We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS

BOOK CITATION INDEX

INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Bucentaur (Bcnt)[1] Gene Family: Gene Duplication and Retrotransposon Insertion

Shintaro Iwashita and Naoki Osada
Iwaki Meisei University / National Institute of Genetics
Japan

1. Introduction

Members of multiple gene families in higher organisms allow for more refined cellular signaling networks and structural organization toward more stable physiological homeostasis. Gene duplication is one the most powerful ways of providing an opportunity to create a novel gene(s) because a novel function might be acquired without the loss of the original gene function (Ohno, 1970). Gene duplication can result from unequal crossing over by recombination, retroposition of cDNA, or whole-genome duplication. Furthermore, a replication-based mechanism of change in gene copy number has been proposed recently (Hastings et al., 2009). Gene duplication generated by retroposition is frequently accompanied by deleterious effects because the insertion of cDNA into the genome is nearly random or unlinks the original gene location resulting in an alteration of the original vital functions of the target genes. Thus retroelements such as transposable elements and endogenous retroviruses have been thought of as "selfish". On the other hand, gene duplication caused by unequal crossing over generally results in tandem alignment, which less frequently disrupts the functions of other genes. Recent genome-wide studies have demonstrated that retroelements can definitely contribute to the creation of individual novel genes and the modulation of gene expression, which allows for the dynamic diversity of biological systems, such as placental evolution (Rawn & Cross, 2008). It is now recognized that tandem duplication and retroposition are among the key factors that initiate the creation of novel gene family members (Brosius, 2005; Sorek, 2007; Kaessmann, 2010). By these mechanisms, species-specific gene duplication can lead to species-specific gene functions, which might contribute to species-specific phenotypes (Zhang, 2003). For example, many genes derived from retroelements are expressed in mammalian placentas, and species-specific gene duplication has occurred multiple times during placental evolution (Rawn & Cross, 2008). If a combination of tandem gene duplication and retroposition of cDNA occurs, there is a good possibility for the creation of a novel gene(s)

---

[1]Although the vertebrate *Bcnt* (Bucentaur) gene is officially called *Cfdp1* (craniofacial developmental protein 1), its biological function remains unclear. So far, solid evidence that the gene is involved in craniofacial development has not been provided except for its unique expression during mouse tooth development (Diekwisch et al., 1999). The authors are concerned that a "wrong" naming may have caused confusion concerning the function of the *Bcnt/Cfdp1* gene. Thus we use the names *Bcnt/Cfdp1*, *p97Bcnt/Cfdp2* and *p97Bcnt-2* in this article.

because a novel function could be acquired with the guarantee that the original gene functions will be retained. This type of evolutionary process has been described, e.g. in the *Jingwei* gene of *Drosophila*, where segmental duplication of a certain gene followed by retroposition of alcohol dehydrogenase (*Adh*) cDNA into one of the copied genes created a new Adh with an altered substrate specificity (Zhang et al., 2004). Furthermore, since the insertion of retrotransposons can speed up the natural mutation process tremendously (Makalowski, 2003), the combined process of tandem duplication followed by retrotransposon insertion has a greater potential to generate a novel gene(s) (Fig. 1). Indeed we have identified an example of this type, the p97Bcnt protein (Nobukuni et al., 1997). The *p97Bcnt/Cfdp2* gene was created in a common ancestor of ruminants by a partial duplication of the ancestral *Bcnt/Cfdp1* gene followed by insertion of an order-specific retrotransposon, Bov B-LINE (Iwashita et al., 2003, 2006). As a result, the paralog recruited an apurinic/apyrimidinic (AP)-endonuclease domain in the middle of the protein. In this article, we summarize the gene organization and protein structures of three Bcnt family members, and describe their biochemical characteristics. We also argue that the process of tandem duplication followed by retroelement insertion generates a high potential for creating novel genes for expanding signaling networks.
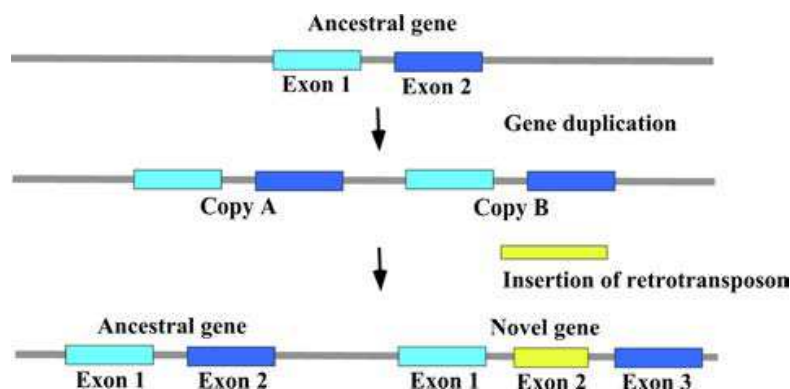

Fig. 1. Mechanism of novel gene creation by a combination of gene duplication and retrotransposition

If segmental duplication followed by retrotransposon insertion occurs, it provides a good opportunity to generate a paralogous gene, because a novel function could be acquired under the guarantee that the original gene function will remain. The schematic is a modification of the original (Makalowski, 2003)

2. Establishment of three *Bcnt* members

2.1 A bovine specific retrotransposon, Bov B-LINE

Autonomous non-long-terminal repeat (non-LTR) retrotransposons, also termed Long-Interspersed Nuclear Elements (LINEs), have been identified in almost all eukaryotic organisms. Based on their structures and type of endonuclease, non-LTR retrotransposons are classified into two subtypes. The major subtype encodes an endonuclease with homology to AP-endonuclease (APE), thus termed APE-type non-LTR retrotransposons. These APE-type elements are now divided into four groups and eleven clades (Zingler et al., 2005). The RTE clade is one of the most widespread and shortest APE-type non-LTR retrotransposons, which are truncated forms of L1 (human LINE 1) lacking both the 5' and 3' regions (Fig. 2).
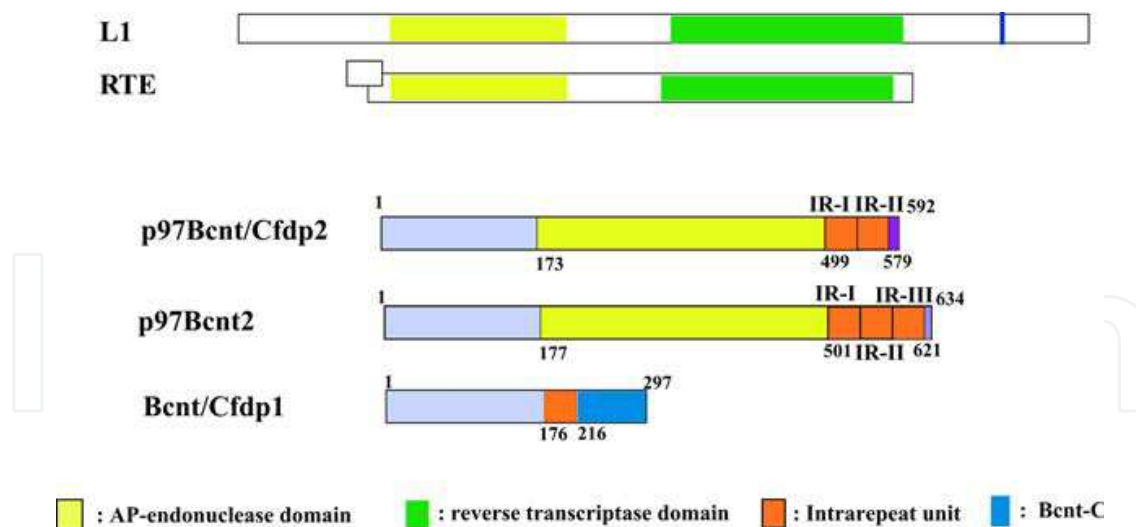
Fig. 2. The structural relationships among the open reading frames of retrotransposable L1 and RTE elements and three Bcnt-related proteins

The L1 and RTE elements have apurinic/apyrimidic (AP)-endonuclease domains (yellow boxes) and reverse transcriptase domains (green boxes). The L1 element has another restriction enzyme-like endonuclease domain in the C-terminal region (dark blue bar). The square to the left of RTE indicates the ambiguity of the 5' region. The assignment of the domains in L1 and RTE is according to Malik & Eickbush, 1998. The numbers above the rectangles of the three Bcnt-related proteins, Bcnt/Cfdp1, p97Bcnt/Cfdp2 and p97Bcnt2, indicate amino acid residue numbers. The latter two contain a region derived from the AP-endonuclease domain of RTE (termed the RTE domain) in the middle of their molecules. As described below, the three proteins have common acidic N-terminal regions (grey boxes) and intramolecular repeat (IR) units consisting of 40 amino acids each (orange boxes). The blue box at the C-terminus of ancestral Bcnt/Cfdp1 indicates a conserved 82-amino acids region (Bcnt-C)

Retrotransposons spread through vertical transmission, but occasionally through horizontal transmission (Gentles et al., 2007). Bov-B LINEs are order-specific RTEs that are found specifically in ruminants, where they were initially reported as a bovine *Alu*-like dimer-driven family; they potentially encode both an AP-enodnuclease domain and a reverse transcriptase domain accompanied by a short interspersed repetitive element (SINE) cassette (Szemraj et al., 1995). It has been suggested that BovB-LINEs were transferred horizontally from squamata into an ancestral ruminant and expanded in all ruminants (Zupunski et al., 2001). *p97Bcnt/Cfdp 2* recruited the AP-endonuclease domain of Bov-B LINE during the creation process in an ancient ruminant.

