of galline embryos express CLCA2 protein (*Connon et al., 2006*). Like in skin, hair and feathers, expression in corneal epithelial cells is plausible as they derive from a common ectodermal origin with keratinocytes. Accordingly, *gCLCA2* specific mRNA was found in the ten-week old chicken eye in our study, while no protein was detected using immunohistochemistry. It will be interesting to explore whether impaired sensitivity for protein detection was the reason for this discrepancy or if CLCA2 expression in corneal epithelial cells depends on the developmental status. Similarly, the detection of *gCLCA2* mRNA in several other tissues, albeit at lower mRNA expression levels, with concomitant lack of detectable CLCA2 protein is in line with similar findings on murine CLCA2 (*Braun et al., 2010*). On the one hand, the *gCLCA2* specific amplicon detected by RT-qPCR is slightly (129 nucleotides) upstream to the sequence that encodes the antibody-binding site for the anti-gCLCA2 antibody used in this study. Therefore, it cannot be fully excluded that these tissues, which apparently express *gCLCA2*-mRNA but lack expression of gCLCA2 protein, may express truncated or otherwise modified variants of *gCLCA2*. On the other hand, this discrepancy underscores established difficulties in exploring gene products that appear at low expression levels or are restricted to small niches of expression sites. Theoretically, non-translated transcripts of *CLCA2* might also have regulatory properties, but to our knowledge, no such mechanisms have ever been proposed for any member of the *CLCA* gene family. For this reason, we propose the most relevant role of CLCA2 in keratinocytes of skin, hair and feather follicles in mammals and birds, respectively. The highly conserved expression in the epidermis may suggest an indispensable role for CLCA2 in keratinocyte function during the course of evolution. It remains to be established, however, to which functional aspect this may pertain, including skin barrier function, osmolar homeostasis, cell signaling or local immunity.

Despite the fact that most of the assessed traits appear conserved between aCLCA2 and maCLCA2, two inconsistencies were found: first, while we failed to detect the gCLCA2 protein in the airways of chicken, it is abundantly expressed in epithelial cells of the respiratory tract of some albeit not all mammals (*Dietert et al., 2015*). Given the established

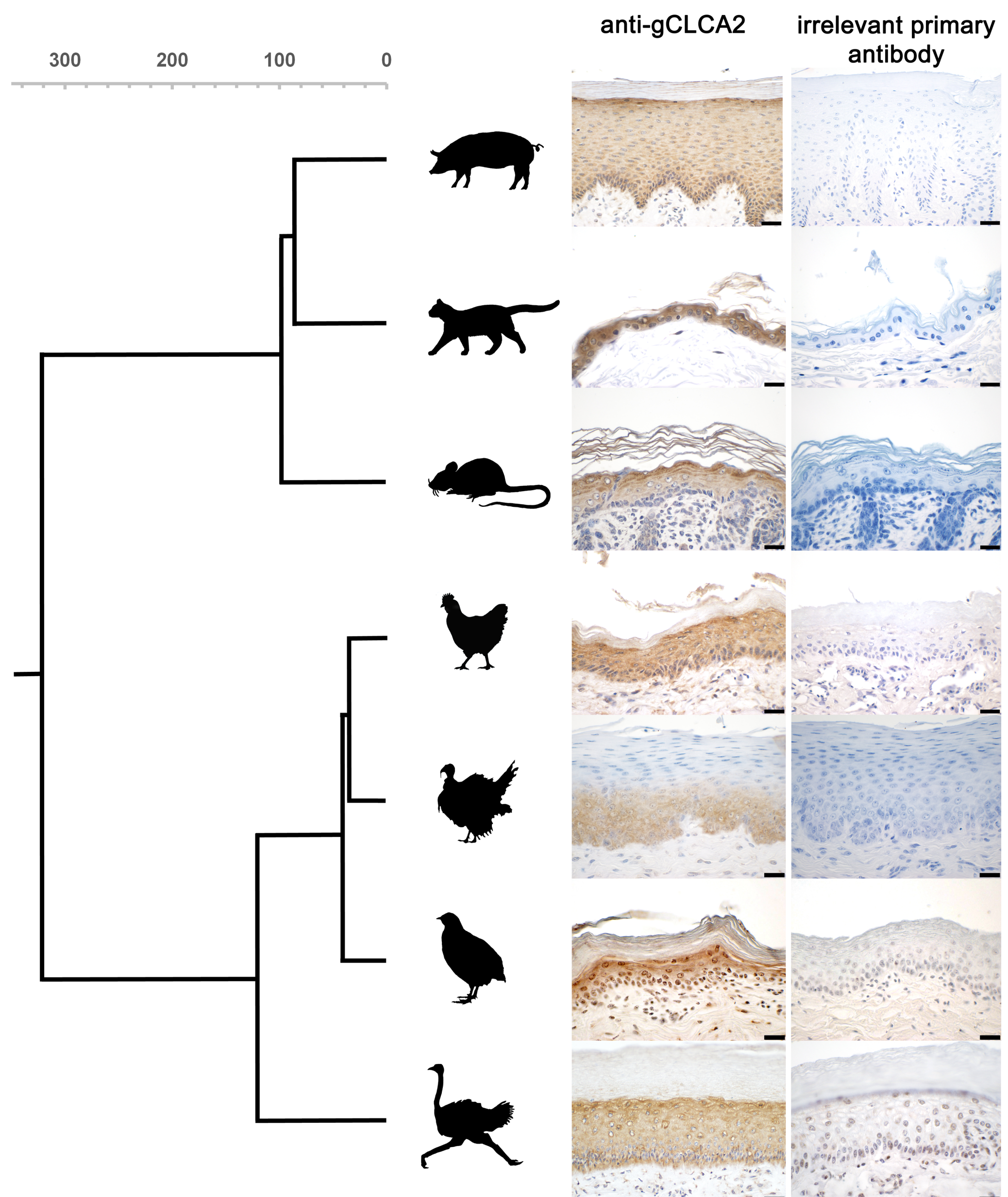


Figure 8 CLCA2 expression is conserved in viable epidermal keratinocytes across avian and mammalian species. Immunohistochemistry using the gC2 antibody and an irrelevant control antibody (anti-porcine CFTR). Biotinylated anti-mouse secondary antibodies and DAB as chromogen (brown) and hematoxylin counterstain (blue). Gray scalebar: species divergence in million years. Black bars: 20 µm. Representative images of sections from three animals per species.

Full-size DOI: 10.7717/peerj.14202/fig-8

evolutionary relationships, this may propose a potential adaptation of CLCA2 in the respiratory tract during mammalian evolution and its subsequent loss in cats or other carnivores. Theoretically, a more ancient function of CLCA2 in the respiratory tract and an independent loss in avian and select mammalian species may be conceived, possibly to be confirmed in a common ancestor of avian and mammals. Even more

complicated to interpret is the presence of IBM motifs in the vWA and fn3 domains of CLCA2 that are thought to mediate cell–cell adhesion via interaction between CLCA2 and beta4-integrin (*Abdel-Ghany et al., 2003*; *Abdel-Ghany et al., 2001*). However, the inconsistency of functional data on this motif and its relatively loose consensus sequence F(S/N)R(I/L/V)(S/T)S raise doubt on whether this motif has any physiological function in CLCA2.

To date, knowledge about functional properties of CLCA2 are still limited, as is the case for CLCA proteins in general. CLCA2 seems to lack significant associations with human diseases, (OMIM Database entry #604003), which is usually the prime driver for generating animal models to investigate more complex functions *in vivo*. This is in contrast to ma *CLCA1*, for which its proposed role as modifier in cystic fibrosis (*Hauber et al., 2003*; *Ritzka et al., 2004*; *Young et al., 2007*) has stimulated the generation of several knockout models in mice (*Erickson et al., 2018*; *Erickson et al., 2015*; *Long et al., 2006*; *Mundhenk et al., 2012*; *Nyström et al., 2018*; *Patel et al., 2006*; *Robichaud et al., 2005*). The lack of a relevant phenotype in any of these models raises general reservations regarding the suitability of deleting a gene for exploring its function, but in the case of *CLCA*, the interpretation was complicated even further. The similar protein architectures of CLCAs and, at least in mice, the overlapping tissue expression patterns of CLCA1 and CLCA2 (*Long et al., 2006*; *Mundhenk et al., 2012*; *Patel et al., 2006*; *Robichaud et al., 2005*) immediately suggested a functional cross-compensation between different CLCA members in the respiratory tract. Given the tissue expression pattern of distinct CLCA members in the epidermis of mice (*Seltmann et al., 2018*), a mutually overlapping and possibly redundant function of certain CLCAs would also question the suitability of the mouse as model organism for studying *CLCA2* function keratinocytes. In chicken, however, the overall architecture of the *CLCA* locus is much less complex than that in mammals (*Bartenschlager et al., 2022*; *Mundhenk et al., 2018*). Moreover, the non-overlapping expression of *gCLCA1* in enterocytes (*Bartenschlager et al., 2022*) and *gCLCA2* in keratinocytes largely excludes the possibility of mutual compensation in chickens. In combination with the high degree of conservation among avian and mammalian *CLCA2* and the recent progress in genome editing techniques in birds (*Chojnacka-Puchta & Sawicka, 2020*; *Morin, Véron & Marcelle, 2017*), the simple structure of the galline *CLCA* gene locus together with its distinct expression pattern might provide a suitable setting for *in vivo* explorations of *CLCA2* functions using genetically edited chickens.

CONCLUSIONS

Our data provide strong evidence for a high conservation of *CLCA2* in mammalian and avian species during evolution. This is in stark contrast to the dynamics and proposed complex functional adaptations in the *CLCA1/3/4* cluster. The slow evolutionary dynamics of *CLCA2* genes insinuates a significant degree of negative selection and a strong functional conservation of *CLCA2* orthologues among birds and mammals, particularly in epidermal keratinocytes. This proposes g*CLCA2* as a suitable object for studying basic functional properties of *CLCA* 2. Furthermore, the simple structure of the *CLCA* locus in birds and

distinct and rather simple expression patterns in chicken may serve as an ideal frame to experimentally address overall CLCA functions in the latter, rather than in mammals.

ACKNOWLEDGEMENTS

The excellent technical support of Annett Kannegießer, Michaela Scholz, Charlene Lamprecht, Nicole Huth, Angela Linke, Cornelia Ziegler, Jana Enders, Sophie Merz, A. Laura Eisenhauer and Mandana J. Poetsch is greatly appreciated. We are grateful for constructive and helpful technical suggestions by Dr. Gerhard Sponder. This work is part of the Ph.D. thesis of Florian Bartenschlager.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Open Access Publication Fund of the Freie Universität Berlin. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Freie Universität Berlin.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Florian Bartenschlager conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored and reviewed drafts of the article, was responsible for data curation, methodology, investigation, project administration and approved the final draft.
- Nikolai Klymiuk conceived and designed the experiments, authored or reviewed drafts of the article, and was responsible for supervision, investigation and approved the final draft.
- Achim D. Gruber conceived and designed the experiments, authored or reviewed drafts of the article, and was responsible for supervision, resources, project administration and approved the final draft.
- Lars Mundhenk conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the article, and was responsible for supervision, project administration, methodology and approved the final draft.

Animal Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

The State Office of Health and Social Affairs Berlin approved the studies (IC 114-ZH70; G 0323/06; T 0104/06).

Data Availability

The following information was supplied regarding data availability:

The raw data, the individual Ct values of gCLCA2 and PGK, a loading control, and the raw, uncropped blot images are available in the Supplementary Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.14202#supplemental-information.

REFERENCES

Abdel-Ghany M, Cheng H-C, Elble RC, Lin H, Di Biasio J, Pauli BU. 2003. The interacting binding domains of the β4 integrin and calcium-activated chloride channels (CLCAs) in metastasis. *Journal of Biological Chemistry* **278**:49406–49416 DOI 10.1074/jbc.M309086200.

Abdel-Ghany M, Cheng H-C, Elble RC, Pauli BU. 2001. The breast cancer β4 integrin and endothelial human CLCA2 mediate lung metastasis. *Journal of Biological Chemistry* **276**:25438–25446 DOI 10.1074/jbc.M100478200.

Akat E, Yenmiş M, Pombal MA, Molist P, Megías M, Arman S, Veselỳ M, Anderson R, Ayaz D. 2022. Comparison of vertebrate skin structure at class level: a review. *The Anatomical Record* Epub ahead of print 27 February 2022 DOI 10.1002/ar.24908.

Bartenschlager F, Klymiuk N, Weise C, Kuropka B, Gruber AD, Mundhenk L. 2022. Evolutionarily conserved properties of CLCA proteins 1, 3 and 4, as revealed by phylogenetic and biochemical studies in avian homologues. *PLOS ONE* **17**:e0266937 DOI 10.1371/journal.pone.0266937.

Bendtsen JD, Nielsen H, Von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: signalP 3.0. *Journal of Molecular Biology* **340**:783–795 DOI 10.1016/j.jmb.2004.05.028.

Bothe MK, Mundhenk L, Beck CL, Kaup M, Gruber AD. 2012. Impaired autoproteolytic cleavage of mCLCA6, a murine integral membrane protein expressed in enterocytes, leads to cleavage at the plasma membrane instead of the endoplasmic reticulum. *Molecules and Cells* **33**:251–257 DOI 10.1007/s10059-012-2217-1.

Braun J, Bothe MK, Mundhenk L, Beck CL, Gruber AD. 2010. Murine mCLCA5 is expressed in granular layer keratinocytes of stratified epithelia. *Histochemistry and Cell Biology* **133**:285–299 DOI 10.1007/s00418-009-0667-0.

Chen H, Liu Y, Jiang C-J, Chen Y-M, Li H, Liu Q-A. 2019. Calcium-activated chloride channel A4 (CLCA4) plays inhibitory roles in invasion and migration through suppressing epithelial-mesenchymal transition via PI3K/AKT signaling in colorectal cancer. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research* **25**:4176 DOI 10.12659/MSM.914195.

Chojnacka-Puchta L, Sawicka D. 2020. CRISPR/Cas9 gene editing in a chicken model: current approaches and applications. *Journal of Applied Genetics* **61**:221–229 DOI 10.1007/s13353-020-00537-9.

Connon CJ, Kawasaki S, Liles M, Koizumi N, Yamasaki K, Nakamura T, Quantock AJ, Kinoshita S. 2006. Gene expression and immunolocalisation of a calcium-activated chloride channel during the stratification of cultivated and developing corneal epithelium. *Cell and Tissue Research* **323**:177–182 DOI 10.1007/s00441-005-0059-2.

Connon CJ, Yamasaki K, Kawasaki S, Quantock AJ, Koizumi N, Kinoshita S. 2004. Calcium-activated chloride channel-2 in human epithelia. *Journal of Histochemistry & Cytochemistry* **52**:415–418 DOI 10.1177/002215540405200313.

Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean Irina M, Austine-Orimoloye O, Azov Andrey G, Barnes I, Bennett R, Berry A, Bhai J, Bignell A, Billis K, Boddu S, Brooks L, Charkhchi M, Cummins C, Da Rin Fioretto L, Davidson C, Dodiya K, Donaldson S, El Houdaigui B, El Naboulsi T, Fatima R, Giron CG, Genez T, Martinez Jose G, Guijarro-Clarke C, Gymer A, Hardy M, Hollis Z, Hourlier T, Hunt T, Juettemann T, Kaikala V, Kay M, Lavidas I, Le T, Lemos D, Marugán JC, Mohanan S, Mushtaq A, Naven M, Ogeh Denye N, Parker A, Parton A, Perry M, Piližota I, Prosovetskaia I, Sakthivel Manoj P, Salam Ahamed Imran A, Schmitt Bianca M, Schuilenburg H, Sheppard D, Pérez-Silva José G, Stark W, Steed E, Sutinen K, Sukumaran R, Sumathipala D, Suner M-M, Szpak M, Thormann A, Tricomi FF, Urbina-Gómez D, Veidenberg A, Walsh Thomas A, Walts B, Willhoft N, Winterbottom A, Wass E, Chakiachvili M, Flint B, Frankish A, Giorgetti S, Haggerty L, Hunt Sarah E, Isley Garth RI, Loveland Jane E, Martin Fergal J, Moore B, Mudge Jonathan M, Muffato M, Perry E, Ruffier M, Tate J, Thybert D, Trevanion Stephen J, Dyer S, Harrison Peter W, Howe Kevin L, Yates Andrew D, Zerbino Daniel R, Flicek P. 2021. Ensembl 2022. *Nucleic Acids Research* **50**:D988–D995 DOI 10.1093/nar/gkab1049.

Di-Poï N, Milinkovitch MC. 2016. The anatomical placode in reptile scale morphogenesis indicates shared ancestry among skin appendages in amniotes. *Science Advances* **2**:e1600708 DOI 10.1126/sciadv.1600708.