2.2 Discovery of a novel protein, p97Bcnt/Cfdp2

The p97Bcnt/Cfdp2 protein was discovered in bovine brain during screening for hybridoma producing monoclonal antibodies (mAbs). In the course of a study on Ras GTPase-activating proteins (GAPs, RAS p21 protein activators, Rasa), we had attempted to generate mAbs to distinguish each GAP from among their family members (Kobayashi et al., 1993; Iwashita & Song, 2008). We used a glutathione-S-transferase (GST) fusion protein of rat Rasa2 (GAP[1m]) as an immunoantigen and screened for hybridomas by western blotting using bovine brain extract. We isolated five independent clones, all of which showed a single broad band with an apparent molecular mass of 97 kDa, exactly the expected size of rat Rasa2. At one time

we thought we had obtained appropriate antibodies, but the target protein was entirely different from Rasa as described below. Although we screened a bovine brain cDNA expression library by western blotting with the obtained mAbs, we could not clone the target molecule. Instead, a 97 kDa protein was isolated from bovine brain extract by affinity chromatography with the antibodies, and the amino acid sequences of its protease-digested fragments were determined. We used redundant primers designed based on the determined peptide sequences as DNA probes, and cloned the target molecule by both "rapid amplification of cDNA ends" (RACE) and screening of a bovine brain cDNA library (Nobukuni et. al., 1997). The obtained clone, which had an open reading frame of 592 amino acids, was named Bcnt after bucentaur, a Greek mythical creature that is half man and half ox, implying a strange protein from bovine brain. The identified protein, named p97Bcnt, Bcnt with a molecular mass of 97kDa, consists of an acidic N-terminal region, a retrotransposon-derived 325-amino acid region (termed the RTE domain), and two 40-amino acid intrarepeat (IR) units. The RTE domain is 72% identical to an order-specific retrotransposon, Bov-B LINE (GenBank accession number AF332697). The relationship between p97Bcnt and its estimated epitope in the mAbs, which enabled us to identify the protein, is summarized in Fig. 3. It provides a reasonable explanation as to why the unique protein was isolated by mAbs generated by a GST-fusion protein of rat Rasa2 as an immunoantigen. The estimated epitope of five independent mAbs maps on a single site in the N-terminal region of p97Bcnt/Cfdp2, and the antibodies recognize neither human BCNT/CFDP1 (Nobukuni et al., 1997) nor bovine Bcnt/Cfdp1 (Iwashita et al., 2003). The junction region of the fusion protein between GST and truncated Rasa2 codes a unique amino acid sequence generated by the extra nucleotides of the multiple cloning sites and a nucleotide linker for plasmid construction. Since Rasa is a highly conserved protein in mammals, the junction region might present strong antigenicity. Generally, it is hard to clone a target molecule by direct DNA screening when interspersed repetitive sequences are involved. Therefore, we first isolated a 97kDa protein that was recognized by the accidentally generated mAbs, determined its amino acid sequence, and then screened a cDNA library with the designed oligonucleotide probes. This led to the identification of a unique protein, p97Bcnt/Cfdp2.

2.3 Identification of three Bcnt-related proteins

Immediately after the identification of *p97Bcnt/Cfdp2*, we isolated its human and mouse counterparts, and examined their differences from *p97Bcnt/Cfdp2* at both the cDNA and genome levels (Nobukuni et al., 1997; Takahashi et al., 1998). The counterparts, called (ancestral) *Bcnt/Cfdp1*, have homologous acidic N-terminal regions and one IR unit of 40-amino acids, but lack the RTE domain. Instead they contain a highly conserved 82-amino acid region at the C-terminus that is not present in p97Bcnt/Cfdp2 (Fig. 2) as will be described below. Subsequently, we found that ruminants have both ancestral *Bcnt/Cfdp1* and *p97Bcnt/Cfdp2*, while other animals have only *Bcnt/Cfdp1*. The pairwise sequence alignment of bovine and human genome DNA revealed that the region encompassing the gene was duplicated in two rounds in bovines (Iwashita et al., 2003). Although automated computational annotation predicted another homolog of *p97Bcnt* (LOC514131) in the bovine genome, its 5′ UTR was different from the full-length cDNA that we isolated. Then we identified another paralog, termed *p97Bcnt-2,* in the adjacent region (Iwashita et al., 2009). The gene product, p97Bcnt2, is highly homologous to p97Bcnct/Cfdp2, comprising an acidic N-terminal region, a 324-amino acid RTE domain, and three IR units instead of the two in p97Bcnt/Cfdp2 in the C-terminal region (Fig. 2).
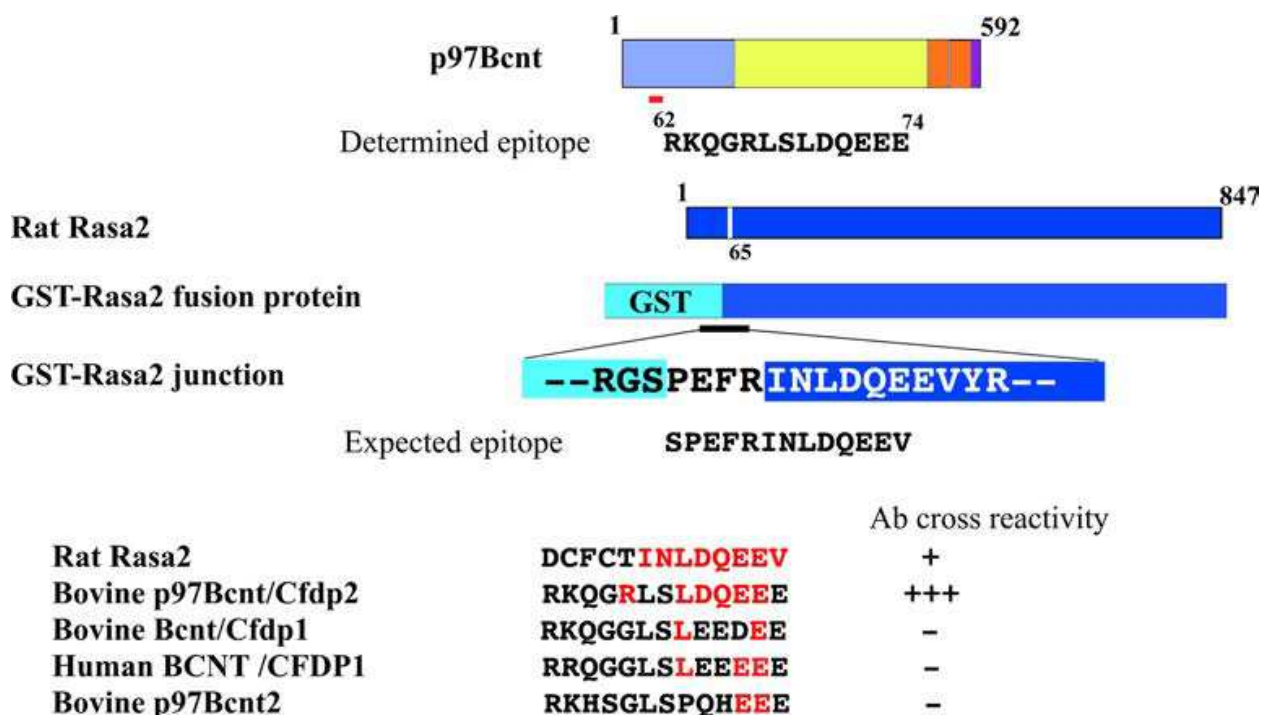
Fig. 3. Epitopes of the monoclonal antibodies that enabled the identification of the p97Bcnt/Cfdp2 protein

A plasmid of a fusion protein of truncated rat Rasa2 (from Ile65 to Ser847) and glutathione S-transferase was constructed using a linker (by Dr. S. Hattori), expressed in *Escherichia coli*, and its protein was purified by glutathione-affinity column chromatography. mAbs against the fusion protein were isolated according to a conventional method. Epitope mapping was carried out using the full-length cDNA of the targeted molecule, hereafter *p97Bcnt*. Fragments of ~300 base pairs in size were expressed in a protein expression vector and screened with the obtained mAbs. Seven positive clones were isolated from among ~9 x 10^3 bacterial colonies, and the sequence common to all clones was determined as the possible epitope for anti-p97Bcnt antibodies (13-amino acids, RKQGRLSLDQEEE, represented by the red bar in the upper part) (Nobukuni et al., 1997). Amino acid sequences corresponding to the epitope region of rat Rasa2, bovine p97Bcnt/Cfdp2, human BCNT/CFDP1, bovine Bcnt/Cfdp1, and bovine p97Bcnt2 are aligned and amino acid residues identical to those in the expected epitope are indicated in red