Dietert K, Mundhenk L, Erickson NA, Reppe K, Hocke AC, Kummer W, Witzenrath M, Gruber AD. 2015. Murine CLCA5 is uniquely expressed in distinct niches of airway epithelial cells. *Histochemistry and Cell Biology* **143**:277–287 DOI 10.1007/s00418-014-1279-x.

Elble RC, Walia V, H-c Cheng, Connon CJ, Mundhenk L, Gruber AD, Pauli BU. 2006. The putative chloride channel hCLCA2 has a single C-terminal transmembrane segment. *Journal of Biological Chemistry* **281**:29448–29454 DOI 10.1074/jbc.M605919200.

Ellegren H. 2005. The avian genome uncovered. *Trends in Ecology & Evolution* **20**:180–186 DOI 10.1016/j.tree.2005.01.015.

Erickson NA, Dietert K, Enders J, Glauben R, Nouailles G, Gruber AD, Mundhenk L. 2018. Soluble mucus component CLCA1 modulates expression of leukotactic cytokines and BPIFA1 in murine alveolar macrophages but not in bone marrow-derived macrophages. *Histochemistry and Cell Biology* **149**:619–633 DOI 10.1007/s00418-018-1664-y.

Erickson N, Gruber A, Mundhenk L. 2020. The family of chloride channel regulator, calcium-activated proteins in the feline respiratory tract: a comparative perspective on airway Diseases in man and animal models. *Journal of Comparative Pathology* **174**:39–53 DOI 10.1016/j.jcpa.2019.10.193.

Erickson NA, Nyström EE, Mundhenk L, Arike L, Glauben R, Heimesaat MM, Fischer A, Bereswill S, Birchenough GM, Gruber AD. 2015. The goblet cell protein Clca1 (alias mClca3 or Gob-5) is not required for intestinal mucus synthesis, structure and barrier function in naive or DSS-challenged mice. *PLOS ONE* **10**:e0131991 DOI 10.1371/journal.pone.0131991.

Evans SR, Thoreson WB, Beck CL. 2004. Molecular and functional analyses of two new calcium-activated chloride channel family members from mouse eye and intestine. *Journal of Biological Chemistry* **279**:41792–41800 DOI 10.1074/jbc.M408354200.

Gibson A, Lewis AP, Affleck K, Aitken AJ, Meldrum E, Thompson N. 2005. hCLCA1 and mCLCA3 are secreted non-integral membrane proteins and therefore are not ion channels. *Journal of Biological Chemistry* **280**:27205–27212 DOI 10.1074/jbc.M504654200.

Gruber AD, Schreur KD, Ji H-L, Fuller CM, Pauli BU. 1999. Molecular cloning and transmembrane structure of hCLCA2 from human lung, trachea, and mammary gland. *American Journal of Physiology-Cell Physiology* **276**:C1261–C1270 DOI 10.1152/ajpcell.1999.276.6.C1261.

Hamalainen L, Bart G, Takabe P, Rauhala L, Deen A, Pasonen-Seppanen S, Karkkainen E, Karna R, Kumlin T, Tammi MI, Tammi RH. 2021. The calcium-activated chloride channel-associated protein rCLCA2 is expressed throughout rat epidermis, facilitates apoptosis and is downmodulated by UVB. *Histochemistry and Cell Biology* **155**:605–615 DOI 10.1007/s00418-021-01962-5.

Hammesfahr B, Odronitz F, Mühlhausen S, Waack S, Kollmar M. 2013. GenePainter: a fast tool for aligning gene structures of eukaryotic protein families, visualizing the alignments and mapping gene structures onto protein structures. *BMC Bioinformatics* **14**:1–11.

Hatje K, Hammesfahr B, Kollmar M. 2013. WebScipio: reconstructing alternative splice variants of eukaryotic proteins. *Nucleic Acids Research* **41**:W504–W509 DOI 10.1093/nar/gkt398.

Hauber HP, Manoukian JJ, Nguyen LH, Sobol SE, Levitt RC, Holroyd KJ, McElvaney NG, Griffin S, Hamid Q. 2003. Increased expression of interleukin-9, interleukin-9 receptor, and the calcium-activated chloride channel hCLCA1 in the upper airways of patients with cystic fibrosis. *The Laryngoscope* **113**:1037–1042 DOI 10.1097/00005537-200306000-00022.

Hirokawa T, Boon-Chieng S, Mitaku S. 1998. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14**:378–379 DOI 10.1093/bioinformatics/14.4.378.

Hoshino M, Morita S, Iwashita H, Sagiya Y, Nagi T, Nakanishi A, Ashida Y, Nishimura O, Fujisawa Y, Fujino M. 2002. Increased expression of the human Ca2+-activated

Cl − channel 1 (CaCC1) gene in the asthmatic airway. *American Journal of Respiratory and Critical Care Medicine* **165**:1132–1136 DOI 10.1164/ajrccm.165.8.2107068.

Hou T, Zhou L, Wang L, Kazobinka G, Zhang X, Chen Z. 2017. CLCA4 inhibits bladder cancer cell proliferation, migration, and invasion by suppressing the PI3K/AKT pathway. *Oncotarget* **8**:93001 DOI 10.18632/oncotarget.21724.

Käll L, Krogh A, Sonnhammer EL. 2007. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Research* **35**:W429–W432 DOI 10.1093/nar/gkm256.

Koegel H, Alzheimer C. 2001. Expression and biological significance of Ca2+-activated ion channels in human keratinocytes. *The FASEB Journal* **15**:145–154 DOI 10.1096/fj.00-0055com.

Lenart A, Dudkiewicz M, Grynberg M, Pawłowski K. 2013. CLCAs-a family of metalloproteases of intriguing phylogenetic distribution and with cases of substituted catalytic sites. *PLOS ONE* **8**:e62272 DOI 10.1371/journal.pone.0062272.

Long AJ, Sypek JP, Askew R, Fish SC, Mason LE, Williams CM, Goldman SJ. 2006. Gob-5 contributes to goblet cell hyperplasia and modulates pulmonary tissue inflammation. *American Journal of Respiratory Cell and Molecular Biology* **35**:357–365 DOI 10.1165/rcmb.2005-0451OC.

Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS. 2020. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Research* **48**:D265–D268 DOI 10.1093/nar/gkz991.

Morin V, Véron N, Marcelle C. 2017. CRISPR/Cas9 in the chicken embryo. In: Sheng G, ed. *Avian and reptilian developmental biology*. *Methods in molecular biology*, vol. 1650. New York, NY: Humana Press DOI 10.1007/978-1-4939-7216-6_7.

Mundhenk L, Alfalah M, Elble RC, Pauli BU, Naim HY, Gruber AD. 2006. Both cleavage products of the mCLCA3 protein are secreted soluble proteins. *Journal of Biological Chemistry* **281**:30072–30080 DOI 10.1074/jbc.M606489200.

Mundhenk L, Erickson NA, Klymiuk N, Gruber AD. 2018. Interspecies diversity of chloride channel regulators, calcium-activated 3 genes. *PLOS ONE* **13**:e0191512 DOI 10.1371/journal.pone.0191512.

Mundhenk L, Johannesson B, Anagnostopoulou P, Braun J, Bothe MK, Schultz C, Mall MA, Gruber AD. 2012. mCLCA3 does not contribute to calcium-activated chloride conductance in murine airways. *American Journal of Respiratory Cell and Molecular Biology* **47**:87–93 DOI 10.1165/rcmb.2010-0508OC.

Nakanishi A, Morita S, Iwashita H, Sagiya Y, Ashida Y, Shirafuji H, Fujisawa Y, Nishimura O, Fujino M. 2001. Role of gob-5 in mucus overproduction and airway hyperresponsiveness in asthma. *Proceedings of the National Academy of Sciences of the United States of America* **98**:5175–5180 DOI 10.1073/pnas.081510898.

Nyström EE, Birchenough GM, van der Post S, Arike L, Gruber AD, Hansson GC, Johansson ME. 2018. Calcium-activated chloride channel regulator 1 (CLCA1) controls mucus expansion in colon by proteolytic activity. *EBioMedicine* **33**:134–143 DOI 10.1016/j.ebiom.2018.05.031.

Olias P, Adam I, Meyer A, Scharff C, Gruber AD. 2014. Reference genes for quantitative gene expression studies in multiple avian species. *PLOS ONE* **9**:e99678 DOI 10.1371/journal.pone.0099678.

Patel AC, Brett TJ, Holtzman MJ. 2009. The role of CLCA proteins in inflammatory airway disease. *Annual Review of Physiology* **71**:425–449 DOI 10.1146/annurev.physiol.010908.163253.

Patel AC, Morton JD, Kim EY, Alevy Y, Swanson S, Tucker J, Huang G, Agapov E, Phillips TE, Fuentes ME. 2006. Genetic segregation of airway disease traits despite redundancy of calcium-activated chloride channel family members. *Physiological Genomics* **25**:502–513 DOI 10.1152/physiolgenomics.00321.2005.

Pawłowski K, Lepistö M, Meinander N, Sivars U, Varga M, Wieslander E. 2006. Novel conserved hydrolase domain in the CLCA family of alleged calcium-activated chloride channels. *Proteins: Structure, Function, and Bioinformatics* **63**:424–439 DOI 10.1002/prot.20887.

Plog S, Grötzsch T, Klymiuk N, Kobalz U, Gruber AD, Mundhenk L. 2012a. The porcine chloride channel calcium-activated family member pCLCA4a mirrors lung expression of the human hCLCA4. *Journal of Histochemistry & Cytochemistry* **60**:45–56 DOI 10.1369/0022155411426455.

Plog S, Klymiuk N, Binder S, Van Hook MJ, Thoreson WB, Gruber AD, Mundhenk L. 2015. Naturally occurring deletion mutants of the pig-specific, intestinal crypt epithelial cell protein CLCA4b without apparent phenotype. *PLOS ONE* **10**:e0140050 DOI 10.1371/journal.pone.0140050.

Plog S, Mundhenk L, Bothe MK, Klymiuk N, Gruber AD. 2010. Tissue and cellular expression patterns of porcine CFTR: similarities to and differences from human CFTR. *Journal of Histochemistry & Cytochemistry* **58**:785–797 DOI 10.1369/jhc.2010.955377.

Plog S, Mundhenk L, Klymiuk N, Gruber AD. 2009. Genomic, tissue expression, and protein characterization of pCLCA1, a putative modulator of cystic fibrosis in the pig. *Journal of Histochemistry & Cytochemistry* **57**:1169–1181 DOI 10.1369/jhc.2009.954594.

Plog S, Mundhenk L, Langbein L, Gruber AD. 2012b. Synthesis of porcine pCLCA2 protein during late differentiation of keratinocytes of epidermis and hair follicle inner root sheath. *Cell and Tissue Research* **350**:445–453 DOI 10.1007/s00441-012-1482-9.

Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER web server: 2018 update. *Nucleic Acids Research* **46**:W200–W204 DOI 10.1093/nar/gky448.

Pult F, Fallah G, Braam U, Detro-Dassen S, Niculescu C, Laube B, Schmalzing G. 2011. Robust post-translocational N-glycosylation at the extreme C-terminus of membrane and secreted proteins in Xenopus laevis oocytes and HEK293 cells. *Glycobiology* **21**:1147–1160 DOI 10.1093/glycob/cwr013.

Ramena G, Yin Y, Yu Y, Walia V, Elble RC. 2016. CLCA2 interactor EVA1 is required for mammary epithelial cell differentiation. *PLOS ONE* **11**:e0147489 DOI 10.1371/journal.pone.0147489.

Range F, Mundhenk L, Gruber A. 2007. A soluble secreted glycoprotein (eCLCA1) is overexpressed due to goblet cell hyperplasia and metaplasia in horses with recurrent airway obstruction. *Veterinary Pathology* **44**:901–911 DOI 10.1354/vp.44-6-901.

Ritzka M, Stanke F, Jansen S, Gruber AD, Pusch L, Woelfl S, Veeze HJ, Halley DJ, Tümmler B. 2004. The CLCA gene locus as a modulator of the gastrointestinal basic defect in cystic fibrosis. *Human Genetics* **115**:483–491 DOI 10.1007/s00439-004-1190-y.

Robichaud A, Tuck SA, Kargman S, Tam J, Wong E, Abramovitz M, Mortimer J, Burston HE, Masson P, Hirota J. 2005. Gob-5 is not essential for mucus overproduction in preclinical murine models of allergic asthma. *American Journal of Respiratory Cell and Molecular Biology* **33**:303–314 DOI 10.1165/rcmb.2004-0372OC.

Seltmann K, Meyer M, Sulcova J, Kockmann T, Wehkamp U, Weidinger S, Keller Uaufdem, Werner S. 2018. Humidity-regulated CLCA2 protects the epidermis from hyperosmotic stress. *Science Translational Medicine* **10**:eaao4650 DOI 10.1126/scitranslmed.aao4650.

Tamura K, Stecher G, Kumar S. 2021. MEGA11: molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution* **38**:3022–3027 DOI 10.1093/molbev/msab120.

Walia V, Ding M, Kumar S, Nie D, Premkumar LS, Elble RC. 2009. hCLCA2 is a p53-inducible inhibitor of breast cancer cell proliferation. *Cancer Research* **69**:6624–6632 DOI 10.1158/0008-5472.CAN-08-4101.

Walia V, Yu Y, Cao D, Sun M, McLean JR, Hollier BG, Cheng J, Mani SA, Rao K, Premkumar L. 2012. Loss of breast epithelial marker hCLCA2 promotes epithelial-to-mesenchymal transition and indicates higher risk of metastasis. *Oncogene* **31**:2237–2246 DOI 10.1038/onc.2011.392.

Young FD, Newbigging S, Choi C, Keet M, Kent G, Rozmahel RF. 2007. Amelioration of cystic fibrosis intestinal mucous disease in mice by restoration of mCLCA3. *Gastroenterology* **133**:1928–1937.e1923 DOI 10.1053/j.gastro.2007.10.007.

Yu Y, Walia V, Elble RC. 2013. Loss of CLCA4 promotes epithelial-to-mesenchymal transition in breast cancer cells. *PLOS ONE* **8**:e83943 DOI 10.1371/journal.pone.0083943.

Yurtsever Z, Sala-Rabanal M, Randolph DT, Scheaffer SM, Roswit WT, Alevy YG, Patel AC, Heier RF, Romero AG, Nichols CG. 2012. Self-cleavage of human CLCA1 protein by a novel internal metalloprotease domain controls calcium-activated chloride channel activation. *Journal of Biological Chemistry* **287**:42138–42149 DOI 10.1074/jbc.M112.410282.

Concluding discussion and outlook

This cumulative thesis aimed to characterize the two recently predicted galline CLCA members (Mundhenk et al. 2018) in terms of exon-intron structure, protein domain architecture, biochemical properties and protein expression. By comparison with the mammalian CLCA clusters 1-4, conserved traits and differences between the two galline CLCAs and mammalian CLCAs were identified. These analyses provided evidence for the most likely evolutionary scenario of the *CLCA* gene family among avian and mammalian species.

1.8 A hypothetical ancestral *CLCA1* gene diversified in mammals but not in birds

Recently published phylogenetic analyses of galline and mammalian CLCA protein sequences revealed a monophyletic group of gCLCA1 and mammalian CLCA1, 3 and 4 separate from a monophyletic group of gCLCA2 and mammalian CLCA2 (Mundhenk et al. 2018). This finding was in accordance to the results obtained from a protein domain specific analysis including the CLCA sequences of twelve avian species and fifteen mammalian species, as it is illustrated in publication 1 (chapter 1.6). Noteworthy, all avian CLCA1 sequences formed a monophyletic group separate from the mammalian CLCA 1, 3 and 4 sequences in each of the analyzed domain. When phylogenetic relationships of the domains were investigated, variable genetic distances between avian CLCA1 and mammalian CLCA1, 3 and 4 were detected (publication 1; chapter 1.6). However, and in accordance to the phylogenetic analysis of full length gCLCA1 and mammalian CLCA1, 3 and 4 protein sequences (Mundhenk et al. 2018), the first nodes of avian CLCA1 and mammalian CLCA3 had the closest distance in most of the analyzed domains (publication 1; chapter 1.6). Based on the phylogenetic trees provided in publication 1 (chapter 1.6), mammalian CLCA3 had a closer genetic distance to mammalian CLCA4 than to CLCA1 in all domains, except for the vWA domain. These aa sequence-based data suggest a diversification of an ancestral CLCA1 that ended up in complex, highly dynamic and often lineage-specific patterns of duplication events in the mammals, which seem to not have happened in birds.

But which properties in terms of gene structure, protein architecture, protein biochemistry and cell type specific expression pattern does an avian CLCA1 possess? And moreover, does an avian CLCA1 combine all the characteristics that are split into three specialized mammalian protein clusters? To address these questions, *in silico* analyses of the galline CLCA1 gene and protein were performed. Furthermore, the gene was cloned, its protein biochemical properties and mRNA expression as well as its cell specific protein expression in tissues were analyzed. The gCLCA1 protein is encoded by 14 exons and has a protein domain architecture as it was described for mammalian CLCA1, 3 and 4 proteins (publication 1; chapter 1.6). This architecture includes a signal sequence and the N-CLCA, vWA, bsr, and fn III domains. In addition, a putative cleavage site similar to that of mammalian CLCA1, 3 and 4 was detected

in silico (publication 1; chapter 1.6). Interestingly, and in accordance to CLCA4 proteins, a TMC was present as described in publication 1 (chapter 1.6). Moreover and as demonstrated for mammalian CLCA4, gCLCA1 is a heavily asparagine-linked glycosylated, cleaved protein (publication 1; chapter 1.6). The N-terminal fragment was secreted, while the C-terminal cleavage product remained associated with the cell membrane (publication 1; chapter 1.6). Interestingly, when the catalytic site responsible for self-cleavage was mutated, the cleavage was completely abrogated in gCLCA1, as it was described only for mammalian CLCA1 (Pawłowski et al. 2006; Bothe et al. 2011; publication 1; chapter 1.6). These findings support hypotheses 1.5.1 and 1.5.2 and indicate that the gene structure, protein domain architecture, heavy glycosylation, cleavage and secretion of the N-terminal fragment are ancient characteristics, which were likely also present in an ancestral CLCA1 (scientific question 1.5.6).

Remarkably, gCLCA1 protein was exclusively found in a single cell type: in analogy to CLCA4 in mammals, it was expressed along the apical surface of enterocytes in villi and crypts along the galline alimentary canal. However, no other cell types, as for example goblet cells which are known to express mammalian CLCA1 were positive for gCLCA1 (publication 1; chapter 1.6). This finding did not support hypothesis 1.5.3 but suggests the expression in enterocytes as an old, conserved trait, which was likely also present in an ancestral CLCA1 (scientific question 1.5.6).

Taken together, the phylogenetic relationship between gCLCA1 and mammalian CLCA1, 3 and 4, as well as their biochemical properties and cell-specific protein expression patterns allow speculations about the evolution of these genes among birds and mammals. Based on the mentioned data of mammalian CLCA1, 3 and 4 as well as the galline CLCA1, a hypothetical ancient *CLCA1* gene was likely expressed in intestinal enterocytes in villi and crypts along the alimentary canal of an avian-mammalian ancestor. Also, the encoded protein might have been heavily glycosylated and cleaved. A TMC likely associated the C-terminal cleavage product with the plasma membrane, while the N-terminal fragment might have been secreted. The hypothetical ancestral CLCA1 putatively underwent genetic duplications in the mammalian lineage, but not in the avian phylum. This likely resulted in an adaptive radiation of these genes only in mammals. For some of the mammalian species, there are preliminary data available that support the model of gene subfunctionalization after duplication. For example, a spatial separation of the expression of the two porcine CLCA4 proteins in villus (CLCA4a) and crypts (CLCA4b) might indicate such a subfunctionalization process on the level of protein expression (Plog et al. 2012a; Plog et al. 2015). In parallel, the pseudogenization of CLCA4b in a porcine subpopulation (Plog et al. 2015) supports the hypothesis on the highly dynamic evolution of these genes in mammals. The spatial expressional separation of both porcine CLCA4 proteins

indicates a subfunctionalization prior to inactivation by pseudogenization of CLCA4b. This suggests the pseudogenization of CLCA4b as a recent event in the history of domestic pigs (Plog et al. 2015).

1.9 The strong conservation of CLCA2 in keratinocytes implicates an important function in these cells

Phylogenetic analyses revealed a close genetic distance not only between the galline CLCA2 and mammalian CLCA2 but even among mammalian CLCA2 aa sequences (Mundhenk et al. 2018). This finding provided first evidence for a high degree of conservation of these orthologues and raised questions on the kind and extent of this conservation regarding exon-intron structure, protein architecture, biochemical properties and protein expression in the mammalian and avian taxa. The three analyzed avian CLCA2 orthologous proteins of ostrich, chicken and quail are all encoded by 14 exons, as it is the case for mammalian CLCA2 (publication 2; chapter 1.7). Interestingly, not only the number of exons in avian and mammalian *CLCA2* were identical, but a majority of exons also corresponded to each other in their number of nucleotides (publication 2; chapter 1.7). These data support the hypothesis 1.5.1 and demonstrate the high degree of conservation in the exon-intron structure of avian and mammalian CLCA2.

A multiple sequence alignment of avian and mammalian CLCA2 protein sequences revealed a typical CLCA2 protein domain architecture in avian CLCA2 as it was reviewed by Patel for mammals (Patel et al. 2009). Both taxa have an N-CLCA, a vWA, a bsr, a fn III and a TMC domain. In contrast to mammalian CLCA1, 3 and 4 (Patel et al. 2009), mammalian CLCA2 proteins lack an intact metal ion dependent adhesion site (MIDAS) in the vWA domain. This lack of an intact MIDAS was also found in all avian CLCA2 aa sequences (publication 2; chapter 1.7). For a second CLCA-typical aa motif, the beta4-integrin binding motif, the results were less consistent. The presence, the location and the number of the beta4-integrin binding motifs were inconsistent among mammalian CLCA2 sequences and completely absent in avian CLCA2 proteins (publication 2; chapter 1.7). These findings support the hypothesis 1.5.4 in terms of protein domain architecture with the exception of the absence of a beta4-integrin binding motif in avian CLCA2. This aa motif appears to be a lineage specific trait with a high degree of variation in mammals. These results suggest this motif to be under divergent selection compared to other parts of the protein in mammals.

Biochemical properties were analyzed for the galline CLCA2 protein as a putative avian prototype. The cloned, heterologous expressed gCLCA2 appeared to be a cleaved, glycosylated protein possessing a TMC (publication 2; chapter 1.7) as it was described for mammalian CLCA2 (Elble et al. 2006; Braun et al. 2010). The conservation of protein

biochemical properties of CLCA2 among chicken and mammalian species support the hypothesis 1.5.4.

Remarkably, the galline CLCA2 protein was exclusively found in keratinocytes of the galline epidermis as an anatomical unit of the cutaneous skin and certain mucus membranes (publication 2; chapter 1.7). This expression spectrum appears to be slightly narrower than that of mammals, where CLCA2 protein was also found in cells of the respiratory tract, with a species-specific variation (Dietert et al. 2015). As it was described for the presence of a beta4-integrin binding motif in mammalian CLCA2 sequences (Abdel-Ghany et al. 2001), the cell specific expression of CLCA2 protein in this anatomical location seems to be a mammal-specific trait with a degree variation depending on the species. Analyses of cutaneous sections of four bird species, including ostrich, chicken, quail, and turkey revealed the presence of their CLCA2 protein orthologues in the epidermis, similar to that of chicken (publication 2; chapter 1.7). The immunohistochemical staining pattern of avian CLCA2 slightly diverged from the one described in mice and men. The murine CLCA2 was only found in the epidermal granular layer (Bothe et al. 2008) and the human CLCA2 was mainly detected in the basal layer (Connon et al. 2004). In contrast to these findings, the avian CLCA2 protein orthologues were also detected in keratinocytes between the basal and the granular layer of the avian epidermis (publication 2; chapter 1.7). Taken together, the expression of CLCA2 protein in keratinocytes seems to be a highly conserved trait among birds and mammals, which supports hypothesis 1.5.5, albeit a slight difference depending on the expression patterns in the epidermal layer has been described in mammals. This finding implicates a significant function of these proteins in epidermal keratinocytes, which might have been unaltered since the mammal and avian lineage diverged approximately 300 million years ago (Namekawa and Lee 2009). Thus, an ancient CLCA2 protein might have been a cleaved, glycosylated transmembrane protein containing an N-CLCA, vWA, bsr and fn III domain. Furthermore, this hypothetical ancient CLCA2 has likely been expressed in keratinocytes. Only slight modifications of this proteins evolved since the mammalian-avian speciation event (research question 1.5.6).

This high degree of conservation allows speculations on a possible function of CLCA2 in the mammalian and avian epidermis that might be common for both of the taxa. Considering the concept of orthologue conjecture (chapter 1.2.2), the function of mammalian and avian CLCA2 proteins in the epidermis could be very similar: all proteins share a high degree of phylogenetic homology, a similar protein architecture, biochemical properties and cell-specific expression. Results from analyses of CLCA2 function in mammals might be transferred to birds and vice versa. For example, an interaction of mammalian CLCA2 with the anion channel TMEM16A, which is known to be also expressed in mammalian epidermal keratinocytes (Choi et al. 2021), has been recently demonstrated *in vitro* (Sharma et al. 2018). Future studies have to reveal if

a similar interaction also applies for avian CLCA2 and TMEM16A. Besides that, proteases are essential components for the maintenance of intactness and differentiation of the integument, including the epidermis (Zeeuwen 2004). Since CLCA2 proteins have an intact HEXXH-protease motif, the role of CLCA2 in these processes should be further investigated using *in vitro* and *in vivo* models.

1.10 The chicken appears to be a promising model organism for functional CLCA2 investigations

As described in the introduction, several murine *CLCA1* knockout models have been generated to study CLCA functions, however none of them developed a clear phenotype under unchallenged conditions (Robichaud et al. 2005; Long et al. 2006; Patel et al. 2006; Mundhenk et al. 2012; Erickson et al. 2015; Erickson et al. 2018; Nyström et al. 2018). This observation hampered functional CLCA analyses in the recent years. Due to the high redundancy and overlapping expression pattern of CLCA proteins in mice and other mammals, the absence of a phenotype in knockout models appears plausible by a compensation of the knockout by other *CLCA* members expressed in the same environment. For example, in the intestine of mice, such a compensation might be accomplished by *CLCA4*. Furthermore and similar to this assumption concerning *CLCA1*, a knockout of *CLCA2* in murine keratinocytes might be compensated by *CLCA3*. This highlights the weaknesses of a murine *CLCA* knockout model and demonstrates the need for another organism with a lower genetic complexity.

In this context, the chicken appears to be characterized by some merits. First, only two *CLCA* genes exist in the galline genome and their gene products are expressed by different cell types (gCLCA1 in enterocytes as described in publication 1, chapter 1.6 and gCLCA2 in keratinocytes as described in publication 2; chapter 1,7). As a consequence, both proteins are present in spatially separated environments (intestine or skin and cornified mucus membranes, respectively). With respect to these findings, the compensation of one knocked out gene by the other homologue appears very unlikely. Second, CLCA2 is strongly conserved among chicken and mammals not only in terms of protein domain architecture, but also in its protein biochemistry and expression. Thus, it does seem to be reasonable that functional insights gained by experiments on galline CLCA2 can be transferred to mammals. Third, there are versatile genome editing techniques available for chicken that allow the generation of such a knockout model on a routine basis (Morin et al. 2017; Chojnacka-Puchta and Sawicka 2020). Factors such as species-specific generation time and economic aspects, including housing requirements, are inferior compared to those of murine models but are still affordable compared to primate models, for example. Furthermore, since the avian reproduction is mediated by eggs, the galline embryo is easily accessible and studies on the role of *CLCA* in differentiation and development can be performed more conveniently compared to mammals.

However, for a couple of reasons, a *gCLCA1* knockout chicken will probably not serve as a useful model for studying mammalian CLCA1, which appears to be the currently most relevant CLCA member in biomedical research (Patel et al. 2009). First, especially the expression in goblet cells and the fully secreted character of mammalian CLCA1 deviate significantly from that of gCLCA1, which is expressed in enterocytes and possesses a transmembrane domain that anchors the C-terminal cleavage product in the plasma membrane. Second, CLCA1 is highly associated with asthma (Nakanishi et al. 2001; Hoshino et al. 2002; Anton et al. 2005; Erickson et al. 2020), cystic fibrosis (Van Der Doef et al. 2010) or COPD (Hegab et al. 2004) in mammals , which are diseases that have not yet been described in the chicken. However, when the function of mammalian CLCA4 in enterocytes is the focus of interest, a *gCLCA1* knockout model might provide some useful data, since both share a comparable expression pattern in the gut and have a similar protein architecture and biochemistry in common. Noteworthy and in contrast to mammals, no other CLCA member is expressed in the galline gut, which might compensate the loss of *gCLCA1* and by this mask a knockout phenotype.

1.11 Outlook and limitations of this cumulative thesis

Early evolutionists such as Carl von Linné, Jean-Baptiste Lamarck, Charles Darwin, Alfred Russel Wallace or Thomas Henry Huxley based their hypotheses on phenotypical traits, which they obtained by exact description and comparison of morphological findings in recent life forms and/or fossils. By this, they substantially formed the modern view on mechanisms influencing macroevolution, which is defined as the formation of systematic classes here (Hautmann 2020). In contrast to these pioneers of evolution, we nowadays are living in a time where we are aware of the existence of genes and their products. We now know about the laws according to which genetic exchange happens between individuals and populations. We can infer the degree of relationship between individuals and species by comparing their gene and genome sequences. As we move forward into the post-genomic era, when genomes of more and more species will be decoded, the information associated with gene products will be the fundamental resource to understand their function and the forces that shaped them. Moreover, this information on gene products, including data on expression, protein architecture, and biochemical properties, will be essential pieces of a puzzle, the whole of which will be a refinement of the modern view of the (molecular) tree of life. With respect to the early evolutionists, this might be accomplished by a similar approach to the one that was already used over 150 years ago, when theories of evolution where published: Exact description and comparison of traits. However, instead of being restricted to macroscopic phenotypes, we will be able to compare data on, for example, gene and protein expression patterns, protein architecture or biochemical properties. This cumulative thesis adopted this approach using the *CLCA* gene family as an example and identified significant similarities but

also important differences of these genes and their associated properties between avian and mammalian species. In particular, the differences between genes and gene products underscore the importance of such analyses when it comes to the selection of an animal model to investigate physiological and pathological functions of a gene and the transferability of the data obtained from this investigation. Once again, it becomes obvious that a clearly defined research question and the specific anatomy, physiology, as well as the genetic background of a species of interest must be the most important factors influencing the selection of an animal model. To some extent, this is still in contrast to the common practice of using established mammalian models such as mice, just based on factors like the availability of established strains, husbandry systems, short reproduction cycles, small body sizes, and other, also economically driven aspects. Using the data obtained in this cumulative thesis, it appears reasonable to choose the chicken as a model organism for a defined range of functional investigations on the *CLCA* gene family. Due to the knowledge of the evolutionary relationship among chicken and mammals, which is available now, the transferability of a defined range of data obtained from a *CLCA* knockout chicken to mammalian *CLCA* becomes arguable.

In terms of modeling the evolutionary background of CLCA, some aspects of this cumulative thesis have to be critically discussed. Accurate modeling of gene evolution requires the analysis of a huge amount of nucleotide and/or protein sequences. In this cumulative thesis, the evolutionary background among avian and mammalian *CLCA* genes was analyzed using only up to equal to or little greater than a dozen protein sequences from each of these taxons, respectively. For comparative analyses of protein biochemical properties and cell specific protein expression, both galline *CLCA* genes were cloned, heterologously expressed and analyzed *in vitro* using immunoblot techniques or in tissues using immunohistochemistry. Since data from these experiments are available only for chicken, there might be a bias towards gallinaceous birds. Furthermore, no data specifying the CLCAs of lower mammalian species such as marsupials are available to date. Due to the gain of more and more whole-genome sequences of animals, there will be a tremendous increase of comparable sequence data from many other species which can be subjected to such phylogenetic investigations in the close future. Furthermore, systematic biochemical and expressional studies focusing on far more bird and mammalian species will be necessary to correct the bias toward gallinaceous birds and higher mammals. These studies will be necessary to refine the phylogenetic relationships, unveil selection pressures and lineage specific variations of CLCAs as well. Such analyses will help to refine the identification of conserved traits and dispensable elements of these proteins. Comparative, functional assays such as the analysis of the ability to activate TMEM16A, the identification of educts for the HEXXH protease motif or the interaction with inflammatory cells should be systematically established and investigated for CLCAs across

various species in the future. Furthermore, it was reported that the most distant animal possessing a mammalian-like CLCA protein is the placozoan species *Trichoplax adhaerens* (Lenart et al. 2013). Besides the modeling of the rather recent evolution of CLCA among birds and mammals, it will be interesting to systematically analyze the CLCA sequences, the cellular expression and biochemical properties of CLCAs from invertebrates like Radiata or Protostomia, vertebrates like fish, amphibians, reptiles and lower mammals to unveil more delicate phylogenetic signals and to derive the most likely evolutionary background of all mammalian-like CLCAs. Such investigations might also provide data that help to elucidate the cause of the complex evolutionary developments of CLCA.