3. Tandem alignment of three *Bcnt* gene family members

The draft bovine genome sequence was published in 2009 (The Bovine Genome Sequencing and Analysis Consortium, 2009). The initial analysis estimated that the bovine genome contains about 22,000 genes, with a core set of 14,345 orthologs shared among seven mammalian species. It has been shown that 3.1% of the bovine genome consists of recently duplicated sequences (judged by sequences ≥ 1 kb in length and ≥ 90% identity), and more than three-quarters (75-90%) of segmental duplications are organized into local tandem duplication clusters (Liu et al., 2009). It is noteworthy that cattle-specific evolutionary breakpoint regions in the chromosomes have a higher density of tandem duplications and enrichment of repetitive elements. Furthermore, it has been pointed out that bovine tandem gene duplication is significantly related to species-specific biological functions such as immunity, digestion, lactation, and reproduction (Liu et al., 2009).
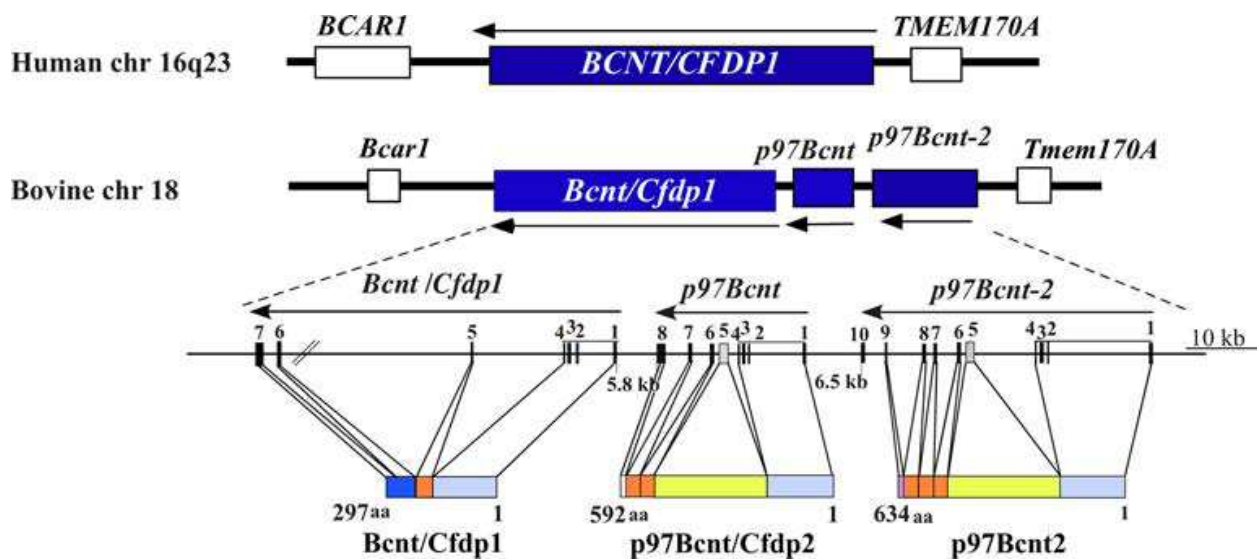
Fig. 4. Bovine *Bcnt/Cfdp1* locus and its corresponding region in the human genome

The organization of bovine *Bcnt/Cfdp1-p97Bcnt/Cfdp2-p97Bcnt-2* is shown schematically. *Bcar1*, Breast cancer anti-estrogen resistance 1 gene and *Tmem170A*, Transmembrane protein 170A gene, are located proximal and distal to the *Bcnt*-gene cluster or *BCNT* in both bovine chromosome 18 (middle part) and human chromosome 16q23 (upper part), respectively. The *Bcnt/Cfdp1, p97Bcnt/Cfdp2,* and *p97Bcnt-2* genes comprise 7, 8, and 10 exons, respectively (lower part); each exon is indicated by a vertical bar and is numbered

The three *Bcnt*-related genes are tandemly aligned on bovine chromosome 18 over a range of more than 177 kb, a syntenic region of human chromosome 16q23 (Fig. 4) and mouse chromosome 8. This gene cluster exists between the proximal breast cancer anti-estrogen resistance 1 gene (*Bcar1*) and the distal transmembrane protein *170A* gene (*Tmem*170A) in bovines, as is the case of *BCNT/CFDP1* in humans and *Bcnt/Cfdp1* in mice. Therefore, the cluster region was generated from an order-specific segmental duplication. It has been suggested that Bov-B LINEs emerged by horizontal transfer from squamata to ancient ruminants (Zupunski et al., 2001), and expanded just after the divergence of ruminantia and Camelidae (Jobse et al., 1995). Bov-B LINEs have further expanded in different lineages during the diversification of ruminant species after splitting from Tragulina, which was confirmed by hybridization with DNA fragments of the RTE domain of *Lesser Malay chevrotain* (Iwashita et al., 2006). This is consistent with the expansion of bovine SINEs (Jobse et al., 1995). *Tragulus javanicus*, the living fossil of the basal ruminant stock, shares a similar *Bcnt/Cfdp1 and p97Bcnt/Cfdp2* gene organization to bovines (Iwashita et al., 2006). Thus the partial gene duplication of the ancestral *Bcnt/Cfdp1* followed by the Bov-B LINE insertion occurred sometime after the Ruminantia-Suina-Tylopoda split and before the Pecora-Tragulina divergence, ~50 million years ago. A phylogenic tree has been constructed based on the N-terminal regions (~ 175 amino acids) encoded by the first four exons and shared among three *Bcnt*-related members. The tree topology suggests *p97Bcnt-2* was created from duplication of ancestor *p97Bcnt/Cfdp2* in an ancient ruminant prior to the Pecora-Tragulina divergence. Furthermore, using the 120-bp sequence corresponding to 40 amino acid residues in IR, duplication of the IR unit in *p97Bcnt/Cfdp2* is estimated to have occured prior to the creation of *p97Bcnt-2,* which has three IR units (Fig. 2). The two units in *p97Bcnt-2* (IR-

II and IR-III) diverged from IR-II in *p97Bcnt/Cfdp2*. We propose a parsimonious scenario for the creation of the three *Bcnt*-related genes in a process comprising 5 steps as shown in Fig. 5 (Iwashita et al., 2009). Self-BLAST search of the 120-kb region from *Tmem 170A* to exon 5 of *Bcnt/Cfdp1* confirms the two-round duplication of this gene cluster. Furthermore, homologous fragments of *Tmem170A* 3' UTR, which is located 6.8-kb distal to *p97Bcnt-2*, distribute at the 3'-region of both *Bcnt/Cfdp1* and *p97Bcnt/Cfdp2*. These data support the above scenario that resulted in the creation of the two paralogs, *p97Bcnt/Cfdp2* and *p97Bcnt-2*. Furthermore, both the processed pseudogene of *Bcnt/Cfdp1* and a 900-bp fragment encompassing the IR-II exon of *p9Bcnt-2* map on bovine chromosome 26 (Iwashita et al., 2009). It is interesting to examine the relationship between the retrotransposon-mediated creation of novel genes and the occurrence of processed pseudogenes.
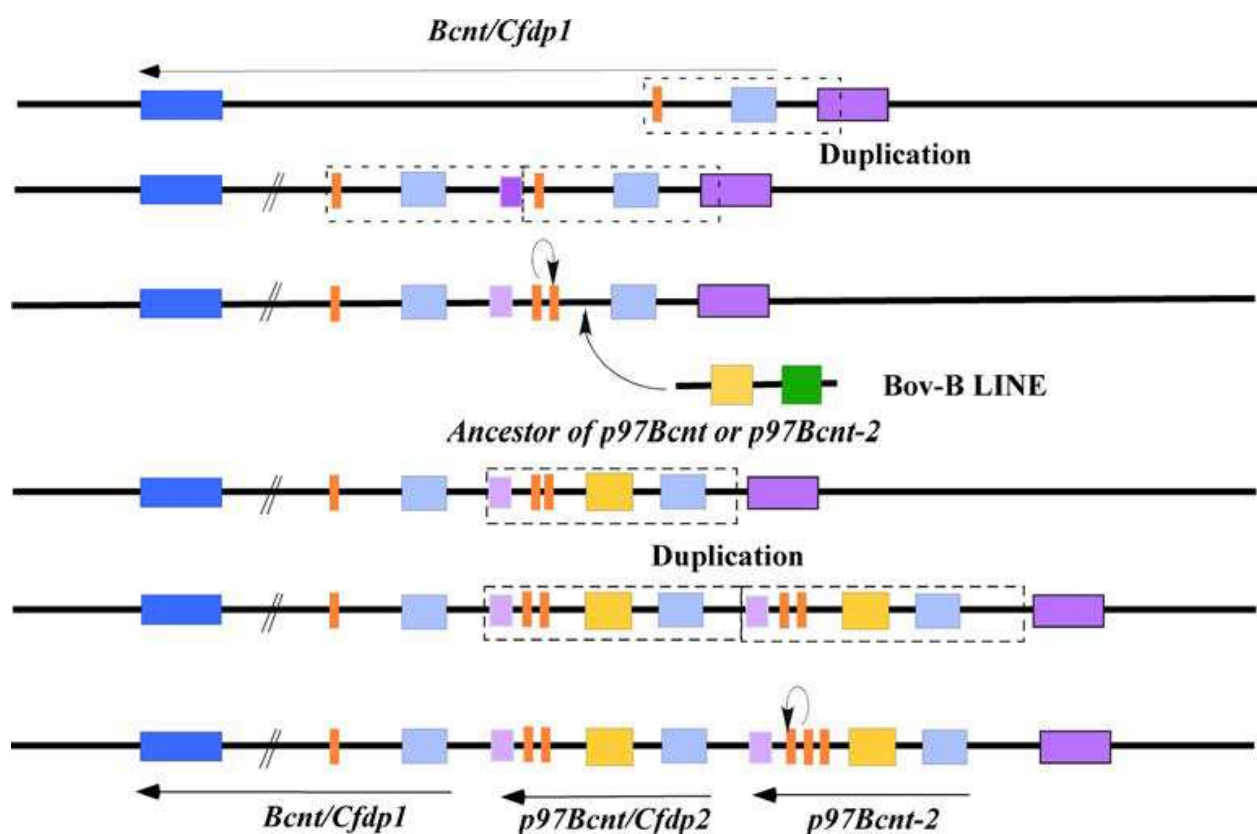
Fig. 5. A scenario for the creation of the two paragolous genes, *p97Bcnt/Cfdp2* and *p97Bcnt-2*