It became obvious that the cell-specific expression pattern is a prominent factor in which CLCA proteins vary significantly among the CLCA clusters. With respect to this finding, it has to be considered that the regulation of protein expression involves two dimensions: First, the gene to be expressed is transcribed and second, the messenger RNA generated by transcription is translated into a protein. Factors influencing the cell specificity of gene transcription are mainly promoters, however, the degree of methylation of cytosine and guanine (CpG) islands and histones, enhancer and silencer sequences or RNA interference also significantly influence the transcription of genes in eukaryotes. Comparative identification and analyses of *CLCA* promoters between avian and mammalian species were not part of this cumulative thesis but might provide an explanation for these differences in cell type-specific expression patterns. Assays and techniques for *CLCA*-specific promoter analyses should be established and such investigations should be done in the future.

At this point, let’s return to the questions posed in the third section of the introduction: Is the complexity and diversity of these genes a unique entity of mammals? And what is the evolutionary background of this complexity? We are aware of the likely evolutionary background of CLCAs between mammalian and avian species at this point. The characteristics of hypothetical ancient *CLCA* genes and their gene products were inferred by this cumulative thesis. There were identified conserved regions of CLCA that did not change in the course of evolution between birds and mammals. Conversely, highly variable sections with impact on the biochemical processing, in which these genes and gene products differ between the species were also found. The first question is yet to be fully answered, however it became obvious for birds that there are species with a significantly less complex *CLCA* genotype compared to mammals.

1.12 Conclusions

The *CLCA* gene family is characterized by highly dynamic developments among birds and mammals, including gene duplications and deletions in mammalian *CLCA3* and *4* genes (Mundhenk et al. 2018). Additionally, there are also highly conserved genes such as *CLCA2*, which are under strong negative selection. The chicken possesses only two *CLCA* members, with a similar exon-intron structure compared to CLCA members of mammals. The galline CLCA1 has the highest congruence with mammalian CLCA4, but also combines properties of mammalian CLCA1 and 3. Galline CLCA2 is highly similar compared to mammalian CLCA2 and might fulfill similar functions in mammals and birds. Due to the simple *CLCA* genotype, the non-overlapping expression pattern of both CLCAs and the high similarity of galline and mammalian CLCA2 proteins, the chicken might be an auspicious knockout model for studying CLCA2 functions *in vivo*. Based on the results of this cumulative thesis, one came to speculate that the ancient CLCAs of a mammalian-avian ancestor might have been cleaved, glycosylated proteins with an N-CLCA, a vWA, a bsr, an fn III and a TMC domain. The gCLCA1 / mammalian CLCA1, 3 and 4 ancestor might have been expressed in enterocytes, while the ancient CLCA2 could have been expressed in keratinocytes.

Summary

Evolutionary Diversity of *CLCA* Genes Among Birds and Mammals

Florian Bartenschlager

The *chloride channel regulators, calcium activated* (*CLCA*) gene family has mainly been associated with cancer and chronic inflammatory airway diseases but the presumably complex cellular functions of these gene products are still widely unknown. The family comprises four distinct genetic clusters in mammals, termed *CLCA1* to *CLCA4*. It is highly complex and diverse and includes amplified or inactivated genes with a high degree of variation between species. For example, in contrast to mice with eight *CLCA*s including one inactivated gene, humans have only three intact *CLCA*s and one inactivated gene. The tissue and cellular expression patterns of different CLCA homologues within a species are also often redundant. This complexity and redundancy of the CLCA members might be a reason, why the function of this gene family has not been revealed yet based on a mammalian model organism. The complexity and redundancy of mammalian CLCAs raise two questions: Are the complexity and diversity of these genes unique features of mammals? And second, what is the evolutionary background of these peculiar developments? To address these questions, data on CLCA homologues obtained from evolutionarily more distant species are needed. Recently, a rather simple *CLCA* gene locus was predicted for the chicken, comprising only two *CLCA* genes. In a phylogenetic analysis, the first galline CLCA gene product, gCLCA1, was found closely related to mammalian CLCA1, 3 and 4. In contrast, the second one, gCLCA2, seemed more closely related to mammalian CLCA2 than to gCLCA1 or mammalian CLCA1, 3 and 4.

In this cumulative thesis, the galline *CLCA* genes and their genomic structures were analyzed and both members were cloned. Their protein architecture and biochemical properties were investigated *in silico* and *in vitro*. In addition, their mRNA as well as their cellular protein expression patterns were analyzed. All data were compared to mammalian CLCA1, 2, 3, and 4. Both avian proteins are encoded by 14 exons and are located in a conserved locus between the *Outer Dense Fiber of Sperm Tails 2-Like (ODF2L)* and *SH3-Domain GRB2-Like Endophilin B1 (SH3GLB1)* genes. gCLCA1 combines many properties of mammalian CLCA1, 3 and 4 as it was shown to be a cleaved protein with a typical CLCA domain architecture. Despite its relatively high phylogenetic distance to mammalian CLCA4, it shares most common traits with this member. This includes heavy asparagine-linked glycosylation, the presence of a transmembrane domain in the carboxy-terminal cleavage product and protein expression in the apical membrane of enterocytes. In contrast, gCLCA2 was highly similar to mammalian CLCA2 in terms of its protein architecture, cleavage and glycosylation. These findings were in line with results from *in silico* analyses of CLCA2 sequences from other avian species.

Furthermore, the presence of a transmembrane domain in the carboxy-terminal cleavage product and its expression in keratinocytes are traits of avian CLCA2, which are also found in mammalian CLCA2. Interestingly, and as opposed to the expression patterns of mammalian CLCA proteins, no overlapping tissue or cellular expression patterns were detected for the two galline CLCA members.

Based on these findings, CLCA2 appears to be highly conserved among birds and mammals. The results allow to speculate that a hypothetical gene ancestor was expressed in keratinocytes of a common ancestral species before mammalian and avian lineages diverged. This high degree of CLCA2 conservation is in contrast to gCLCA1 and mammalian CLCA1, 3 and 4. During evolution, a hypothetical ancient ancestor of *gCLCA1* / mammalian *CLCA1, 3*, and *4*, was likely expressed by enterocytes of a common ancestral species of mammals and birds. The hypothetical ancestral gene seems to have expanded by gene duplications in the mammalian lineage, which did not occur in birds. Besides that, these findings cannot only be used to unveil the evolutionary history of the *CLCA* family but should be taken into account with regard to the selection of an animal model for the functional analysis of these genes. The chicken might serve as a promising species for knockout models to study CLCA2 functions *in vivo*. Results obtained from such a knockout chicken are likely transferable to mammalian CLCA2 due to the high degree of conservation. A chicken *gCLCA1* knockout model might provide data, which might be transferred most likely to mammalian *CLCA4* genes in the gut, as both share a similar cell type specific protein expression in this microenvironment, a similar protein architecture as well as similar biochemical properties.

At the transition to the post-genomic era with publically accessible information on gene structures as well as nucleotide and protein sequences of various species, comprehensive analyses of gene families across species have become possible. The comparison of such data in combination with the comparison of gene related information, including cellular expression patterns and biochemical properties, is a powerful approach to enlighten the evolutionary background of proteins. Furthermore, it might be beneficial to identify the most suitable animal model for further functional and biomedical studies.

Zusammenfassung

Evolutionäre Diversität der *CLCA* Gene zwischen Vögeln und Säugetieren

Florian Bartenschlager

Die *CLCA*, *engl. „chloride channel regulator, calcium-activated*“ Genfamilie wird hauptsächlich mit Krebserkrankungen sowie chronisch entzündlichen Atemwegserkrankungen in Zusammenhang gebracht. Die mutmaßlich komplexen Funktionen dieser Gene sind bisher jedoch noch weitgehend unbekannt. Die *CLCA* Genfamilie umfasst bei Säugetieren vier verschiedene Cluster, die als *CLCA1* bis *CLCA4* bezeichnet werden. Sie zeichnet sich durch eine außerordentliche Komplexität und Vielfältigkeit aus und beinhaltet tierartlich variierend amplifizierte und inaktivierte Gene. So besitzt der Mensch beispielsweise drei intakte *CLCA*s sowie ein inaktiviertes Gen, wohingegen die Maus über 8 CLCAs verfügt, von denen ebenfalls eines als inaktiviertes Gen vorliegt. Darüber hinaus sind die Gewebe- und Zellexpressionsmuster der CLCA-Homologen auch innerhalb einer Spezies häufig redundant. Diese Komplexität und Redundanz von CLCA könnten Gründe sein, welche die Aufdeckung der Funktion dieser Genfamilie im Säugetiermodell bisher erschwerte. Damit stellen sich zwei Fragen: Ist die Komplexität dieser Genfamilie eine Eigenart der Säugetiere? Und was ist der evolutionäre Hintergrund für diese Entwicklungen? Um diese Fragen zu beantworten, werden Daten über CLCA benötigt, die von evolutionär weiter entfernten Arten stammen. Kürzlich wurde für das Huhn ein relativ einfacher *CLCA* Genlokus vorhergesagt, welcher nur zwei *CLCA* Gene umfasst. Das erste CLCA Genprodukt des Huhns, gCLCA1, zeichnet sich durch eine besondere phylogenetische Nähe zu CLCA1, 3 und 4 der Säugetiere aus. Im Gegensatz dazu ist das zweite CLCA Genprodukt, gCLCA2, näher mit Säuger-CLCA2 verwandt als mit gCLCA1 oder Säuger-CLCA1, 3 oder 4.

In dieser kumulativen These wurden die gallinen *CLCA* Gene sowie deren genomische Struktur analysiert und beide Homologe wurden kloniert. Darüber hinaus wurde deren Proteinarchitektur und die biochemischen Eigenschaften *in silico* und *in vitro* untersucht. Weiterhin wurden die mRNA- und zellulären Proteinexpressionsmuster im Vergleich zu CLCA1, 2, 3 und 4 bei Säugetieren analysiert. Beide Vogelproteine werden von 14 Exons kodiert und befinden sich an einem konservierten Ort zwischen den *ODF2L, engl. „Outer Dense Fiber of Sperm Tails 2-Like”* und *SH3GLB1, engl „SH3-Domain GRB2-Like Endophilin B1“* Genen. gCLCA1 vereint mit seiner Proteinspaltung sowie der Proteinarchitektur viele Eigenschaften von CLCA1, 3 und 4 der Säugetiere. Trotz einer relativ großen phylogenetischen Entfernung zu Säuger-CLCA4 weist es jedoch die meisten gemeinsamen Merkmale mit diesem CLCA-Vertreter auf. Hierzu zählen eine starke, an Asparagin gekoppelte Glykosylierung, eine Transmembrandomäne im carboxyterminalen Spaltprodukt und eine

Proteinexpression in der apikalen Membran von Enterozyten. Im Gegensatz dazu ist gCLCA2 dem Säugetier-CLCA2 in Bezug auf dessen Phylogenie, Proteinarchitektur, Spaltung und Glykosylierung sehr ähnlich. Diese Befunde ließen sich mittels *in silico* Analysen auch für weitere aviäre CLCAs nachweisen. Daneben sind die Präsenz einer Transmembrandomäne im carboxyterminalen Spaltprodukt und die Expression in Keratinozyten Merkmale des aviären CLCA2, die auch bei Säugetier-CLCA2 vorzufinden sind. Bemerkenswerterweise fanden sich im Gegensatz zu den Säuger-CLCAs bei den gallinen CLCA-Mitgliedern keine überlappenden gewebe- und zellspezifischen Expressionsmuster.

Unter Betrachtung dieser Befunde erscheint CLCA2 bei Vögeln und Säugetieren stark konserviert zu sein. Weiterhin lässt sich auf Basis der in dieser kumulativen These erhobenen Daten spekulieren, dass ein hypothetischer gemeinsamer genetischer Vorfahre wahrscheinlich in Keratinozyten eines gemeinsamen Vorfahren von Vögeln und Säugern exprimiert wurde. Dieser hohe Grad an Konservierung von CLCA2 steht im Gegensatz zu demjenigen von gCLCA1 und Säugetier-CLCA1, 3 und 4. Im Laufe der Evolution scheint ein hypothetischer genetischer Vorfahre von gCLCA1/Säugetier-CLCA1, 3 und 4 vermutlich in Enterozyten eines gemeinsamen Vorfahren von Vögeln und Säugern exprimiert worden zu sein. Dieser hypothetische genetische Vorfahre scheint durch Genduplikationen bei Säugetieren expandiert zu sein, nicht jedoch bei Vögeln. Neben diesen Aussagen zu einem wahrscheinlichen evolutionären Szenario der *CLCA* Genfamilie zwischen Vogel und Säuger lassen sich die Ergebnisse jedoch auch zur Auswahl eines Modellorganismus zur funktionalen Analyse dieser Gene nutzen. Hierbei erscheint das Huhn zur Untersuchung der CLCA2-Funktion *in vivo* ein vielversprechender Kandidat für knockout-Modelle, deren Untersuchungsergebnisse durch den Grad an Konservierung mit hoher Wahrscheinlichkeit auf Säugetiere übertragbar sind. Ein *gCLCA1*-Knockoutmodell könnte hingegen Daten liefern, die hochwahrscheinlich auf *CLCA4* Gene von Säugetieren im Darm übertragbar sind, da beide in dieser anatomischen Lokalisation ein vergleichbares zellspezifisches Expressionsmuster sowie eine ähnliche Proteinarchitektur und biochemische Eigenschaften besitzen.

Im Anbruch des postgenomischen Zeitalters, in dem Genstrukturen sowie Nukleotid- und Proteinsequenzen verschiedener Spezies öffentlich leicht zugänglich sind, werden umfassende, vergleichende Analysen von Genfamilien über verschiedene Spezies hinweg möglich. In Kombination mit dem Vergleich von genbezogenen Daten wie zellulären Expressionsmustern oder biochemischen Eigenschaften der Genprodukte ist dies ein wirkungsvolles Vorgehen, um den evolutionären Hintergrund von Proteinen aufzudecken und das geeignetste Tiermodell für weitere wissenschaftliche Fragestellung auszuwählen.

References

Abdel-Ghany, M., H.-C. Cheng, R. C. Elble and B. U. Pauli (2001):

The breast cancer β4 integrin and endothelial human CLCA2 mediate lung metastasis.

Journal of Biological Chemistry, 276, 25438-25446. DOI: 10.1074/jbc.M100478200.

Abdel-Ghany, M., H.-C. Cheng, R. C. Elble and B. U. Pauli (2002):

Focal adhesion kinase activated by β4 integrin ligation to mCLCA1 mediates early metastatic growth.

Journal of Biological Chemistry, 277, 34391-34400. DOI: 10.1074/jbc.M205307200.

Agnel, M., T. Vermat and J.-M. Culouscou (1999):

Identification of three novel members of the calcium-dependent chloride channel (CaCC) family predominantly expressed in the digestive tract and trachea.

FEBS Letters, 455, 295-301. DOI: 10.1016/s0014-5793(99)00891-1.

Alevy, Y. G., A. C. Patel, A. G. Romero, D. A. Patel, J. Tucker, W. T. Roswit, C. A. Miller, R. F. Heier, D. E. Byers, T. J. Brett and M. J. Holtzman (2012):

IL-13-induced airway mucus production is attenuated by MAPK13 inhibition.

Journal of Clinical Investigation, 122, 4555-4568. DOI: 10.1172/JCI64896.

Altenhoff, A. M., R. A. Studer, M. Robinson-Rechavi and C. Dessimoz (2012):

Resolving the ortholog conjecture: orthologs tend to be weakly, but significantly, more similar in function than paralogs.

PLoS computational biology, 8, e1002514. DOI: 10.1371/journal.pcbi.1002514.

Anton, F., I. Leverkoehne, L. Mundhenk, W. B. Thoreson and A. D. Gruber (2005):

Overexpression of eCLCA1 in small airways of horses with recurrent airway obstruction.

Journal of Histochemistry & Cytochemistry, 53, 1011-1021. DOI: 10.1369/jhc.4A6599.2005.

Beckley, J. R., B. U. Pauli and R. C. Elble (2004):

Re-expression of detachment-inducible chloride channel mCLCA5 suppresses growth of metastatic breast cancer cells.

Journal of Biological Chemistry, 279, 41634-41641. DOI: 10.1074/jbc.M408334200.

Berry, K. N. and T. J. Brett (2020):

Structural and Biophysical Analysis of the CLCA1 VWA Domain Suggests Mode of TMEM16A Engagement.

Cell Reports, 30, 1141-1151. DOI: https://doi.org/10.1016/j.celrep.2019.12.059.

REFERENCES

Bothe, M. K., J. Braun, L. Mundhenk and A. D. Gruber (2008):

Murine mCLCA6 is an integral apical membrane protein of non-goblet cell enterocytes and co-localizes with the cystic fibrosis transmembrane conductance regulator.

Journal of Histochemistry & Cytochemistry, 56, 495-509. DOI: 10.1369/jhc.2008.950592.

Bothe, M. K., L. Mundhenk, M. Kaup, C. Weise and A. D. Gruber (2011):

The murine goblet cell protein mCLCA3 is a zinc-dependent metalloprotease with autoproteolytic activity.

Molecules and Cells, 32, 535-541. DOI: 10.1007/s10059-011-0158-8.