A parsimonious scenario for the creation of the three *Bcnt*-related family genes includes 5 steps as follows: (1) partial gene duplication of the ancestral *Bcnt/Cfdp1*, leaving the Bcnt-C region by segmental duplication; (2) insertion of a Bov-B LINE in intron 5 of one of the duplicated copies, recruitment of the AP-endonuclease domain of the retrotransposon, and generation of the ancestor of *p97Bcnt/Cfdp2* or *p97Bcnt-2*; (3) segmental duplication of the IR unit of ancestor *p97Bcnt*; (4) further gene duplication of the ancestor *p97Bcnt* to generate the nascent *p97Bcnt-2*; and, finally, (5) segmental duplication of the IR unit of the nascent *p97Bcnt-2* to create *p97Bcnt-2*. Nucleotide regions corresponding to the acidic N-terminal regions, IR units, Bcnt-C, and Tmem170A, are symbolically indicated by boxes colored grey, orange, dark blue, and purple, respectively. Bov-B LINE has an AP-endonuclease domain (in yellow) and reverse transcriptase domain (in green)

4. Characteristics of three *Bcnt*-related gene products

4.1 Ancestral Bcnt protein with highly conserved C-terminal region (Bcnt-C)

It has been proposed that duplicated genes yield genetic redundancy, which should result in either the acquisition of a gene with a novel function or the degeneration of one of the duplicated genes. Two paralogs, *p97Bcnt/Cfdp2* and *p97Bcnt-2* genes, were created via a process of tandem duplication followed by retrotransposon insertion. We expect that the three Bcnt-related proteins may play a role in more refined cellular signaling in ruminants. The vertebrate Bcnt/Cfdp1 protein includes a highly conserved 82-amino acid region at the C-terminus, termed Bcnt-C, which is not present in either p97Bcnt/Cfdp2 or p97Bcnt2 (Iwashita et al., 2003; 2009) (Fig. 2). Bcnt-C, known as the BCNT superfamily, is found in most eukaryotes, including yeast, and is classified into Pfam 07572 in the Pfam database (http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=pfam07572). Although the functions of the BCNT family members remain mostly unclear, a vertebrate Bcnt/Cfdp1 was recently identified as a centromere protein, CENP-29, in DT40 cells, a chicken B cell line transformed by avian leukosis virus (Ohta et al., 2010). Furthermore, a yeast ortholog, Swc5 /YBR231C/AOR1, is a component of the chromatin-remodeling complex SWR1 in *Saccharomyces cerevisiae* (budding yeast)(Wu et al., 2009). The SWR1 complex mediates the ATP-dependent exchange of histone H2A for the H2A variant HZT1, and the Swc5 null mutant shows phenotypes of decreased resistance to macromolecule synthesis inhibitors such as hydroxyurea and cycloheximide, and increased heat sensitivity in budding yeast. These data indicate that the yeast *Bcnt* ortholog is not essential for survival, but contributes to maintaining physiological homeostasis at the transcriptional level.

Whereas Bcnt-C is highly conserved among almost all eukaryotes, the N-terminal regions are less conserved. For example, the amino acids in Drosophila Bcnt (YETI) are ~50% identitical to those of bovine Bcnt/Cfdp1 in the C-terminal region, while the N-terminal region shows only ~22 % identity. Thus, although YETI is reported to bind to microtubule-based motor kinesin-I (Wisniewski, et al., 2003), a reevaluation is needed to confirm whether vertebrate Bcnt functions in intracellular trafficking because its interaction is mediated via its N-terminal region. One characteristic of the three Bcnt-related proteins is their different numbers of IR units: Bcnt/Cfdp1, p97Bcnt/Cfdp2, and p97Bcnt2 have one, two, and three IR units, respectively. Sequences homologous to the 40 amino acid IR unit are found in zebra fish (*Danio rerio*) and nematodes, but not in yeast. These IR units comprise intrinsically disordered regions that might present scaffolds for protein-protein interaction as described later.

4.2 Intrinsic disorder of the three Bcnt-related proteins and cellular localization

The three Bcnt-related proteins move more slowly in sodium dodecyl sulfate acrylamide gel electrophoresis (SDS-PAGE) than expected, resulting in apparently higher molecular masses than those calculated. For example, bovine brain Bcnt/Cfdp1 has 297 amino acids and a calculated molecular mass of 33.3 kDa, but appears around 45 kDa in SDS-PAGE (Fig. 6). The situation is exactly the same for both the p97Bcnt/Cfdp2 and p97Bcnt2 proteins, which have calculated molecular masses of 66. 3 and 70.8 kDa, respectively (Iwashita et al., 2003, 2009). This might be caused by their physical properties in that the three Bcnt-related proteins are intrinsically disordered proteins (IDPs). It has been shown that many biologically active proteins lack a stable three-dimensional (3-D) structure; such proteins are referred to as IDPs (Dunker et al., 2008). IDPs are common to the three domains of life, and, especially in multicellular eukaryotic proteins, account for more than 70% of total proteins. They are involved in the regulation of various signalings through protein-protein
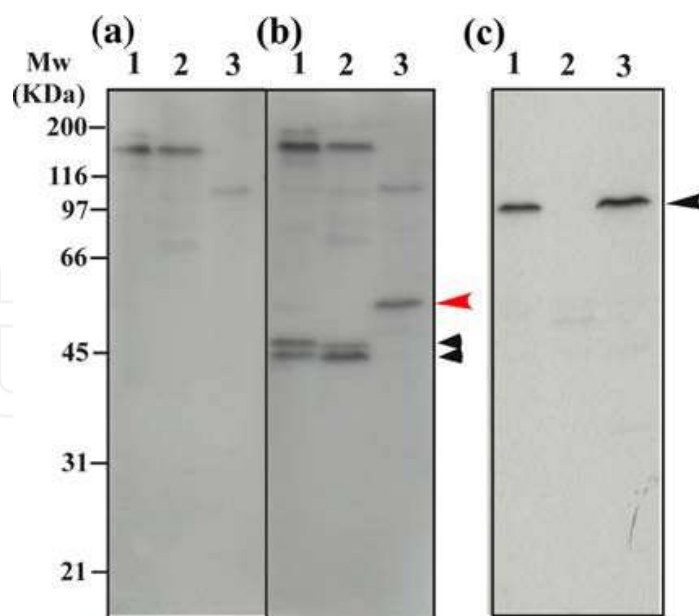
Fig. 6. Unique mobility of the Bcnt/Cfdp1 and p97Bcnt/Cfdp2 proteins in SDS-PAGE

Extracts of bovine brain (1), rat brain (2) and MDBK cells, a bovine kidney epithelial cell line (3) were separated in SDS polyacrylamide gels and subjected to immunoblotting with anti-Bcnt-C peptide antibody in the presence (**a**) or absence (**b**) of antigen peptide at a final concentration of 100 μM, or with anti-p97Bcnt monoclonal antibodies (**c**). The two small black arrows indicate Bcnt/Cfdp1 with an apparent molecular mass of 45 kDa appearing as a doublet, probably due to phosphorylation (Iwashita et. al., 2003); the red large arrow indicates Bcnt/Cfdp1 with an apparent molecular mass of 53 kDa as described below

interactions that are frequently triggered by posttranslational modifications within the regions of intrinsic disorder (Dunker et al., 2008). IDPs may function as hub proteins via the formation of complexes with cellular proteins, which are then modulated by protein modifications such as phosphorylation, acetylation, ubiquitination, or degradation.

By computational prediction, Bcnt/Cfdp1, p97Bcnt/Cfdp2, and p97Bcnt2 are all suggested to comprise intrinsically disordered regions, except for the core of RTE domains that correspond to AP-endonuclease in the two paralogs (Fig. 7). This computational prediction is partially supported by an NMR study of the 3-D structure of the N-terminal 40 amino acid residues of the Bcnt-C region prepared in *Escherichia coli* using ^{15}N-labeled amino acids. The spectrum revealed a lack of fixed tertiary structure (courtesy of Dr. T. Kohno). Furthermore, the Bcnt/Cfdp1 protein forms a tight protein complex with cellular proteins in bovine placenta even in the presence of a detergent, CHAPSO, when evaluated by gel filtration chromatography on Sephacryl S-300 HR followed by western blotting. Both Bcnt/Cfdp1 and p97Bcnt/Cfdp2 are phosphoproteins that are potentially phosphorylated on serine residues by casein kinase II *in vitro* (Iwashita et. al., 1999). Recently, the two phosphorylated serine residues in human BCNT, [116]S in the N-terminal region and [250]S in the C-terminal region, were identified by mass spectrometric analysis (Dephoure et al., 2008). This phosphorylation is cell cycle independent. It should be noted that these two phosphorylated serine residues reside in amino acid sequences WASF and WESF, respectively, which implies a unique motif for specific phosphorylation. Phosphorylation on these motifs might be expected to play a role in switching, such as switching the cation-π mediated protein-ligand interaction (Zacharias & Dougherty, 2002). These characteristics suggest that the three Bcnt-related family members are hub-like molecules.
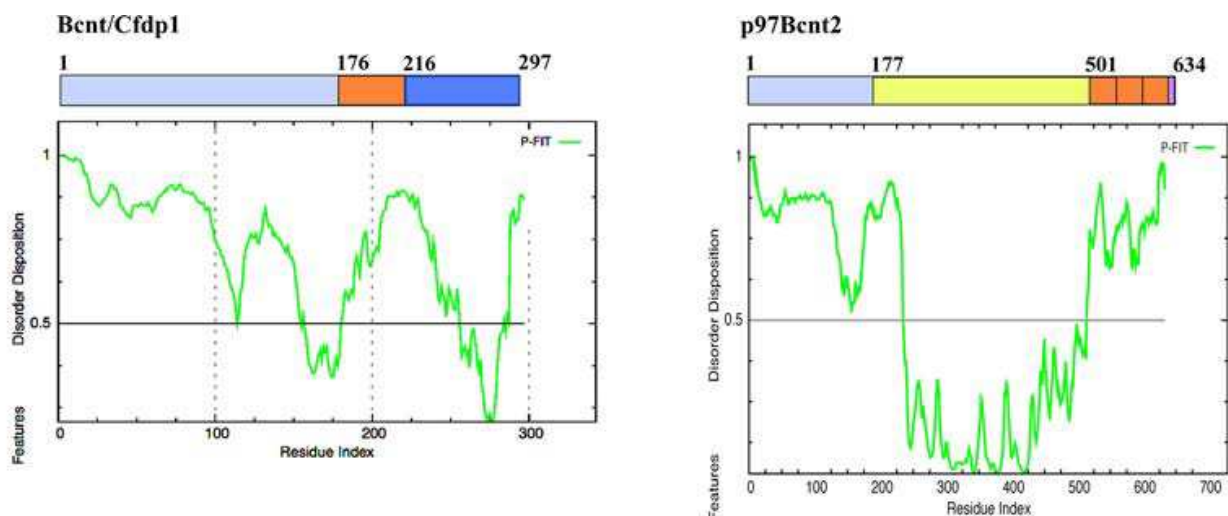
Fig. 7. Characteristics of intrinsic disorder of three Bcnt-related proteins

The Bcnt/Cfdp1, p97Bcnt/Cfdp2, and p97Bcnt2 proteins are predicted to comprise intrinsically disordered regions. Amino acid sequences of the three Bcnt-related proteins were subjected to analysis by a soft server of DisProt (Sickmeier et al., 2007), and individual profiles were obtained. The data for p97Bcnt/Cfdp2 are not shown, but are quite similar to those for p97Bcnt2. Vertical axes indicate the disorder probability of each amino acid residue, and the horizontal axes indicate the number of amino acid residues. Schematic domain structures of Bcnt/Cfdp1 and p97Bcnt2 are shown for each profile for comparision. Similar results were obtained using another program, Anchor (Mészáros et al., 2009)

We have found that the Bcnt/Cfdp1 protein from MDBK cells, a bovine kidney epithelial cell line (Madin & Darby, 1965; Iwashita et al., 1999), migrates at around 53 kDa in SDS-PAGE, significantly bigger than the rat or bovine brain proteins (Fig. 6). The same shift is observed in many other ruminant organs such as bovine placenta, testis and goat kidney, but not in all rat organs. Although we have yet not determined the cause, clarification of this anomaly could shed light on the role of Bcnt/Cfdp1, because the modification may be related to Bcnt/Cfdp1 function. Whereas the~175 amino acid N-terminal regions of the three Bcnt-related proteins are acidic as a whole, they contain several arginine/lysine-rich elements, including a putative nuclear targeting motif of Arg-Lys-Arg-Lys (~61-64th). Therefore we examined the cellular distribution of the three Bcnt-related proteins in MDBK cells. The three were localized in both the cytosolic and nuclear fractions, and, in addition, both p97Bcnt/Cfdp2 and p97Bcnt2 were found significantly in the chromatin fractions (Fig. 8). These results suggest that Bcnt family members have the potential to function as shuttle molecules between the cytosol and nuclei. The nuclear localizations of p97Bcnt/Cfdp2 and p97Bcnt2 are consistent with their protein structure domains; the two paralogs include AP-endonuclease domains in the middle of the molecule as described in more detail below. On the other hand, either the 45 kDa (all rat organs and bovine brain) or 53 kDa (MDBK cells) Bcnt/Cfdp1 is scarcely found in the chromatin fraction, although chicken Bcnt/Cfdp1 has been reported as a centromere protein in a transformed cell line (Ohta et al., 2010).

4.3 RTE domains of p97Bcnt/Cfdp2 and p97Bcnt2
AP-endonuclease is well known to function as an abasic endonuclease in the base excision repair pathway. It possesses multiple enzymic activities as a 3'-5' DNA exonuclease,
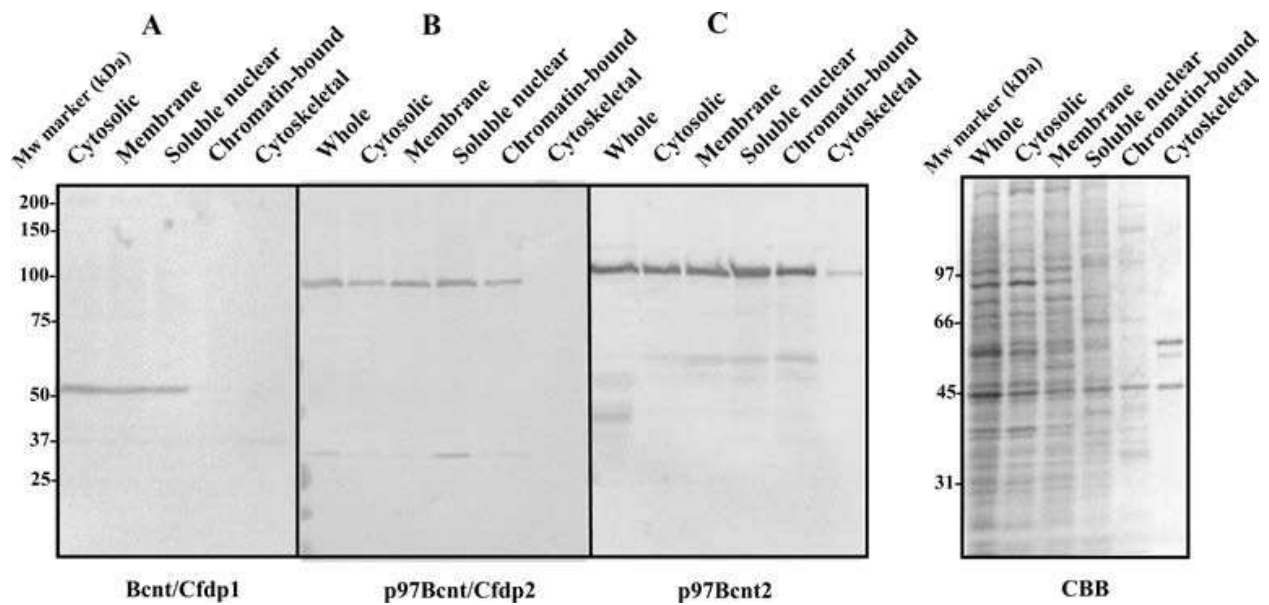
Fig. 8. Subcellular distribution of the three Bcnt-related proteins in MDBK cells

Subcellular fractionation of cultured MDBK cells was carried out successively using a Subcellular Protein Fractionation Kit from Pierce. Constant volume amounts of each fraction were assessed by immunoblotting. A: anti-Bcnt-C peptide antibody (Iwashita et al., 2003), B: anti-p97Bcnt monoclonal antibodies (Nobukuni et al., 1997) and C: anti-p97Bcnt2 peptide antibody (Iwashita et al., 2009). The right panel shows the Coomassie Brilliant Blue staining pattern. The subcellular fractions are identified at the top of the panels. The effectiveness of cellular fractionation was evaluated by immunoblotting using three antigens; anti-p120GAP (a marker for the cytosolic fraction, Kobayashi et al., 1993), anti-Topoisomerase II (a marker for the nuclear fraction, Iwashita et al., 1999), and anti-actin (a marker for the cytoskeleton). The data are consistent with previously reported results (not shown)