Braun, J., M. K. Bothe, L. Mundhenk, C. L. Beck and A. D. Gruber (2010):

Murine mCLCA5 is expressed in granular layer keratinocytes of stratified epithelia.

Histochemistry and Cell Biology, 133, 285-299. DOI: 10.1007/s00418-009-0667-0.

Bustin, S. A., S.-R. Li and S. Dorudi (2001):

Expression of the Ca^{2+}-activated chloride channel genes CLCA1 and CLCA2 is downregulated in human colorectal cancer.

DNA and Cell Biology, 20, 331-338. DOI: 10.1089/10445490152122442.

Chen, H., Y. Liu, C.-J. Jiang, Y.-M. Chen, H. Li and Q.-A. Liu (2019):

Calcium-activated chloride channel A4 (CLCA4) plays inhibitory roles in invasion and migration through suppressing epithelial-mesenchymal transition via PI3K/AKT signaling in colorectal cancer.

Medical Science Monitor: International Medical Journal of Experimental and Clinical Research, 25, 4176. DOI: 10.12659/MSM.914195.

Ching, J. C., L. Lobanova and M. E. Loewen (2013):

Secreted hCLCA1 is a signaling molecule that activates airway macrophages.

PloS One, 8, e83130. DOI: 10.1371/journal.pone.0083130.

Choi, M. R., H. D. Kim, S. Cho, S. H. Jeon, D. H. Kim, J. Wee and Y. D. Yang (2021):

Anoctamin1 Induces Hyperproliferation of HaCaT Keratinocytes and Triggers Imiquimod-Induced Psoriasis-Like Skin Injury in Mice.

International Journal of Molecular Sciences, 22, 7145. DOI: 10.3390/ijms22137145.

Chojnacka-Puchta, L. and D. Sawicka (2020):

CRISPR/Cas9 gene editing in a chicken model: current approaches and applications.

Journal of Applied Genetics, 61, 221-229. DOI: 10.1007/s13353-020-00537-9.

Clark, K., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell and E. W. Sayers (2016):

REFERENCES

GenBank.

Nucleic Acids Research, 44, D67-72. DOI: 10.1093/nar/gkv1276.

Connon, C. J., S. Kawasaki, M. Liles, N. Koizumi, K. Yamasaki, T. Nakamura, A. J. Quantock and S. Kinoshita (2006):

Gene expression and immunolocalisation of a calcium-activated chloride channel during the stratification of cultivated and developing corneal epithelium.

Cell and Tissue Research, 323, 177-182. DOI: 10.1007/s00441-005-0059-2.

Connon, C. J., K. Yamasaki, S. Kawasaki, A. J. Quantock, N. Koizumi and S. Kinoshita (2004):

Calcium-activated chloride channel-2 in human epithelia.

Journal of Histochemistry & Cytochemistry, 52, 415-418. DOI: 10.1177/002215540405200313.

Corren, J. (2013):

Role of interleukin-13 in asthma.

Current Allergy and Asthma Reports, 13, 415-420. DOI: 10.1007/s11882-013-0373-9.

Cunningham, F., J. E. Allen, J. Allen, J. Alvarez-Jarreta, M R. Amode, Irina M. Armean, O. Austine-Orimoloye, Andrey G. Azov, I. Barnes, R. Bennett, A. Berry, J. Bhai, A. Bignell, K. Billis, S. Boddu, L. Brooks, M. Charkhchi, C. Cummins, L. Da Rin Fioretto, C. Davidson, K. Dodiya, S. Donaldson, B. El Houdaigui, T. El Naboulsi, R. Fatima, C. G. Giron, T. Genez, Jose G. Martinez, C. Guijarro-Clarke, A. Gymer, M. Hardy, Z. Hollis, T. Hourlier, T. Hunt, T. Juettemann, V. Kaikala, M. Kay, I. Lavidas, T. Le, D. Lemos, J. C. Marugán, S. Mohanan, A. Mushtaq, M. Naven, Denye N. Ogeh, A. Parker, A. Parton, M. Perry, I. Piližota, I. Prosovetskaia, Manoj P. Sakthivel, Ahamed Imran A. Salam, Bianca M. Schmitt, H. Schuilenburg, D. Sheppard, José G. Pérez-Silva, W. Stark, E. Steed, K. Sutinen, R. Sukumaran, D. Sumathipala, M.-M. Suner, M. Szpak, A. Thormann, F. F. Tricomi, D. Urbina-Gómez, A. Veidenberg, Thomas A. Walsh, B. Walts, N. Willhoft, A. Winterbottom, E. Wass, M. Chakiachvili, B. Flint, A. Frankish, S. Giorgetti, L. Haggerty, Sarah E. Hunt, Garth R. Iisley, Jane E. Loveland, Fergal J. Martin, B. Moore, Jonathan M. Mudge, M. Muffato, E. Perry, M. Ruffier, J. Tate, D. Thybert, Stephen J. Trevanion, S. Dyer, Peter W. Harrison, Kevin L. Howe, Andrew D. Yates, Daniel R. Zerbino and P. Flicek (2022):

Ensembl 2022.

Nucleic Acids Research, 50, D988-D995. DOI: 10.1093/nar/gkab1049.

Cunningham, S. A., M. S. Awayda, J. K. Bubien, Ismailov, II, M. P. Arrate, B. K. Berdiev, D. J. Benos and C. M. Fuller (1995):

Cloning of an epithelial chloride channel from bovine trachea.

Journal of Biological Chemistry, 270, 31016-31026. DOI: 10.1074/jbc.270.52.31016.

Czelusniak, J., M. Goodman, N. D. Moncrief and S. M. Kehoe (1990):

Maximum parsimony approach to construction of evolutionary trees from aligned homologous sequences.

Methods in Enzymology, 183, 601-615. DOI: 10.1016/0076-6879(90)83039-c.

Dietert, K., L. Mundhenk, N. A. Erickson, K. Reppe, A. C. Hocke, W. Kummer, M. Witzenrath and A. D. Gruber (2015):

Murine CLCA5 is uniquely expressed in distinct niches of airway epithelial cells.

Histochemistry and Cell Biology, 143, 277-287. DOI: 10.1007/s00418-014-1279-x.

Dietert, K., K. Reppe, L. Mundhenk, M. Witzenrath and A. D. Gruber (2014):

mCLCA3 modulates IL-17 and CXCL-1 induction and leukocyte recruitment in murine *Staphylococcus aureus* pneumonia.

PloS One, 9, e102606. DOI: 10.1371/journal.pone.0102606.

Dolinski, K. and D. Botstein (2007):

Orthology and functional conservation in eukaryotes.

Annual Review of Genetics, 41, 465-507. DOI: 10.1146/annurev.genet.40.110405.090439.

Elble, R. C., G. Ji, K. Nehrke, J. DeBiasio, P. D. Kingsley, M. I. Kotlikoff and B. U. Pauli (2002):

Molecular and functional characterization of a murine calcium-activated chloride channel expressed in smooth muscle.

Journal of Biological Chemistry, 277, 18586-18591. DOI: 10.1074/jbc.M200829200.

Elble, R. C., V. Walia, H.-c. Cheng, C. J. Connon, L. Mundhenk, A. D. Gruber and B. U. Pauli (2006):

The putative chloride channel hCLCA2 has a single C-terminal transmembrane segment.

Journal of Biological Chemistry, 281, 29448-29454. DOI: 10.1074/jbc.M605919200.

Elble, R. C., J. Widom, A. D. Gruber, M. Abdel-Ghany, R. Levine, A. Goodwin, H. C. Cheng and B. U. Pauli (1997):

Cloning and characterization of lung-endothelial cell adhesion molecule-1 suggest it is an endothelial chloride channel.

Journal of Biological Chemistry, 272, 27853-27861. DOI: 10.1074/jbc.272.44.27853.

Erickson, N., A. Gruber and L. Mundhenk (2020):

The family of chloride channel regulator, calcium-activated proteins in the feline respiratory tract: a comparative perspective on airway Diseases in man and animal models.

Journal of Comparative Pathology, 174, 39-53. DOI: 10.1016/j.jcpa.2019.10.193.

Erickson, N. A., K. Dietert, J. Enders, R. Glauben, G. Nouailles, A. D. Gruber and L. Mundhenk (2018):

REFERENCES

Soluble mucus component CLCA1 modulates expression of leukotactic cytokines and BPIFA1 in murine alveolar macrophages but not in bone marrow-derived macrophages.

Histochemistry and Cell Biology, 149, 619-633. DOI: 10.1007/s00418-018-1664-y.

Erickson, N. A., E. E. Nyström, L. Mundhenk, L. Arike, R. Glauben, M. M. Heimesaat, A. Fischer, S. Bereswill, G. M. Birchenough and A. D. Gruber (2015):

The goblet cell protein Clca1 (alias mClca3 or Gob-5) is not required for intestinal mucus synthesis, structure and barrier function in naive or DSS-challenged mice.

PLoS One, 10, e0131991. DOI: 10.1371/journal.pone.0131991.

Evans, C. M., D. S. Raclawska, F. Ttofali, D. R. Liptzin, A. A. Fletcher, D. N. Harper, M. A. McGing, M. M. McElwee, O. W. Williams, E. Sanchez, M. G. Roy, K. N. Kindrachuk, T. A. Wynn, H. K. Eltzschig, M. R. Blackburn, M. J. Tuvim, W. J. Janssen, D. A. Schwartz and B. F. Dickey (2015):

The polymeric mucin Muc5ac is required for allergic airway hyperreactivity.

Nature Communications, 6, 6281. DOI: 10.1038/ncomms7281.

Evans, S. R., W. B. Thoreson and C. L. Beck (2004):

Molecular and functional analyses of two new calcium-activated chloride channel family members from mouse eye and intestine.

The Journal of Biological Chemistry, 279, 41792-41800. DOI: 10.1074/jbc.M408354200.

Fang, G., N. Bhardwaj, R. Robilotto and M. B. Gerstein (2010):

Getting started in gene orthology and functional analysis.

PLoS Computational Biology, 6, e1000703. DOI: 10.1371/journal.pcbi.1000703.

Felsenstein, J. (1981):

Evolutionary trees from DNA sequences: a maximum likelihood approach.

Journal of Molecular Evolution, 17, 368-376. DOI: 10.1007/BF01734359.

Felsenstein, J. (1984):

Distance Methods for Inferring Phylogenies: A Justification.

Evolution, 38, 16-24. DOI: 10.1111/j.1558-5646.1984.tb00255.x.

Fitch, W. M. (1970):

Distinguishing Homologous from Analogous Proteins.

Systematic Biology, 19, 99-113. DOI: 10.2307/2412448.

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan and J. Postlethwait (1999):

Preservation of duplicate genes by complementary, degenerative mutations.

Genetics, 151, 1531-1545. DOI: 10.1093/genetics/151.4.1531.

Francino, M. P. (2005):

An adaptive radiation model for the origin of new gene functions.

Nature Genetics, 37, 573-577. DOI: 10.1038/ng1579.

Furuya, M., S. B. Kirschbaum, A. Paulovich, B. U. Pauli, H. Zhang, J. S. Alexander, A. G. Farr and A. Ruddell (2010):

Lymphatic endothelial murine chloride channel calcium-activated 1 is a ligand for leukocyte LFA-1 and Mac-1.

The Journal of Immunology, 185, 5769-5777. DOI: 10.4049/jimmunol.1002226.

Gabaldón, T. and E. V. Koonin (2013):

Functional and evolutionary implications of gene orthology.

Nature Reviews. Genetics, 14, 360-366. DOI: 10.1038/nrg3456.

Gandhi, R., R. Elble, A. Gruber, K. Schreur, H.-L. Ji, C. Fuller and B. Pauli (1998):

Molecular and functional characterization of a calcium-sensitive chloride channel from mouse lung.

Journal of Biological Chemistry, 273, 32096-32101. DOI: 10.1074/jbc.273.48.32096.

Gaspar, K., K. Racette, J. Gordon, M. Loewen and G. Forsyth (2000):

Cloning a chloride conductance mediator from the apical membrane of porcine ileal enterocytes.

Physiological Genomics, 3, 101-111. DOI: 10.1152/physiolgenomics.2000.3.2.101.

Gerber, V., H. De Feijter-Rupp, J. Wagner, P. Venta, J. Harkema and N. Robinson (2009):

Differential association of MUC5AC and CLCA1 expression in small cartilaginous airways of RAO-affected and control horses.

Equine Veterinary Journal, 41, 817-823. DOI: 10.2746/042516409x443305.

Gibb, A. A., M. P. Lazaropoulos and J. W. Elrod (2020):

Myofibroblasts and Fibrosis: Mitochondrial and Metabolic Control of Cellular Differentiation.

Circulation Research, 127, 427-447. DOI: 10.1161/CIRCRESAHA.120.316958.

Gibson, A., A. P. Lewis, K. Affleck, A. J. Aitken, E. Meldrum and N. Thompson (2005):

hCLCA1 and mCLCA3 are secreted non-integral membrane proteins and therefore are not ion channels.

Journal of Biological Chemistry, 280, 27205-27212. DOI: 10.1074/jbc.M504654200.

Gonzalez-Silva, C., J. Vera, M. R. Bono, C. González-Billault, B. Baxter, A. Hansen, R. Lopez, E. A. Gibson, D. Restrepo and J. Bacigalupo (2013):

Ca^{2+}-activated Cl- channels of the ClCa family express in the cilia of a subset of rat olfactory sensory neurons.

PLoS One, 8, e69295. DOI: 10.1371/journal.pone.0069295.

Gruber, A. D., R. C. Elble, H.-L. Ji, K. D. Schreur, C. M. Fuller and B. U. Pauli (1998):

Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca^{2+}-activated Cl− channel proteins.

Genomics, 54, 200-214. DOI: 10.1006/geno.1998.5562.

Gruber, A. D. and B. U. Pauli (1999):

Tumorigenicity of human breast cancer is associated with loss of the Ca2+-activated chloride channel CLCA2.

Cancer Research 59: 5488-5491.

URL: https://aacrjournals.org/cancerres/article/59/21/5488/505622/Tumorigenicity-of-Human-Breast-Cancer-Is.

Gruber, A. D., K. D. Schreur, H.-L. Ji, C. M. Fuller and B. U. Pauli (1999):

Molecular cloning and transmembrane structure of hCLCA2 from human lung, trachea, and mammary gland.

American Journal of Physiology, 276, C1261-C1270. DOI: 10.1152/ajpcell.1999.276.6.C1261.

Hahn, M. W. (2009):

Distinguishing Among Evolutionary Models for the Maintenance of Gene Duplicates.

The Journal of Heredity, 100, 605-617. DOI: 10.1093/jhered/esp047.

Hamalainen, L., G. Bart, P. Takabe, L. Rauhala, A. Deen, S. Pasonen-Seppanen, E. Karkkainen, R. Karna, T. Kumlin, M. I. Tammi and R. H. Tammi (2021):

The calcium-activated chloride channel-associated protein rCLCA2 is expressed throughout rat epidermis, facilitates apoptosis and is downmodulated by UVB.

Histochemistry and Cell Biology, 155, 605-615. DOI: 10.1007/s00418-021-01962-5.

Hautmann, M. (2020):

What is macroevolution?

Palaeontology 63: 1-11. DOI: doi.org/10.1111/pala.12465.

Hashimoto, K., B. S. Graham, S. B. Ho, K. B. Adler, R. D. Collins, S. J. Olson, W. Zhou, T. Suzutani, P. W. Jones and K. Goleniewska (2004):

REFERENCES

Respiratory syncytial virus in allergic lung inflammation increases Muc5ac and gob-5.

American Journal of Respiratory and Critical Care Medicine, 170, 306-312. DOI: 10.1164/rccm.200301-030OC.

Hegab, A., T. Sakamoto, Y. Uchida, A. Nomura, Y. Ishii, Y. Morishima, M. Mochizuki, T. Kimura, W. Saitoh and H. Massoud (2004):

CLCA1 gene polymorphisms in chronic obstructive pulmonary disease.

Journal of Medical Genetics, 41, e27-e27. DOI: 10.1136/jmg.2003.012484.

Hoshino, M., S. Morita, H. Iwashita, Y. Sagiya, T. Nagi, A. Nakanishi, Y. Ashida, O. Nishimura, Y. Fujisawa and M. Fujino (2002):

Increased expression of the human Ca^{2+}-activated Cl− channel 1 (*CaCC1*) gene in the asthmatic airway.

American Journal of Respiratory and Critical Care Medicine, 165, 1132-1136. DOI: 10.1164/ajrccm.165.8.2107068.

Hou, T., L. Zhou, L. Wang, G. Kazobinka, X. Zhang and Z. Chen (2017):

CLCA4 inhibits bladder cancer cell proliferation, migration, and invasion by suppressing the PI3K/AKT pathway.

Oncotarget, 8, 93001-93013. DOI: 10.18632/oncotarget.21724.