3'-phosphodiesterase, 3'-phosphatase, and RNase H (Barzilay et al., 1995). Many organisms possess two functional AP-endonucleases, which are thought to be important for cell viability. In contrast to non-vertebrate AP-endonuclease, vertebrate AP-endonuclease, which has an extra 6 kDa N-terminal region of intrinsic disorder, plays a role not only in repairing DNA damage, but also in regulating the redox state of various proteins that modulate transcription factors such as AP-1 (Fos/Jun), NF-κB, HIF-1, and p53 (Tell et al., 2009; Busso et al., 2010); thus it is termed APE/Ref-1 (AP-endonuclease/Redox effector factor 1). This is natural considering that DNA damage is one of the most vital stresses faced by living organisms. The extra N-terminal region of human AP-endonuclease (APE1) contains multiple arginine/lysine rich elements, and provides a scaffold for protein-protein interaction for DNA repair proteins such as Pol B and XRCC, and transcription factors including STAT3, YB-1, and nucleophosmin (NPM1) (Vascotto et al., 2009; Busso et. al., 2010). Although we have not yet found evidence that p97Bcnt/Cfdp2 and p97Bcnt2 possess any of these activities, they have several characteristics common to mammalian AP-endonuclease with intrinsic disorder regions at both the N- and C-termini. Amino acid sequences in a part of the RTE domains are well conserved in all ruminants so far examined including *Lesser Malay chevrotain* (Iwashita et al., 2009). The central 239-amino acid region of the RTE domain (termed the core RTE domain) corresponds exactly to Endonuclease/Exonuclease/Phosphatase family members (http://www.ncbi.nlm.nih.gov/

cdd?term=Pfam03372). The amino acid sequences of p97Bcnt/Cfdp2 and p97Bcnt2 were compared with those of three canonical AP-endonucleases: human APEX1, *Archaeoglobus* Af_Exo, and *Neiserria* Nape (Fig. 9). Although the comparison revealed low overall identity (~20%) in the core RTE domains, eight amino acid residues involved in catalytic activity and at least 6 amino acids participating in substrate binding are conserved among the molecules. Furthermore, their 3-D structures could be remodeled with high accuracy, revealing the characteristics of Exo III or AP-endonuclease.

```
                     *      *                              * *
p97Bcnt    EYCIGTWNVRSMNPGKLDVVKQEMERINIDILGISELKWTG------MGELNSDDHYIYY 295
p97Bcnt2   EYCIGTWNVRSMNPGKLDVVKQEMERINIDILGISELKWTG------MGELNSDDHYIYY 297
APEX1      TLKICSWNVDGLRAWIKKKGLDWVKEEAPDILCLQETKCSENKLPAELQELPGLSHQYWS 120
Nape       MLKIISANVNGIRSAYKKGFYEYIAASGADIVCVQELKAQEADLSADMKNPHGMHGHWHC 60
Af_Exo     MLKIATFNVNSIRSR-LHIVIPWLKENKPDILCMQETKVENRK----FPEADFHRIGYHV 55

                   * *                                              *
p97Bcnt    CGQQSLRRNGVALIVNKRVRNAIIGCNLKNDRMISVRFQGKPFNLTVIQVYAPTPY-AEE 354
p97Bcnt2   CGQQSLRRNGVAIIVNKRVRNAVIGCNLKNDRMISVRFQGKPFNLTVIQVYAPTPY-APE 356
APEX1      APSDKEGYSGVGLLSRQCPLKVSYGIGDEEHDQEGRVIVAEFDSFVLVTAYVPNAG-RGL 179
Nape       AE--KRGYSGVAVYSKRKPDNVQIGMGIEEFDREGRFVRCDFGRLSVISLYLPSGS-SAE 117
Af_Exo     VFSGSKGRNGVAIASLEEPEDVSFGLDSEPKD-EDRLIRAKIAGIDVINTYVPQGFKIDS 114

                           *                       *
p97Bcnt    GE---VYRFYEDLQHLLE-ITPKIDVLFIIGDWNAKVGSQEIPGITG-RFGLGMQNEAGR 409
p97Bcnt2   PE---VYRFYEDLQHLLE-ITPKIDVLFIIGDWNAKVGSQEIPRITG-KFGLGVQNEAGR 411
APEX1      VRLEYRQRWDEAFRKFLKGLASR-KPLVLCGDLNVAHEEIDLRNPKGNKKNAGFTPQERQ 238
Nape       ERQQVKYRFLDAFYPMLEAMKNEGRDIVVCGDWNIAHQNIDLKNWKGNQKNSGFLPEERE 177
Af_Exo     EKYQYKLQWLERLYHYLQKTVDFRSFAVWCGDMNVAPEPIDVHSPDKLKNHVXFHEDARR 174

                      * *
p97Bcnt    RLIEFCHHNRLVITNTLFQQPSRRLYTWTSP-DGRYRD-----QIDYIICRQRWRSSVQS 463
p97Bcnt2   RLIEFCYQNRLVIANTLFQQHKRRLYTWTSP-DGRYRD-----QIDYIICRQRWRSSVQS 465
APEX1      GFGELLQAVPLADSFRHLYPNTPYAYTFWTY-MMNARSKNVGWRLDYFLLSHSLLPALCD 297
Nape       WIGKVIHKLGWTDMWRTLYPDVP-GYTWWSN-RGQAYAKDVGWRIDYQMVTPELAAKAVS 235
Af_Exo     AYKKILELG-FVDVLRKIHPNER-IYTFYDYRVKGAIERGLGWRGDAILATPPLAERCVD 232

              *
p97Bcnt    AKTRPG---ADCGSDHKLLIAKF----- 483
p97Bcnt2   AKTRPG---ADCGSDHKLLIAKF----- 485
APEX1      SKIRS----KALGSDHCPITLYLAL--- 318
Nape       AHVYK----DEKFSDHAPLVVEYDYAAE 259
Af_Exo     CYADIKPRLAEKPSDHLPLVAVFDV--- 257
```

Fig. 9. Highly conserved amino acid residues of the core RTE domains critical for AP-endonuclease activity

The amino acid sequences of the core RTE domains of p97Bcnt／Cfdp 2 (241-483th) and p97Bcnt2 (243-485th), APEX1 (human APE, 61-318th), Nape (*Neisseria*, 1-259th, Carpenter et al., 2007) and Af_Exo (*Archaeoglobus fulgidus*, 1-257th, Schmiedel et al., 2009) were aligned by the ClustalW2 program of EMBL-EBI. Residues critical for the catalytic activity of canonical AP-endonucleases are shown in red bold, and amino acid substitutions in the core RTE domains between p97Bcnt/Cfdp2 and p97Bcnt2 are indicated in blue bold

The 3-D structure of AP-endonuclease is evolutionarily well conserved and comprises two domains, each containing six-stranded β sheets decorated by helixes on the concave site (Barzilay et al., 1995). The predicted 3-D structures of both RTE domains of p97Bcnt/Cfdp2 and p97Bcnt2 are quite similar to each other, and present possible DNA-binding sites between α-helix domains and opposite both the N-terminal and C-terminal regions of the RTE domain. Next we superimposed the structure of p97Bcnt2 onto that of p97Bcnt/Cfdp2 to examine the structural relationship between the two domains (Fig. 10). Whereas the N-terminal ~60-amino acid region is variable between the two, there are only 12 amino acid differences in the 239-amino acid core RTE domain. It is noteworthy that 7 of the 12 different residues are located in the neighborhood of the predicted active sites. It is characteristic that the enzymatic properties of AP-endonuclease change significantly with subtle changes in the neighborhood of the active cavity. For example, a single amino acid substitution restores

Neisserial AP-endonuclease activity from the exonuclease (Carpenter et al., 2007), and a spontaneous substitution of Val to Gly in the C-terminal Archaeglobus AP-endonuclease, which participates in forming an abasic DNA binding pocket, is accompanied by an increase in non-specific endonuclease activity (Schmiedel et al., 2009). This is probably because AP-endonuclease possesses multiple enzymatic activities as described above. Thus it could be expected that p97Bcnt/Cfdp2 and p97Bcnt2 would have different enzymatic properties, with each compensating for the function of the other.
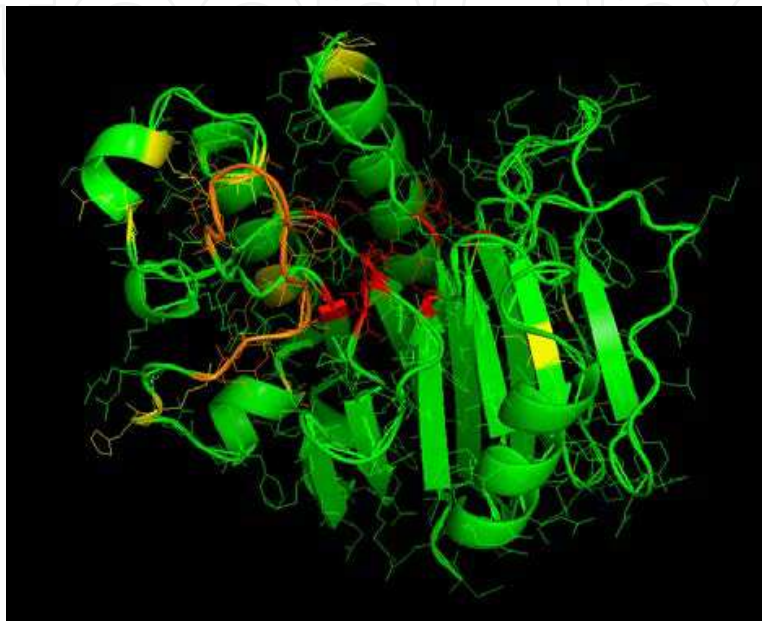

Fig. 10. 3-D comparison of the two core RTE domains of p97Bcnt/Cfdp2 and p97Bcnt2

3-D structures of the two core RTE domains of p97Bcnt/Cfdp2 and p97Bcnt2 were remodeled by the I-TASSER server (Roy et al., 2010), and p97Bcnt2 was superimposed on the p97Bcnt axis fixing the main alpha chain. RMSD (root mean square deviation) of 243 amino acid residues was 0.68 Å. From the data of the top templates of 2jc5A (Carpenter et al, 2007) and 2voaA (Schmiedel et al., 2009), the catalytic sites and DNA binding sites in red, or a loop involved in the targeting site in orange from a template of 2v0rA (Repanas et al., 2007) were identified. Twelve amino acid substitutions between the two domains are shown in yellow. These drawings were obtained using PyMOL. Analysis was carried out courtesy of Dr. M. Tanio, National Institutes of Natural Sciences, Okazaki

To explore further whether the nucleotide substitutions in p97Bcnt2 reflect natural selection in the two paralogs, we examined the d_N (non-synonymous substitution per site)/d_S (synonymous substitution per site) values for both the core RTE domain (243-483th amino acids) and the remaining regions (177-242th and 484-500th amino acids). The d_N/d_S values are 0.029/0.160 in the core region and 0.166/0.414 in the other regions. Although there are more non-synonymous and synonymous substitutions outside the core RTE domain of p97bcnt2, the d_N/d_S values are < 1, suggesting no definite attribution to positive selection. On the other hand, the d_N/d_S value in the core RTE domain is much lower than that of the other regions, suggesting that selective constraints have been substantially strong in the core RTE domain (Iwashita etal., 2009). These data suggest that the recruited RTE domains in both p97Bcnt/Cfdp2 and p97Bcnt2 have played a crucial role in the duplicated novel genes.