Hu, D., D. Ansari, Q. Zhou, A. Sasor, K. S. Hilmersson, M. Bauden, Y. Jiang and R. Andersson (2018):

Calcium-activated chloride channel regulator 1 as a prognostic biomarker in pancreatic ductal adenocarcinoma.

BMC Cancer, 18, 1096. DOI: 10.1186/s12885-018-5013-2.

Hu, D., D. Ansari, M. Bauden, Q. Zhou and R. Andersson (2019):

The Emerging Role of Calcium-activated Chloride Channel Regulator 1 in Cancer.

Anticancer Research, 39, 1661-1666. DOI: 10.21873/anticanres.13271.

Huan, C., K. S. Greene, B. Shui, G. Spizz, H. Sun, R. M. Doran, P. J. Fisher, M. S. Roberson, R. C. Elble and M. I. Kotlikoff (2008):

mCLCA4 ER processing and secretion requires luminal sorting motifs.

American Journal of Physiology, 295, C279-287. DOI: 10.1152/ajpcell.00060.2008.

Huelsenbeck, J. P., B. Larget, R. E. Miller and F. Ronquist (2002):

Potential applications and pitfalls of Bayesian inference of phylogeny.

Systematic Biology, 51, 673-688. DOI: 10.1080/10635150290102366.

REFERENCES

Hughes, A. L. (1994):

The evolution of functionally novel proteins after gene duplication.

Proceedings of the Royal Scociety. Biological Sciences, 256, 119-124. DOI: 10.1098/rspb.1994.0058.

Hughes, K., M. Blanck, S. Pensa and C. J. Watson (2016):

Stat3 modulates chloride channel accessory protein expression in normal and neoplastic mammary tissue.

Cell Death & Disease, 7, e2398. DOI: 10.1038/cddis.2016.302.

Innan, H. and F. Kondrashov (2010):

The evolution of gene duplications: classifying and distinguishing between models.

Nature Reviews Genetics, 11, 97-108. DOI: 10.1038/nrg2689.

Ishibashi, K., J. Yamazaki, K. Okamura, Y. Teng, K. Kitamura and K. Abe (2006):

Roles of CLCA and CFTR in electrolyte re-absorption from rat saliva.

Journal of Dental Research, 85, 1101-1105. DOI: 10.1177/154405910608501207.

Jeong, S. M., H. K. Park, I. S. Yoon, J. H. Lee, J. H. Kim, C. G. Jang, C. J. Lee and S. Y. Nah (2005):

Cloning and expression of Ca^{2+}-activated chloride channel from rat brain.

Biochemical and Biophysical Research Communications 334: 569-576. DOI: 10.1016/j.bbrc.2005.06.122.

Kaessmann, H., N. Vinckenbosch and M. Long (2009):

RNA-based gene duplication: mechanistic and evolutionary insights.

Nature Reviews. Genetics, 10, 19-31. DOI: 10.1038/nrg2487.

Kamada, F., Y. Suzuki, C. Shao, M. Tamari, K. Hasegawa, T. Hirota, M. Shimizu, N. Takahashi, X. Mao and S. Doi (2004):

Association of the hCLCA1 gene with childhood and adult asthma.

Genes & Immunity, 5, 540-547. DOI: 10.1038/sj.gene.6364124.

Keith, B. A., J. C. Ching and M. E. Loewen (2019):

Von Willebrand Factor Type A domain of hCLCA1 is sufficient for U-937 macrophage activation.

Biochemistry and Biophysics Reports, 18, 100630. DOI: 10.1016/j.bbrep.2019.100630.

Kellis, M., B. W. Birren and E. S. Lander (2004):

Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*.

Nature, 428, 617-624. DOI: 10.1038/nature02424.

Kent, W. J., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler and D. Haussler (2002):

The human genome browser at UCSC.

Genome Research, 12, 996-1006. DOI: 10.1101/gr.229102.

Komiya, T., Y. Tanigawa and S. Hirohashi (1999):

Cloning and identification of the gene gob-5, which is expressed in intestinal goblet cells in mice.

Biochemical and Biophysical Research Communications, 255, 347-351. DOI: 10.1006/bbrc.1999.0168.

Kumar, S., M. Suleski, J. M. Craig, A. E. Kasprowicz, M. Sanderford, M. Li, G. Stecher and S. B. Hedges (2022):

TimeTree 5: An Expanded Resource for Species Divergence Times.

Molecular Biology and Evolution, 39, msac174. DOI: 10.1093/molbev/msac174.

Kuperman, D. A., X. Huang, L. L. Koth, G. H. Chang, G. M. Dolganov, Z. Zhu, J. A. Elias, D. Sheppard and D. J. Erle (2002):

Direct effects of interleukin-13 on epithelial cells cause airway hyperreactivity and mucus overproduction in asthma.

Nature Medicine, 8, 885-889. DOI: 10.1038/nm734.

Kuperman, D. A., C. C. Lewis, P. G. Woodruff, M. W. Rodriguez, Y. H. Yang, G. M. Dolganov, J. V. Fahy and D. J. Erle (2005):

Dissecting asthma using focused transgenic modeling and functional genomics.

Journal of Allergy and Clinical Immunology, 116, 305-311. DOI: 10.1016/j.jaci.2005.03.024.

Lee, C. (2013):

Chapter 10 - Structural Genomic Variation in the Human Genome.

In: Genomic and Personalized Medicine (Second Edition). G. S. Ginsburg and H. F. Willard, Academic Press: 123-132. DOI: 10.1056/NEJMcibr0708452.

Lee, D., S. Ha, Y. Kho, J. Kim, K. Cho, M. Baik and Y. Choi (1999):

Induction of mouse Ca^{2+}-sensitive chloride channel 2 gene during involution of mammary gland.

Biochemical and Biophysical Research Communications, 264, 933-937. DOI: 10.1006/bbrc.1999.1583.

REFERENCES

Lee, R. M., K. H. Han and J. S. Han (2016):

rbCLCA1 is a putative metalloprotease family member: localization and catalytic domain identification.

Amino Acids, 48, 707-720. DOI: 10.1007/s00726-015-2119-6.

Lee, R. M., R. H. Ryu, S. W. Jeong, S. J. Oh, H. Huang, J. S. Han, C. H. Lee, C. J. Lee, L. Y. Jan and S. M. Jeong (2011):

Isolation and Expression Profile of the Ca-Activated Chloride Channel-like Membrane Protein 6 Gene in *Xenopus laevis*.

Laboratory Animal Research, 27, 109-116. DOI: 10.5625/lar.2011.27.2.109.

Leggatt, R. A. and G. K. Iwama (2003):

Occurrence of polyploidy in the fishes.

Reviews in Fish Biology and Fisheries, 13, 237-246. DOI: 10.1023/B:RFBF.0000033049.00668.fe.

Lenart, A., M. Dudkiewicz, M. Grynberg and K. Pawłowski (2013):

CLCAs-a family of metalloproteases of intriguing phylogenetic distribution and with cases of substituted catalytic sites.

PLoS One, 8, e62272. DOI: 10.1371/journal.pone.0062272.

Leverkoehne, I. and A. D. Gruber (2002):

The murine mCLCA3 (alias gob-5) protein is located in the mucin granule membranes of intestinal, respiratory, and uterine goblet cells.

Journal of Histochemistry & Cytochemistry, 50, 829-838. DOI: 10.1177/002215540205000609.

Li, B., Y. P. Jiang, J. Zhu and L. Meng (2020):

MiR-501-5p acts as an energetic regulator in head and neck squamous cell carcinoma cells growth and aggressiveness via reducing CLCA4.

Molecular Biology Reports, 47, 2181-2187. DOI: 10.1007/s11033-020-05317-6.

Li, X., J. K. Cowell and K. Sossey-Alaoui (2004):

CLCA2 tumour suppressor gene in 1p31 is epigenetically regulated in breast cancer.

Oncogene, 23, 1474-1480. DOI: 10.1038/sj.onc.1207249.

Li, X., W. Hu, J. Zhou, Y. Huang, J. Peng, Y. Yuan, J. Yu and S. Zheng (2017):

CLCA1 suppresses colorectal cancer aggressiveness via inhibition of the Wnt/beta-catenin signaling pathway.

Cell Communication and Signalling, 15, 38. DOI: 10.1186/s12964-017-0192-z.

Lieber, M. R. (2010):

The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway.

Annual Review of Biochemistry, 79, 181-211. DOI: 10.1146/annurev.biochem.052308.093131.

Liu, C. L. and G. P. Shi (2019):

Calcium-activated chloride channel regulator 1 (CLCA1): More than a regulator of chloride transport and mucus production.

World Allergy Organization Journal, 12, 100077. DOI: 10.1016/j.waojou.2019.100077.

Liu, Z., M. Chen, L. K. Xie, T. Liu, Z. W. Zou, Y. Li, P. Chen, X. Peng, C. Ma, W. J. Zhang and P. D. Li (2018):

CLCA4 inhibits cell proliferation and invasion of hepatocellular carcinoma by suppressing epithelial-mesenchymal transition via PI3K/AKT signaling.

Aging, 10, 2570-2584. DOI: 10.18632/aging.101571.

Long, A. J., J. P. Sypek, R. Askew, S. C. Fish, L. E. Mason, C. M. Williams and S. J. Goldman (2006):

Gob-5 contributes to goblet cell hyperplasia and modulates pulmonary tissue inflammation.

American Journal of Respiratory Cell and Molecular Biology, 35, 357-365. DOI: 10.1165/rcmb.2005-0451OC.

Long, M., E. Betrán, K. Thornton and W. Wang (2003):

The origin of new genes: glimpses from the young and old.

Nature Reviews Genetics, 4, 865-875. DOI: 10.1038/nrg1204.

MABLE, B. K. (2004):

'Why polyploidy is rarer in animals than in plants': myths and mechanisms.

Biological Journal of the Linnean Society, 82, 453-466. DOI: 10.1111/j.1095-8312.2004.00332.x.

Magadum, S., U. Banerjee, P. Murugan, D. Gangapur and R. Ravikesavan (2013):

Gene duplication as a major force in evolution.

Journal of Genetics, 92, 155-161. DOI: 10.1007/s12041-013-0212-8.

Mei, L., L. He, S. S. Wu, B. Zhang, Y. J. Xu, Z. X. Zhang, J. P. Zhao and H. L. Zhang (2013):

Murine calcium-activated chloride channel family member 3 induces asthmatic airway inflammation independently of allergen exposure.

Chinese Medical Journal, 126, 3283-3288. DOI: 10.3760/cma.j.issn.0366-6999.20130620.

REFERENCES

Mo, H. Y., J. H. Lee, M. S. Kim, N. J. Yoo and S. H. Lee (2020):

Frameshift Mutations and Loss of Expression of CLCA4 Gene are Frequent in Colorectal Cancers With Microsatellite Instability.

Applied Immunohistochemistry & Molecular Morphology, 28, 489-494. DOI: 10.1097/PAI.0000000000000777.

Morin, V., N. Véron and C. Marcelle (2017):

CRISPR/Cas9 in the chicken embryo.

In: Avian and Reptilian Developmental Biology. Methods in Molecular Biology/Editor: Guojun Sheng, Vol 1650, P. 113-123

New York: Humana Press - Online ISBN 978-1-4939-7216-6. DOI: 10.1007/978-1-4939-7216-6_7.

Mundhenk, L., M. Alfalah, R. C. Elble, B. U. Pauli, H. Y. Naim and A. D. Gruber (2006):

Both cleavage products of the mCLCA3 protein are secreted soluble proteins.

Journal of Biological Chemistry, 281, 30072-30080. DOI: 10.1074/jbc.m606489200.

Mundhenk, L., N. A. Erickson, N. Klymiuk and A. D. Gruber (2018):

Interspecies diversity of chloride channel regulators, calcium-activated 3 genes.

PloS One, 13, e0191512. DOI: 10.1371/journal.pone.0191512.

Mundhenk, L., B. Johannesson, P. Anagnostopoulou, J. Braun, M. K. Bothe, C. Schultz, M. A. Mall and A. D. Gruber (2012):

mCLCA3 does not contribute to calcium-activated chloride conductance in murine airways.

American Journal of Respiratory Cell and Molecular Biology, 47, 87-93. DOI: 10.1165/rcmb.2010-0508OC.

Mura, C. V., R. Delgado, M. G. Delgado, D. Restrepo and J. Bacigalupo (2017):

A CLCA regulatory protein present in the chemosensory cilia of olfactory sensory neurons induces a Ca($^{2+}$)-activated Cl(-) current when transfected into HEK293.

BMC Neuroscience, 18, 61. DOI: DOI: 10.1186/s12868-017-0379-7.

Musrap, N., A. Tuccitto, G. S. Karagiannis, P. Saraon, I. Batruch and E. P. Diamandis (2015):

Comparative Proteomics of Ovarian Cancer Aggregate Formation Reveals an Increased Expression of Calcium-activated Chloride Channel Regulator 1 (CLCA1).

Journal of Biological Chemistry, 290, 17218-17227. DOI: 10.1074/jbc.M115.639773.

Nakanishi, A., S. Morita, H. Iwashita, Y. Sagiya, Y. Ashida, H. Shirafuji, Y. Fujisawa, O. Nishimura and M. Fujino (2001):

REFERENCES

Role of gob-5 in mucus overproduction and airway hyperresponsiveness in asthma.

Proceedings of the National Academy of Sciences of the United States of America, 98, 5175-5180. DOI: 10.1073/pnas.081510898.

Nakano, T., H. Inoue, S. Fukuyama, K. Matsumoto, M. Matsumura, M. Tsuda, T. Matsumoto, H. Aizawa and Y. Nakanishi (2006):

Niflumic acid suppresses interleukin-13-induced asthma phenotypes.

American Journal of Respiratory and Critical Care Medicine, 173, 1216-1221. DOI: 10.1164/rccm.200410-1420OC.

Namekawa, S. H. and J. T. Lee (2009):

XY and ZW: Is Meiotic Sex Chromosome Inactivation the Rule in Evolution?

PLOS Genetics, 5, e1000493. DOI: 10.1371/journal.pgen.1000493.

Nehrt, N. L., W. T. Clark, P. Radivojac and M. W. Hahn (2011):

Testing the Ortholog Conjecture with Comparative Functional Genomic Data from Mammals.

PLOS Computational Biology, 7, e1002073. DOI: 10.1371/journal.pcbi.1002073.

Nyström, E. E., L. Arike, E. Ehrencrona, G. C. Hansson and M. E. Johansson (2019):

Calcium-activated chloride channel regulator 1 (CLCA1) forms non-covalent oligomers in colonic mucus and has mucin 2–processing properties.

Journal of Biological Chemistry, 294, 17075-17089. DOI: 10.1074/jbc.RA119.009940.

Nyström, E. E., G. M. Birchenough, S. van der Post, L. Arike, A. D. Gruber, G. C. Hansson and M. E. Johansson (2018):

Calcium-activated chloride channel regulator 1 (CLCA1) controls mucus expansion in colon by proteolytic activity.

EBioMedicine 33: 134-143. DOI: 10.1016/j.ebiom.2018.05.031.

Ohno, S. (1970):

Evolution by gene duplication.

1st. ed., Berlin, Heidelberg:

Springer Spektrum

ISBN: 978-3-642-86659-3. DOI: 10.1007/978-3-642-86659-3.

Pan, X., Q. Wang, C. Xu, L. Yan, S. Pang and J. Gan (2019):

Prognostic value of chloride channel accessory mRNA expression in colon cancer.

Oncology letters, 18, 2967-2976. DOI: 10.3892/ol.2019.10615.

Patel, A. C., T. J. Brett and M. J. Holtzman (2009):

The role of CLCA proteins in inflammatory airway disease.

Annual Review of Physiology, 71, 425-449. DOI: 10.1146/annurev.physiol.010908.163253.

Patel, A. C., J. D. Morton, E. Y. Kim, Y. Alevy, S. Swanson, J. Tucker, G. Huang, E. Agapov, T. E. Phillips and M. E. Fuentes (2006):

Genetic segregation of airway disease traits despite redundancy of calcium-activated chloride channel family members.

Physiological Genomics, 25, 502-513. DOI: 10.1152/physiolgenomics.00321.2005.

Pawłowski, K., M. Lepistö, N. Meinander, U. Sivars, M. Varga and E. Wieslander (2006):

Novel conserved hydrolase domain in the CLCA family of alleged calcium-activated chloride channels.

Proteins: Structure, Function, and Bioinformatics, 63, 424-439. DOI: 10.1002/prot.20887.

Plog, S., L. Mundhenk, N. Klymiuk and A. D. Gruber (2009):

Genomic, tissue expression, and protein characterization of pCLCA1, a putative modulator of cystic fibrosis in the pig.

Journal of Histochemistry & Cytochemistry, 57, 1169-1181. DOI: 10.1369/jhc.2009.954594.

Plog, S., T. Grötzsch, N. Klymiuk, U. Kobalz, A. D. Gruber and L. Mundhenk (2012a):

The porcine chloride channel calcium-activated family member pCLCA4a mirrors lung expression of the human hCLCA4.