5. Perspectives

Living organisms have evolved so as to acquire various anti-stress systems in response not only to exogenous stresses, but also to the intrinsic stresses faced by multicelluar organisms for physiological homeostasis. DNA repair systems in all organisms, immune systems in vertebrates, and the placental systems in mammals are some of the most fruitfully acquired systems. Several lines of evidence have shown that the expression of various integrated retrotransposons is induced by environmental stimuli, such as ultraviolet light, heat shock, or macromolecule synthesis inhibitors (Liu et al., 1995; Morales et al., 2003; Häsler et. al., 2007). Although the induction mechanism of expression has not been fully elucidated, these stresses may enhance promoter activity (Morales et al., 2003) or release the suppressive state of expression, resulting in the creation of new genetic materials. If the retrotransposon induction process is combined with tandem gene duplication, it is a much more efficient way to create a novel gene. Under such stressful conditions, novel genetic materials may play a role in adaptation to new environments.

Among the three Bcnt-related proteins in ruminants, the two paralogs p97Bcnt/Cfdp2 and p97Bcnt2 were generated by partial segmental duplication of the ancestral *Bcnt/Cfdp1* gene, followed by the insertion of an order-specific retrotransposon, resulting in the recruitment of the AP-endonuclease domain of the retrotransposon. Based on the 3-D remodeled structures of these recruited RTE domains and comparison of their protein sequences, they probably retain AP-endonuclease activity. Mammalian AP-endonuclease plays a role not only in direct DNA repair, but also in stimulating pathways for anti-stress activities. Although the latter activity depends mostly on the N-terminal 6-kDa region (Tell et al., 2009; Busso et al., 2010) which is dissimilar to those of p97Bcnt/Cfdp2 and p97Bcnt2, they contain intrinsically disordered regions other than the core RTE domains in both the N-terminal and C-terminal regions, including outside of the core RTE domains. These regions may serve as scaffolds for cellular protein-protein interactions and create novel functions as chimeric genes. The two paralogs distribute in both the cytosolic and nuclear fractions. The properties of intrinsically disordered proteins other than the AP-endonuclease domain and wide cellular distribution are similar to those of the APE1/Ref-1 molecule. Based on these considerations, we conclude that p97Bcnt/Cfdp2 and p97Bcnt2 have recruited the AP-endonuclease domain of a retrotransposon, which originally played an essential role in the integration of the retrotransposon into the genome, and that the two paralogs may have utilized AP-endonuclease activity to suppress cellular stress for survival. The cellular stress that might have induced the retrotransposition of Bov-B LINEs would increase the probability that the newly created genes would become fixed in a population. In addition, because these novel genes have chimeric origins, the original regulatory network of the ancestral *Bcnt/Cfdp1* gene may also have been modified to some extent. Therefore, we hypothesize that the two novel genes have become additional components of pre-existing regulatory networks for anti-stress activities. Although this hypothesis cannot explain why molecules containing the AP-endocnuclease domain, such as p97Bcnt/Cfdp2 and p97Bcnt2, are so rare despite the advantage of being able to regulate cellular activity, it will be intriguing to examine the functions of the three Bcnt-related proteins based on this working hypothesis.

6. Conclusion

The *Bcnt/Cfdp1* gene comprises a unique gene family with three members in ruminants. The two paralogs, *p97Bcnt/cfdp2* and *p97Bcnt-2*, were created in ancient ruminants by a partial

duplication of the ancestral *Bcnt/Cfdp1* gene followed by the insertion of an order-specific retrotransposon, Bov-B LINE. This type of combined process provides great potential to generate a novel gene because a novel function can be acquired under the guarantee of the original gene function. The ancestral Bcnt/Cfdp1 protein contains a highly conserved C-terminus of 82-amino acids (Bcnt-C) that is not present in either p97Bcnt/Cfdp2 or p97Bcnt2. Bcnt-C is found in all eukaryotes where it is known as the BCNT superfamily. A chicken Bcnt/Cfdp1 is a centromere protein while the yeast ortholog is a component of the chromatin-remodeling complex, suggesting that the ancestral Bcnt/Cfdp1 protein plays a role in the regulation of gene expression. The two paralogs, p97Bcnt/Cfdp2 and p97Bcnt2, recruited an AP-endonuclease domain of the retrotransposon during their generation process as a ~325 amino acid region (RTE domain) in the middle of the molecule. The three Bcnt-related proteins distribute in both the cytosolic and nuclear fractions, and include intrinsically disordered regions other than the core of RTE domains of the two paralogs. The 3-D structures of the core RTE domains can be remodeled as canonical AP-endonucleases with identical catalytic amino acid residues. Although as yet there is no direct evidence for it, the two paralogs probably retain AP-endonuclease activity. Because AP-endonuclease/ Redox effector factor 1 is one of the major regulators of cellular responses to various stresses, we propose that the recruited AP-endonuclease domains, which may have emerged in response to cellular stresses, may be utilized by the paralogs in cellular regulation. Therefore, the three Bcnt-related family members provide a good opportunity to examine dynamic changes in signaling networks that accompany novel genes.

7. Acknowledgements

8. References

Barzilay, G., Walker, L.J., Robson, C.N., & Hickson, I.D. 1995. Site-directed mutagenesis of the human DNA repair enzyme HAP1: identification of residues important for AP endonuclease and RNase H activity. *Nucleic Acids Res.*, 23, pp. 1544-1550

Brosius, J. 2005. Echoes from the past—are we still in an RNP world? *Cytogenet. Genome Res.*, 110, pp. 8-24

Busso, C.S., Lake, M.W., & Izumi, T. 2010. Posttranslational modification of mammalian AP endonuclease (APE1). *Cell. Mol. Life Sci.*, 67, pp. 3609-3620

Carpenter, E.P., Corbett, A., Thomson, H., Adacha, J., Jensen, K., Bergeron, J., Kasampalidis, I., Exley, R., Winterbotham, M., Tang, C., Baldwin, G.S., & Freemont, P. 2007. AP endonuclease paralogues with distinct activities in DNA repair and bacterial pathogenesis. *EMBO J.*, 26, pp. 1363-1372

Dephoure, N., Zhou, C., Villén, J., Beausoleil, S.A., Bakalarski, C.E., Elledge, S.J., & Gygi, S.P. 2008. A quantitative atlas of mitotic phosphorylation. *Proc. Natl. Acad. Sci. USA*, 105, pp. 10762-10767

Diekwisch, T.G., Marches, F., Williams, A., & Luan, X. 1999. Cloning, gene expression, and characterization of CP27, a novel gene in mouse embryogenesis. *Gene,* 235, pp. 19-30

Dosztányi, Z., Mészáros, B., & Simon, I. 2009, ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics.* 25, pp. 2745-2746

Dunker, A.K., Oldfield, C.J., Meng, J., Romero, P., Yang, J.Y., Chen, J.W., Vacic, V., Obradovic, Z., & Uversky, V.N. 2008. The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics* 9, Suppl 2:S1.

Fritz, G., & Kaina, B., 1999. Phosphorylation of the DNA repair protein APE/REF-1 by CKII affects redox regulation of AP-1. *Oncogene*, 18, pp. 1033-1040

Gentles, A.J., Wakefield, M.J., Kohany, O., Gu, W., Batzer, M.A., Pollock, D.D., & Jurka, J. 2007. Evolutionary dynamics of transposable elements in the short-tailed opossum Monodelphis domestica. *Genome Res.*, 17, pp. 992-1004

Hastings, P. J.. Lupski, J. R., Rosenberg,S.M., & Ira,G. 2009. Mechanisms of change in gene copy number. *Nature Reviews. Genetics,* 10, pp. 551-564

Iwashita, S., Nobukuni, T., Tanaka, S., Kobayashi, M., Iwanaga, T., Tamate, H.B., Masui, T., Takahashi, I., & Hashimoto, K. 1999. Partial nuclear localization of a bovine phosphoprotein, BCNT, that includes a region derived from a LINE repetitive sequence in Ruminantia. *Biochim. Biophys. Acta*, 1427, pp. 408-416

Iwashita, S., Osada, N., Itoh, T., Sezaki, M., Oshima, K., Hashimoto, E., Kitagawa-Arita, Y., Takahashi, I., Masui, T., Hashimoto, K., & Makalowski, W. 2003. A transposable element-mediated gene divergence that directly produces a novel type bovine Bcnt protein including the endonuclease domain of RTE-1. *Mol. Biol. Evol.*, 20, pp. 1556-1563

Iwashita, S., Ueno, S., Nakashima, K., Song, S.-Y., Ohshima, K., Tanaka, K., Endo, H., Kimura, J., Kurohmaru, M., Fukuta, K., David, L., & Osada, N. 2006. A tandem gene duplication followed by recruitment of a retrotransposon created the paralogous bucentaur gene (bcntp97) in the ancestral ruminant. *Mol. Biol. Evol.*, 23, pp. 798-806

Iwashita, S., & Song, S.-Y. 2008. RasGAPs: a crucial regulator of extracellular stimuli for homeostasis of cellular functions. *Mol. BioSyst.*, 4, pp. 213-222

Iwashita, S., Nakashima, K., Sasaki, M., Osada, N., & Song, S.-Y. 2009. Multiple duplication of the bucentaur gene family, which recruits the APE-like domain of retrotransposon: Identification of a novel homolog and distinct cellular expression. *Gene*, 435, pp. 88-95

Jobse, C., Buntjer, J.B., Haagsma, N., Breukelman, H.J., Beintema, J.J., & Lenstra, J.A. 1995. Evolution and recombination of bovine DNA repeats. *J. Mol. Evol.*, 41, pp. 277-283

Kaessmann, H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Res.* 20, pp. 1313-1326

Kobayashi, M., Hashimoto, N., Hoshino, M., Hattori, S., & Iwashita, S. 1993. Differential contribution of Mr 120 kDa rasGTPase-activating protein and neurofibromatosis type 1 gene product during the transition from growth phase to arrested state in human fibroblasts accompanied by a unique rasGTPase-activating activity. *FEBS Lett.* 327, pp. 177-182

Liu, W.M., Chu, W.M., Choudary, P.V., & Schmid, C.W. 1995. Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic Acids Res.*, 23, pp. 1758-1765

Liu, G.E., Ventura, M., Cellamare, A., Chen, L., Cheng, Z., Zhu, B., Li, C., Song, J., & Eichler, E.E. 2009. Analysis of recent segmental duplications in the bovine genome. *BMC Genomics.*, 10, pp.571

Makalowski, W. 2003. Not junk after all. *Science*, 300, pp. 1246-1247

Malik, H.S., & Eickbush, T.H. 1998. The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINEs. *Mol. Biol. Evol.* 15, pp. 1123-1134

Madin, S.H., & Darby, N.B.Jr. 1958. Established kidney cell lines of normal adult bovine and, ovine origin. *Proc. Soc. Exp. Biol. Med.* 98, pp. 574-576

Morales, J.F., Snow, E.T., & Murnane, J.P. 2003. Environmental factors affecting transcription of the human L1 retrotransposon. II. Stressors. *Mutagenesis*, 18, pp. 151- 158

Nobukuni, T., Kobayashi, M., Omori, A., Ichinose, S., Iwanaga, T., Takahashi, I., Hashimoto, K., Hattori, S., Kaibuchi, K., Miyata, Y., Masui, T., & Iwashita, S. 1997. An Alu-linked repetitive sequence corresponding to 280 amino acids is expressed in a novel bovine protein, but not in its human homologue. *J. Biol. Chem.*, 272, pp. 2801-2807

Ohno, S. 1970. *Evolution by gene duplication.* Springer-Verlag, New York

Ohta, S., Bukowski-Wills, J.-C., Sanchez-Pulido, L., Alves, de Lima Alves, F., Wood, L., Chen, Z.A. Platani, M., Fischer, L., Hudson, D. F., Ponting, C.P., Fukagawa, T., Earnshaw, W. C., & Rappsilber, J. 2010. The protein composition of mitotic chromosomes determined using multiclassifier combinatorial proteomics. *Cell,* 142, pp. 810–821

Rawn, S.M., & Cross, J.C. 2008. The evolution, regulation, and function of placenta-specific genes. *Annu. Rev. Cell Dev. Biol.*, 24, pp. 159-181

Repanas, K., Zingler, N., Layer, L.E., Schumann, G.G., Perrakis, A., & Weichenrieder, O. 2007. Determinants for DNA target structure selectivity of the human LINE-1 retrotransposon endonuclease. *Nucleic Acids Res.*, 35, pp. 4914-4926

Roy, A., Kucukural, A., & Zhang, Y. 2010, I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.*, 5, pp. 725-738

Schmiedel, R., Kuettner, E.B., Keim, A., Sträter, N., & Greiner-Stöffele, T. 2009. Structure and function of the abasic site specificity pocket of an AP endonuclease from *Archaeoglobus fulgidus*. *DNA Repair (Amst)*, 8, pp. 219-231

Sickmeier, M., Hamilton, J.A., LeGall, T., Vacic, V., Cortese, M.S., Tantos, A., Szabo, B, Tompa, P., Chen, J., Uversky, V.N., Obradovic, Z., & Dunker, A.K. 2007. DisProt: the Database of Disordered Proteins. *Nucleic Acids Res.*, 35, D786-793

Sorek, R. 2007. The birth of new exons: Mechanisms and evolutionary consequences. *RNA*, 13, pp. 1603–1608

Szemraj, J., Plucienniczak, G., Jaworski, J., & Plucienniczak, A. 1995. Bovine Alu-like sequences mediate transposition of a new site-specific retroelement. *Gene*, 152, pp. 261-264

Takahashi, I., Nobukuni, T., Ohmori, H., Kobayashi, M., Tanaka, S., Ohshima, K., Okada, N., Masui, T., Hashimoto, K., & Iwashita, S. 1998. Existence of a bovine LINE repetitive insert that appears in the cDNA of bovine protein BCNT in ruminant, but not in human, genomes. *Gene*, 211, pp. 387-394

Tell, G., Quadrifoglio, F., Tiribelli, C., & Kelley, M.R. 2009. The many functions of APE1/Ref-1: not only a DNA repair enzyme. *Antioxid. Redox Signal.*, 11, pp. 601-620

The Bovine Genome Sequencing and Analysis Consortium. 2009. The Genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*, 324, pp. 522-528

Vascotto, C., Cesaratto, L., Zeef, L.A., Deganuto, M., D'Ambrosio, C., Scaloni, A., Romanello, M., Damante, G., Taglialatela, G., Delneri, D., Kelley, M.R., Mitra, S., Quadrifoglio, F., & Tell, G. 2009. Genome-wide analysis and proteomic studies reveal APE1/Ref-1 multifunctional role in mammalian cells. *Proteomics*, 9, pp. 1058-1074

Wisniewski, T.P., Tanzi, C.L., & Gindhart, J.G. 2003. The Drosophila kinesin-I associated protein YETI binds both kinesin subunits. *Biol. Cell*, 95, pp. 595-602

Wu, W.-H., Wu, C.-H., Ladurner, A., Mizuguchi, G., Wei, D., Xiao, H., Luk, E., Ranjan, A., & Wu, C. 2009. N-terminus of Swr1 binds to histone H2AZ and provides a platform for subunit assembly in the chromatin remodeling complex. *J. Biol. Chem.*, 284, pp. 6200–6207

Zacharias, N., & Dougherty, D.A. 2002. Cation-π interactions in ligand recognition and catalysis. *Trends Pharmacol. Sci.*, 23, pp. 281-287

Zhang, J. 2003. Evolution by gene duplication: an update. *Trends Ecol. Evol.*, 18, pp. 292-298

Zhang, J., Dean, A.M., Brunet, F., & Long, M. 2004. Evolving protein functional diversity in new genes of *Drosophila*. *Proc. Natl. Acad. Sci. USA*, 101, pp. 16246-16250

Zingler, N., Weichenrieder, O., & Schumann, G.G. 2005. APE-type non-LTR retrotransposons: determinants involved in target site recognition. *Cytogenet. Genome Res.*, 110, pp. 250-268

Zupunski, V., Gubensek, F., & Kordis, D. 2001. Evolutionary dynamics and evolutionary history in the RTE clade of non-LTR retrotransposons. *Mol. Biol. Evol.*, 18, pp. 1849-1863
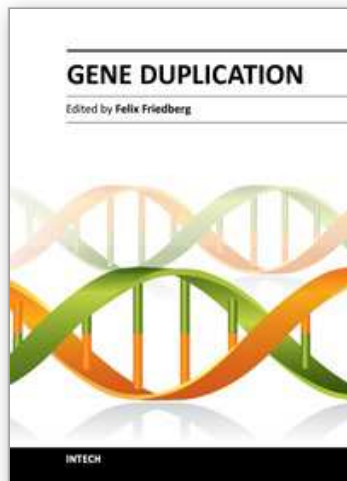
Gene Duplication

Edited by Prof. Felix Friedberg

The book Gene Duplication consists of 21 chapters divided in 3 parts: General Aspects, A Look at Some Gene Families and Examining Bundles of Genes. The importance of the study of Gene Duplication stems from the realization that the dynamic process of duplication is the "sine qua non" underlying the evolution of all living matter. Genes may be altered before or after the duplication process thereby undergoing neofunctionalization, thus creating in time new organisms which populate the Earth.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Shintaro Iwashita and Naoki Osada (2011). Bucentaur (Bcnt) Gene Family: Gene Duplication and Retrotransposon Insertion, Gene Duplication, Prof. Felix Friedberg (Ed.), ISBN: 978-953-307-387-3, InTech, Available from: http://www.intechopen.com/books/gene-duplication/bucentaur-bcnt-gene-family-gene-duplication-and-retrotransposon-insertion

INTECH
open science | open minds