Journal of Histochemistry & Cytochemistry, 60, 45-56. DOI: 10.1369/0022155411426455.

Plog, S., L. Mundhenk, L. Langbein and A. D. Gruber (2012b):

Synthesis of porcine pCLCA2 protein during late differentiation of keratinocytes of epidermis and hair follicle inner root sheath.

Cell and Tissue Research, 350, 445-453. DOI: 10.1007/s00441-012-1482-9.

Plog, S., N. Klymiuk, S. Binder, M. J. Van Hook, W. B. Thoreson, A. D. Gruber and L. Mundhenk (2015):

Naturally Occurring Deletion Mutants of the Pig-Specific, Intestinal Crypt Epithelial Cell Protein CLCA4b without Apparent Phenotype.

PloS One, 10, e0140050. DOI: 10.1371/journal.pone.0140050.

Porretti, J., G. N. Dalton, C. Massillo, G. D. Scalise, P. L. Farré, R. Elble, E. N. Gerez, P. Accialini, A. M. Cabanillas, K. Gardner, P. De Luca and A. De Siervi (2018):

CLCA2 epigenetic regulation by CTBP1, HDACs, ZEB1, EP300 and miR-196b-5p impacts prostate cancer cell adhesion and EMT in metabolic syndrome disease.

International Journal of Cancer, 143, 897-906. DOI: 10.1002/ijc.31379.

Proulx, S. R. and P. C. Phillips (2006):

Allelic divergence precedes and promotes gene duplication.

Evolution, 60, 881-892. DOI: 10.1111/j.0014-3820.2006.tb01168.x.

Qiang, Y. Y., C. Z. Li, R. Sun, L. S. Zheng, L. X. Peng, J. P. Yang, D. F. Meng, Y. H. Lang, Y. Mei, P. Xie, L. Xu, Y. Cao, W. W. Wei, L. Cao, H. Hu, Q. Yang, D. H. Luo, Y. Y. Liang, B. J. Huang and C. N. Qian (2018):

Along with its favorable prognostic role, CLCA2 inhibits growth and metastasis of nasopharyngeal carcinoma cells via inhibition of FAK/ERK signaling.

Journal of Experimental & Clinical Cancer Research, 37, 34. DOI: 10.1186/s13046-018-0692-8.

Qin, Y., Y. Jiang, A. S. Sheikh, S. Shen, J. Liu and D. Jiang (2016):

Interleukin-13 stimulates MUC5AC expression via a STAT6-TMEM16A-ERK1/2 pathway in human airway epithelial cells.

International Immunopharmacology, 40, 106-114. DOI: 10.1016/j.intimp.2016.08.033.

Ran, S. and D. Benos (1991):

Isolation and functional reconstitution of a 38-kDa chloride channel protein from bovine tracheal membranes.

Journal of Biological Chemistry, 266, 4782-4788. DOI: 10.1016/S0021-9258(19)67717-3.

Range, F., L. Mundhenk and A. Gruber (2007):

A soluble secreted glycoprotein (eCLCA1) is overexpressed due to goblet cell hyperplasia and metaplasia in horses with recurrent airway obstruction.

Veterinary Pathology, 44, 901-911. DOI: 10.1354/vp.44-6-901.

Ratjen, F., S. C. Bell, S. M. Rowe, C. H. Goss, A. L. Quittner and A. Bush (2015):

Cystic fibrosis.

Nature Reviews. Disease Primers, 1, 15010. DOI: 10.1038/nrdp.2015.10.

Rice, A., P. Šmarda, M. Novosolov, M. Drori, L. Glick, N. Sabath, S. Meiri, J. Belmaker and I. Mayrose (2019):

The global biogeography of polyploid plants.

Nature Ecology & Evolution, 3, 265-273. DOI: 10.1038/s41559-018-0787-9.

Ritzka, M., F. Stanke, S. Jansen, A. D. Gruber, L. Pusch, S. Woelfl, H. J. Veeze, D. J. Halley and B. Tümmler (2004):

REFERENCES

The CLCA gene locus as a modulator of the gastrointestinal basic defect in cystic fibrosis.

Human Genetics, 115, 483-491. DOI: 10.1007/s00439-004-1190-y.

Robichaud, A., S. A. Tuck, S. Kargman, J. Tam, E. Wong, M. Abramovitz, J. Mortimer, H. E. Burston, P. Masson and J. Hirota (2005):

Gob-5 is not essential for mucus overproduction in preclinical murine models of allergic asthma.

American Journal of Respiratory Cell and Molecular Biology, 33, 303-314. DOI: 10.1165/rcmb.2004-0372OC.

Rogers, D. F. (2000):

Mucus pathophysiology in COPD: differences to asthma, and pharmacotherapy.

Monaldi Archives for Chest Disease, 55, 324-332.

Roussa, E., P. Wittschen, N. A. Wolff, B. Torchalski, A. D. Gruber and F. Thévenod (2010):

Cellular distribution and subcellular localization of mCLCA1/2 in murine gastrointestinal epithelia.

Journal of Histochemistry & Cytochemistry, 58, 653-668. DOI: 10.1369/jhc.2010.955211.

Saitou, N. and M. Nei (1987):

The neighbor-joining method: a new method for reconstructing phylogenetic trees.

Molecular Biology and Evolution, 4, 406-425. DOI: 10.1093/oxfordjournals.molbev.a040454.

Sala-Rabanal, M., Z. Yurtsever, C. G. Nichols and T. J. Brett (2015):

Secreted CLCA1 modulates TMEM16A to activate Ca(2+)-dependent chloride currents in human cells.

Elife, 4, e05875. DOI: 10.7554/eLife.05875.

Sasaki, Y., R. Koyama, R. Maruyama, T. Hirano, M. Tamura, J. Sugisaka, H. Suzuki, M. Idogawa, Y. Shinomura and T. Tokino (2012):

CLCA2, a target of the p53 family, negatively regulates cancer cell migration and invasion.

Cancer Biology & Therapy, 13, 1512-1521. DOI: 10.4161/cbt.22280.

Schmid, M., B. J. Evans and J. P. Bogart (2015):

Polyploidy in Amphibia.

Cytogenetic and Genome Research, 145, 315-330. DOI: 10.1159/000431388.

Seltmann, K., M. Meyer, J. Sulcova, T. Kockmann, U. Wehkamp, S. Weidinger, U. auf dem Keller and S. Werner (2018):

Humidity-regulated CLCA2 protects the epidermis from hyperosmotic stress.

Science Translational Medicine, 10, eaao4650. DOI: 10.1126/scitranslmed.aao4650.

Shao, T., Y. Xue and M. Fang (2021):

Epigenetic Repression of Chloride Channel Accessory 2 Transcription in Cardiac Fibroblast: Implication in Cardiac Fibrosis.

Frontiers in Cell and Developmental Biology, 9, 771466. DOI: 10.3389/fcell.2021.771466.

Sharma, A., G. Ramena, Y. Yin, L. Premkumar and R. C. Elble (2018):

CLCA2 is a positive regulator of store-operated calcium entry and TMEM16A.

PLoS One, 13, e0196512. DOI: 10.1371/journal.pone.0196512.

Shinmura, K., H. Igarashi, H. Kato, Y. Kawanishi, Y. Inoue, S. Nakamura, H. Ogawa, T. Yamashita, A. Kawase, K. Funai and H. Sugimura (2014):

CLCA2 as a novel immunohistochemical marker for differential diagnosis of squamous cell carcinoma from adenocarcinoma of the lung.

Disease Markers, 2014, 619273. DOI: 10.1155/2014/619273.

Sokal, R. R. (1958):

A statistical method for evaluating systematic relationships.

University of Kansas, Scicence Bulletin, 38, 1409-1438.

Song, J., X. Zhang, Z. Qi, G. Sun, S. Chi, Z. Zhu, J. Ren, Z. Qiu, K. Liu, L. Myatt and R. Z. Ma (2009):

Cloning and Characterization of a Calcium-Activated Chloride Channel in Rat Uterus1.

Biology of Reproduction, 80, 788-794. DOI: 10.1095/biolreprod.108.071258.

Sonnhammer, E. L. and E. V. Koonin (2002):

Orthology, paralogy and proposed classification for paralog subtypes.

Trends in Genetics, 18, 619-620. DOI: 10.1016/s0168-9525(02)02793-2. DOI: 10.1016/s0168-9525(02)02793-2

Spofford, J. B. (1969):

Heterosis and the evolution of duplications.

The American Naturalist, 103, 407-432.

Stamboulian, M., R. F. Guerrero, M. W. Hahn and P. Radivojac (2020):

The ortholog conjecture revisited: the value of orthologs and paralogs in function prediction.

Bioinformatics, 36, i219-i226. DOI: 10.1093/bioinformatics/btaa468.

REFERENCES

Sung, H., J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal and F. Bray (2021):

Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries.

CA: A Cancer Journal for Clinicians, 71, 209-249. DOI: 10.3322/caac.21660.

Tanikawa, C., H. Nakagawa, Y. Furukawa, Y. Nakamura and K. Matsuda (2012):

CLCA2 as a p53-inducible senescence mediator.

Neoplasia, 14, 141-149. DOI: 10.1593/neo.111700.

Tatusov, R. L., E. V. Koonin and D. J. Lipman (1997):

A genomic perspective on protein families.

Science, 278, 631-637. DOI: 10.1126/science.278.5338.631.

Thai, P., Y. Chen, G. Dolganov and R. Wu (2005):

Differential regulation of MUC5AC/Muc5ac and hCLCA-1/mGob-5 expression in airway epithelium.

American Journal of Respiratory Cell and Molecular Biology, 33, 523-530. DOI: 10.1165/rcmb.2004-0220RC.

Toda, M., M. K. Tulic, R. C. Levitt and Q. Hamid (2002):

A calcium-activated chloride channel (HCLCA1) is strongly related to IL-9 expression and mucus production in bronchial epithelium of patients with asthma.

Journal of Allergy and Clinical Immunology 109: 246-250. DOI: https://doi.org/10.1067/mai.2002.121555.

Van Der Doef, H., M. Slieker, D. Staab, B. Alizadeh, M. Seia, C. Colombo, C. Van Der Ent, R. Nickel, H. Witt and R. Houwen (2010):

Association of the CLCA1 p. S357N variant with meconium ileus in European patients with cystic fibrosis.

Journal of Pediatric Gastroenterology and Nutrition 50: 347-349. DOI: 10.1097/MPG.0b013e3181afce6c.

Walia, V., M. Ding, S. Kumar, D. Nie, L. S. Premkumar and R. C. Elble (2009):

hCLCA2 is a p53-inducible inhibitor of breast cancer cell proliferation.

Cancer Research, 69, 6624-6632. DOI: 10.1158/0008-5472.CAN-08-4101.

Wang, K., F. Q. Wen, Y. L. Feng, X. M. Ou, D. Xu, J. Yang and Z. P. Deng (2007):

Increased expression of human calcium-activated chloride channel 1 gene is correlated with mucus overproduction in Chinese asthmatic airway.

Cell Biology International 31: 1388-1395. DOI: 10.1016/j.cellbi.2007.06.004.

Weaver, S., S. D. Shank, S. J. Spielman, M. Li, S. V. Muse and S. L. Kosakovsky Pond (2018):

Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes.

Molecular Biology and Evolution, 35, 773-777. DOI: 10.1093/molbev/msx335.

Wei, F.-Z., S.-W. Mei, Z.-J. Wang, J.-N. Chen, H.-Y. Shen, F.-Q. Zhao, J. Li, Z. Liu and Q. Liu (2020):

Differential Expression Analysis Revealing CLCA1 to Be a Prognostic and Diagnostic Biomarker for Colorectal Cancer.

Frontiers in Oncology, 10, 573295. DOI: 10.3389/fonc.2020.573295.

Winpenny, J. P., L. L. Marsey and D. W. Sexton (2009):

The CLCA gene family: putative therapeutic target for respiratory diseases.

Inflammation & Allergy Drug Targets, 8, 146-160. DOI: 10.2174/187152809788462590.

Woodruff, P. G., H. A. Boushey, G. M. Dolganov, C. S. Barker, Y. H. Yang, S. Donnelly, A. Ellwanger, S. S. Sidhu, T. P. Dao-Pick, C. Pantoja, D. J. Erle, K. R. Yamamoto and J. V. Fahy (2007):

Genome-wide profiling identifies epithelial cell genes associated with asthma and with treatment response to corticosteroids.

Proceedings of the National Academy of Sciences of the United States of America, 104, 15858-15863. DOI: 10.1073/pnas.0707413104.

Woodruff, P. G., B. Modrek, D. F. Choy, G. Jia, A. R. Abbas, A. Ellwanger, L. L. Koth, J. R. Arron and J. V. Fahy (2009):

T-helper type 2-driven inflammation defines major subphenotypes of asthma.

American Journal of Respiratory and Critical Care Medicine, 180, 388-395. DOI: 10.1164/rccm.200903-0392OC.

Yamazaki, J., K. Okamura, K. Ishibashi and K. Kitamura (2005):

Characterization of CLCA protein expressed in ductal cells of rat salivary glands.

Biochimica et Biophysica Acta (BBA) - Biomembranes, 1715, 132-144. DOI: https://doi.org/10.1016/j.bbamem.2005.08.001.

Yamazaki, J., K. Okamura, K. Uehara and M. Hatta (2013):

CLCA splicing isoform associated with adhesion through β1-integrin and its scaffolding protein: specific expression in undifferentiated epithelial cells.

Journal of Biological Chemistry, 288, 4831-4843. DOI: 10.1074/jbc.M112.396481.

REFERENCES

Yang, B., L. Cao, B. Liu, C. D. McCaig and J. Pu (2013):

The transition from proliferation to differentiation in colorectal cancer is regulated by the calcium activated chloride channel A1.

PLoS One, 8, e60861. DOI: 10.1371/journal.pone.0060861.

Yang, B., L. Cao, J. Liu, Y. Xu, G. Milne, W. Chan, S. D. Heys, C. D. McCaig and J. Pu (2015):

Low expression of chloride channel accessory 1 predicts a poor prognosis in colorectal cancer.

Cancer, 121, 1570-1580. DOI: 10.1002/cncr.29235.

Yoon, I.-S., S. M. Jeong, S. N. Lee, J.-H. Lee, J.-H. Kim, M. K. Pyo, J.-H. Lee, B.-H. Lee, S.-H. Choi, H. Rhim, H. Choe and S.-Y. Nah (2006):

Cloning and Heterologous Expression of a Ca^{2+}-Activated Chloride Channel Isoform from Rat Brain.

Biological and Pharmaceutical Bulletin, 29, 2168-2173. DOI: 10.1248/bpb.29.2168.

Yu, Y., V. Walia and R. C. Elble (2013):

Loss of CLCA4 promotes epithelial-to-mesenchymal transition in breast cancer cells.

PloS One, 8, e83943. DOI: 10.1371/journal.pone.0083943.

Zeeuwen, P. L. (2004):

Epidermal differentiation: the role of proteases and their inhibitors.

European Journal of Cell Biology, 83, 761-773. DOI: 10.1078/0171-9335-00388.

Zhang, J. (2003):

Evolution by gene duplication: an update.

Trends in Ecology & Evolution, 18, 292-298. DOI: 10.1016/S0169-5347(03)00033-8.

Zhang, P., Y. Lin and Y. Liu (2021):

CLCA2 suppresses the proliferation, migration and invasion of cervical cancer.

Experimental and therapeutic medicine, 22, 776-776. DOI: 10.3892/etm.2021.10208.

Zhou, Y., Q. Dong, J. Louahed, C. Dragwa, D. Savio, M. Huang, C. Weiss, Y. Tomer, M. P. McLane and N. C. Nicolaides (2001):

Characterization of a calcium-activated chloride channel as a shared target of Th2 cytokine pathways and its potential involvement in asthma.

American Journal of Respiratory Cell and Molecular Biology, 25, 486-491. DOI: 10.1165/ajrcmb.25.4.4578.

Zhu, D. and B. Pauli (1991):

Generation of monoclonal antibodies directed against organ-specific endothelial cell surface determinants.

Journal of Histochemistry & Cytochemistry, 39, 1137-1142. DOI: 10.1177/39.8.1856462.

Appendix

The nomenclature of the cloned bovine and rat *CLCA3* genes in the respective original publications is not according to the NCBI Genebank entries and nomenclature found in both is not indicative of the associated CLCA cluster. This causes confusion and hampers interpretation of the data. In order to solve these problems, protein sequences were retrieved from the NCBI Genbank and protein sequences were extracted from the literature. Based on these sequences, phylogenetic trees were generated using the Maximum Likelihood method.

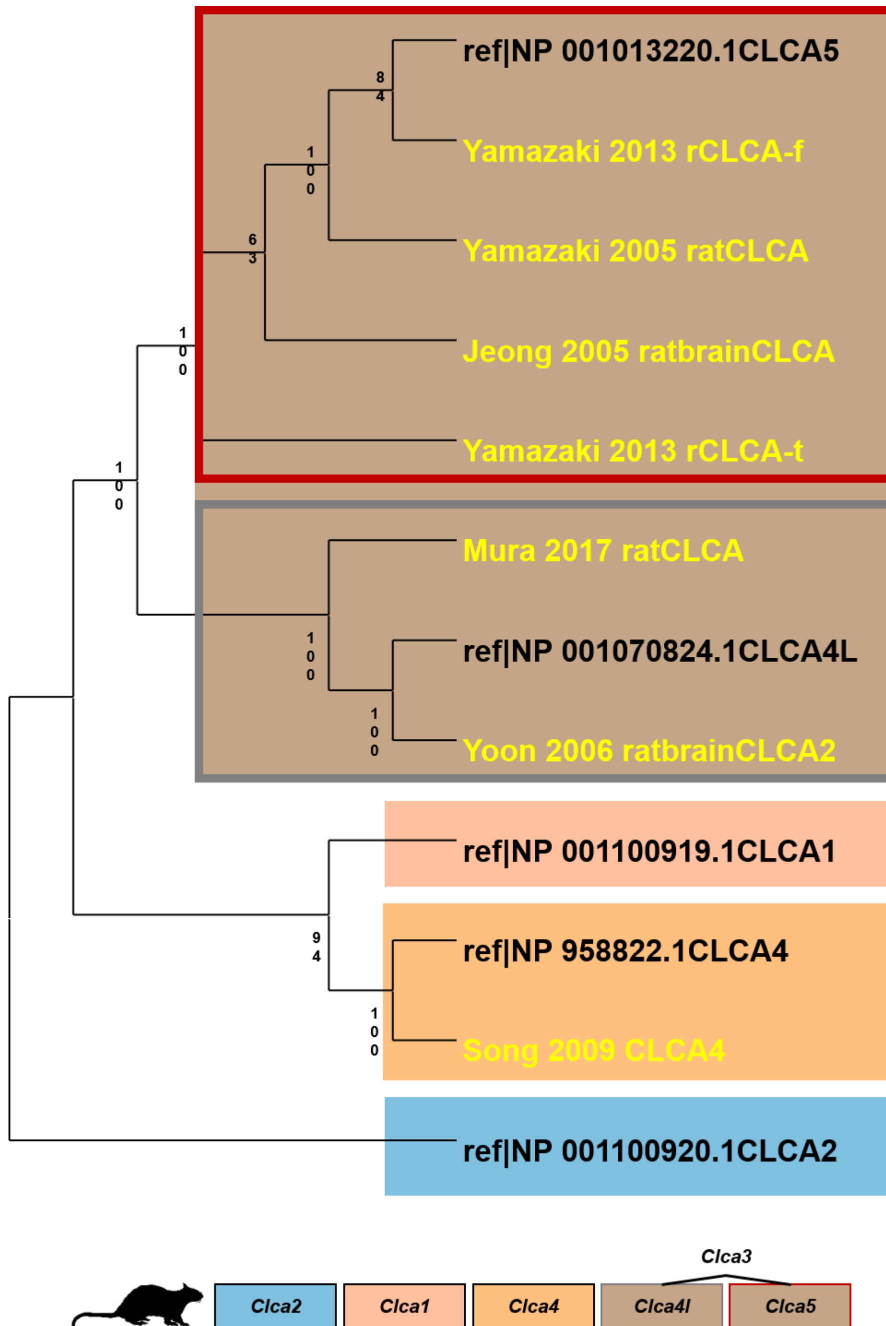


Figure 3. Relationship between published rat CLCA amino acid sequences in the literature compared to those found in the NCBI Genbank

The evolutionary history was inferred by using the Maximum Likelihood method and Whelan and Goldman model. The tree with the highest log likelihood (-4325.61) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model, and then selecting the topology with superior log likelihood value. The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 15.69% sites). This analysis involved 12 amino acid sequences. All positions containing gaps and missing data were eliminated (complete deletion option). There were a total of 459 positions in the final dataset. Evolutionary analyses were conducted in MEGA X. Yellow font indicates the published sequence in the literature, black font that of the NCBI Genbank sequence. This analysis indicates that Clca4l was cloned two times independently, while Clca5 was cloned four times. Both belong to the CLCA3 cluster. Differences between the sequences found in the literature compared to the ones from NCBI Genbank might be due SNP based mutations or splice products. Clca4 was cloned only one time.

Publications mentioned in figure 3 are listed below:

Mura, C. V., R. Delgado, M. G. Delgado, D. Restrepo and J. Bacigalupo (2017):

A CLCA regulatory protein present in the chemosensory cilia of olfactory sensory neurons induces a $Ca^{(2+)}$-activated Cl(-) current when transfected into HEK293.

BMC Neuroscience, 18, 61. DOI: 10.1186/s12868-017-0379-7.

Song, J., X. Zhang, Z. Qi, G. Sun, S. Chi, Z. Zhu, J. Ren, Z. Qiu, K. Liu, L. Myatt and R. Z. Ma (2009):

Cloning and Characterization of a Calcium-Activated Chloride Channel in Rat Uterus1.

Biology of Reproduction, 80, 788-794. DOI: 10.1095/biolreprod.108.071258.

Yamazaki, J., K. Okamura, K. Ishibashi and K. Kitamura (2005):

Characterization of CLCA protein expressed in ductal cells of rat salivary glands.

Biochimica et Biophysica Acta (BBA) – Biomembranes, 1715, 132-144. DOI: 10.1016/j.bbamem.2005.08.001.

Yamazaki, J., K. Okamura, K. Uehara and M. Hatta (2013):

CLCA splicing isoform associated with adhesion through β1-integrin and its scaffolding protein: specific expression in undifferentiated epithelial cells.

Journal of Biological Chemistry, 288, 4831-4843. DOI: 10.1074/jbc.M112.396481.

Yoon, I.-S., S. M. Jeong, S. N. Lee, J.-H. Lee, J.-H. Kim, M. K. Pyo, J.-H. Lee, B.-H. Lee, S.-H. Choi, H. Rhim, H. Choe and S.-Y. Nah (2006):

Cloning and Heterologous Expression of a Ca^{2+}-Activated Chloride Channel Isoform from Rat Brain.

Biological and Pharmaceutical Bulletin, 29, 2168-2173. DOI: 10.1248/bpb.29.2168 .

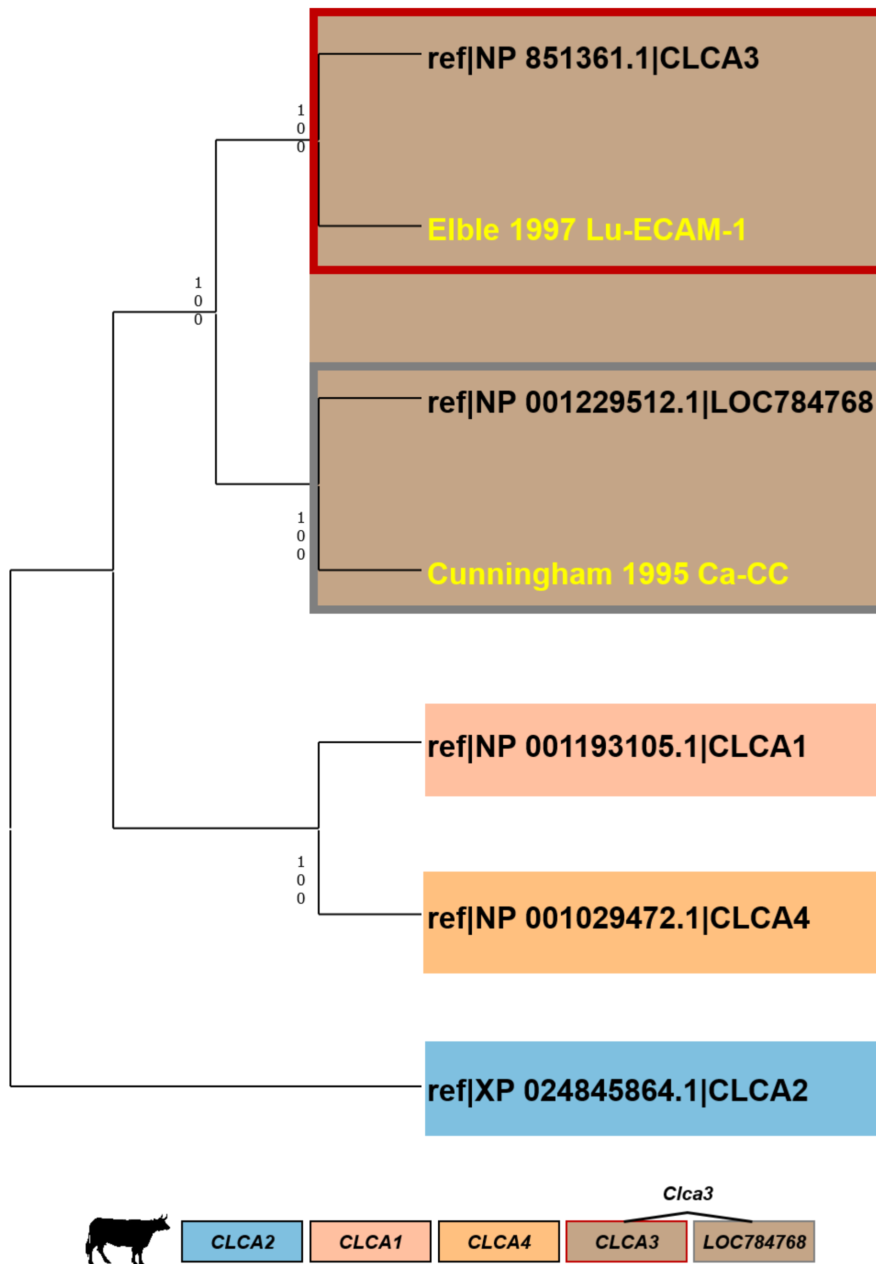


Figure 4. Relationship between published bovine CLCA amino acid sequences in the literature compared to those found in the NCBI Genbank

The evolutionary history was inferred by using the Maximum Likelihood method and Whelan And Goldman model. The tree with the highest log likelihood (-5400.08) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.6586)). This analysis involved 7 amino acid sequences. All positions containing gaps and missing data were eliminated (complete deletion option). There were a total of 638 positions in the final dataset. Evolutionary analyses were conducted in MEGA X. Yellow font indicates the published sequence in the literature, black font that of the NCBI Genbank sequence. This analysis indicates that CLCA3 and LOC784768 were cloned.

Publications mentioned in figure 4 are listed below:

Cunningham, S. A., M. S. Awayda, J. K. Bubien, Ismailov, II, M. P. Arrate, B. K. Berdiev, D. J. Benos and C. M. Fuller (1995):

Cloning of an epithelial chloride channel from bovine trachea.

Journal of Biological Chemistry, 270, 31016-31026. DOI: 10.1074/jbc.270.52.31016.

Elble, R. C., J. Widom, A. D. Gruber, M. Abdel-Ghany, R. Levine, A. Goodwin, H. C. Cheng and B. U. Pauli (1997):

Cloning and characterization of lung-endothelial cell adhesion molecule-1 suggest it is an endothelial chloride channel.

Journal of Biological Chemistry, 272, 27853-27861. DOI: 10.1074/jbc.272.44.27853.

List of own publications regarding this topic

Research papers:

Bartenschlager, F., Klymiuk, N., Weise, C., Kuropka, B., Gruber, A. D., Mundhenk, L. (2022): Evolutionarily conserved properties of CLCA proteins 1, 3 and 4, as revealed by phylogenetic and biochemical studies in avian homologues. PloS One 17: e0266937.

Bartenschlager, F., Klymiuk, N., Gruber A.D., Mundhenk L. (2022): Genomic, biochemical and expressional properties reveal strong conservation of the *CLCA2* gene in birds and mammals. PeerJ 10:e14202.

Presentations at scientific conferences with published abstracts:

Poster presentations:

Bartenschlager, F., Mundhenk, L., Klymiuk, N., Gruber, A. D. (2019). Is the *CLCA* gene family a unique among mammals? *12. Doktorandensymposium & DRS Präsentationsseminar "Biomedical Sciences" Berlin – 27.09.2019. In: Doktorandensymposium & DRS Präsentationsseminar "Biomedical Sciences" Berlin: Mensch und Buch Verlag, ISBN: 978-3-96729-006-6, P45.*

Bartenschlager, F., Klymiuk, N., Gruber, A. D., & Mundhenk, L. (2020). Evolution auf molekularer Ebene–die *CLCA*-Genfamilie im Vergleich zwischen Säuger und Vogel. *Tierärztliche Praxis Ausgabe K: Kleintiere/Heimtiere, 48(03), P15.*

Bartenschlager, F., Klymiuk, N., Gruber, A. D., & Mundhenk, L. (2021). Evolution auf molekularer Ebene II: Gemeinsamkeiten von CLCA2 bei Sauropsiden und Säugern. *Tierärztliche Praxis Ausgabe K: Kleintiere/Heimtiere, 49(03), P08.*

Hoppe, J., **Bartenschlager, F**., Mundhenk, L., Jovanovic, V., & Gruber, A. D. (2021). Gut recherchiert? – Eine Annäherung an die Rolle von CLCA1 in Gelenken durch die Nutzung molekularer Datenbanken. *Tierärztliche Praxis Ausgabe K: Kleintiere/Heimtiere, 49(03), P12.*

Oral presentations:

Bartenschlager, F. Klymiuk, N., Gruber, A. D., & Mundhenk, L. (2021). The *CLCA* gene family – an exclusive entity of the mammals? *9. Leipziger Doktorand:innenforum : 14. Juli 2021 : Programm & Abstracts — Belinda Euring, Maxi Harzer, Christin Kamutzki, Sophia Kreyer, Franziska C. Wagner, Sophie Öhlmann (Hrsg.). 1. Auflage. Leipzig: Universität Leipzig Veterinärmedizinische Fakultät, 2021, urn:nbn:de:bsz:15-qucosa2-753877, P24.*

Acknowledgements

First and foremost I thank **Prof. Dr. Achim D. Gruber, Ph.D.**, Chair of the Institute of Veterinary Pathology, Freie Universität Berlin, for giving me the chance to work in this outstandingly diverse institution which is built by unique people that are passionate about research, pathology and veterinary medicine. I am grateful for your great support on this project, the trust that has been placed in me, your mentoring, the space that I received to develop my own ideas and for sharing your expertise at all times.

My special gratitude goes to my principle investigator **PD Dr. Lars Mundhenk,** who reinvented the terms mentoring and supervision by creating a working environment that was saturated with outstanding scientific discourses, sophisticated criticism, never ending patience, creativity, optimism, motivation and passionate support. The essence of this work is deeply influenced by his person.

Furthermore, I thank my mentor **Prof. Dr. Nikolai Klymiuk** for many exceptionally educational scientific debates, sharing every aspect of his expertise and for motivating me to think outside the box.

I highly appreciate the great support of **Dr. Chris Weise** and **Dr. Benno Kuropka**, who fascinated me for the mass spectrometry technique and significantly improved the quality of this work by profound scientific discussions and suggestions.

All members of the histolab are highly appreciated for their outstanding histological preparations and flexibility. There are no challenges for you, just solutions.

I sincerely thank **Dr. Svenja Lilge** for the best last minute proof reading, for being an outstanding friend, motivator and colleague.

Special thanks go to **Dr. Gerhard Sponder**, who supervised the best research externship during my time as undergraduate student and by this, inspired me to molecular biology. I am grateful for sharing all of your knowledge with me at any time but also for being an awesome colleague and friend.

I sincerely thank **Dr. Olivia Kershaw** and **Prof. Dr. Robert Klopfleisch** for providing the best training in pathology, sharing all your never-ending pathology knowledge with me, for the scientific discussions and for all the little and larger chats, your motivation and support.

Furthermore I express my great gratitude to my dear and precious friends and colleagues **Sina Feyer** as well as **Dr. Sophie Merz**, **Hannah Pischon**, **PhD, Moritz Radbruch**, **Dr. Anne Petrick** and **Dr. Christof Bertram, PhD** for all of your moral support, for motivating me, for the

best lunch and coffee breaks, for sharing all of your knowledge with me and for making the days at work fantastic.

I also thank **all my doctoral colleagues** and **all other members of the institute** for the great cooperation, the nice chats and for being the embodiment of the term *team*.

Finally yet importantly, I express my deepest gratitude to my **family** for their understanding, unconditional love, trust in me and their never-ending support.

Funding

The authors received specific funding for the article processing charges from the Open Access Publication Fund of Freie Universität Berlin.

Conflict of interest

The author of this thesis declares no conflict of interest.

Declaration of originality

Hereby, I declare that the present thesis has been prepared by myself. I assure that I exclusively used the mentioned sources and facilities.

Berlin, 07.02.2023 Florian Bartenschlager