ISSN 1672-9145 **Acta Biochimica et Biophysica Sinica** 2005, **37**(3): 186–191 **CN** 31-1940/Q

Isolation, Sequence Analysis and Expression Profile of a Novel Swine Gene Differentially Expressed in the Longissimus Dorsi Muscle Tissues from Landrace×Large White Cross-combination

Yong-Gang LIU, Yuan-Zhu XIONG*, and Chang-Yan DENG

Key Laboratory of Swine Genetics and Breeding, Ministry of Agriculture, Huazhong Agricultural University, Wuhan 430070, China

Abstract The mRNA differential display technique was performed to investigate the differences in gene expression in the Longissimus dorsi muscle tissues from Landrace×Large White cross-combination. One novel gene that was differentially expressed was identified using semi-quantitative reverse transcriptase-polymerase chain reaction (RT-PCR) and its complete cDNA sequence was obtained using the rapid amplification of cDNA ends (RACE) method. The nucleotide sequence of the gene is not homologous to any of the known porcine genes. The sequence prediction analysis revealed that the open reading frame of this gene encodes a protein of 260 amino acids that contains the putative conserved domain of the carbonic anhydrase, and this protein has high homology with the carbonic anhydrase III (CA-III) of four species—mouse (91%), horse (91%), rat (89%) and human (86%)—so that it can be defined as swine carbonic anhydrase III. The phylogenetic tree analysis revealed that the swine CA-III has a closer genetic relationship with the horse CA-III than with those of mouse, rat and human. The tissue expression analysis indicated that the swine CA-III gene is generally expressed in most tissues. Our experiment is the first to establish the primary foundation for further research on the swine CA-III gene.

Key words mRNA differential display; rapid amplification of cDNA ends (RACE); novel gene; carbonic anhydrase III (CA-III)

The mRNA differential display technique first described by Liang and Pardee [1] is a fast and efficient method for isolating and characterizing altered gene expression in different cell types. It has been statistically shown that 80–120 primer combinations are sufficient to cover all the transcript populations in a cell [2]. This technique possesses the following advantages over other similar techniques: it is based on simple and established methods; more than two samples can be compared simultaneously; and only a small amount of starting material is needed [3]. Our previous experiment used 90 primer combinations to perform the differential display polymerase chain reaction (DD-PCR) and silver stain display to analyze the gene expression differences in two pig purebreds, Large White (LW) and Landrace (LD), and two types of reciprocal F_1 hybrids, Large White (♂)×Landrace (♀) (LWLD) and Landrace (♂)×Large White (♀) (LDLW). The results revealed that many gene expression differences exist in the gene expression of this high heterosis cross-combination between the purebreds and hybrids [4].

In the present study, we conducted further research on those genes that are differentially expressed between the purebreds and hybrids. We isolated their full-length cDNA sequences and did some necessary sequence analysis. It is necessary to clarify why these genes are differentially expressed between the parents and hybrids.

Materials and Methods

Sample collection and total RNA extraction

The cross-breeding population, Landrace×Large White,

Received: October 18, 2004 Accepted: January 22, 2005

This work is supported by a grant from the National Basic Research Program of China (No. G2000016105).

*Corresponding author: Tel, 86-27-87287390; Fax, 86-27-87394184; E-mail, liuyg4567@163.com

was constructed in March, 2002. All the Longissimus dorsi muscle samples were collected from 180-day-old pigs slaughtered in August, 2002. For each breed, the total RNA was extracted from five male and five female pigs using the Total RNA extraction kit (Gibco, USA). Before the first-strand cDNA synthesis, DNase I treatment of the total RNA was necessary if the total RNA samples were seriously contaminated with genomic DNA.

Differential display

The mRNA differential display and the silver stain display were performed as previously described elsewhere [4].

Semi-quantitative RT-PCR

Semi-quantitative RT-PCR was performed as previously described elsewhere [5,6]. To eliminate the effect of cDNA concentration, we repeated the RT-PCR five times using 1 μl, 2 μl, 3 μl, 4 μl, and 5 μl cDNA as templates, respectively. We selected the housekeeping gene *G3PDH* (glyceraldehyde-3-phosphate dehydrogenase) as the internal control. The control primers used were: 5′-ACCACAGTCCATGCCATCAC-3′ (*G3PDH* 5′ primer) and 5′-TCCACCACCCTGTTGCTGTA-3′ (*G3PDH* 3′ primer).

To ensure that no false positive PCR fragments were generated from pseudogenes in the contaminating genomic DNA, *G3PDH* primers were derived from different exons in the same gene. PCR primer combinations were tested using porcine genomic DNA as a negative control and an approximately 780 bp PCR fragment was amplified when cDNA was contaminated by genomic DNA. The following expressed sequence tag (EST) or gene specific primers were used to perform the PCR for identification and tissue expression profile analysis: 5′-AATTTCGCACAACTTCAG -3′ (forward primer) and 5′-TGGAAAGATAGCAGAGGC-3′ (reverse primer). The PCR reactions were optimized for a number of cycles to ensure product intensity within the linear phase of amplification.

5′- and 3′-RACE

5′- and 3′-RACE were performed using BD SMART™ RACE cDNA amplification kit (BD Biosciences, USA). The gene-specific primers (GSPs) were: 5′-TCATTCACTCCTGCGTGTTCTAGGA-3′ (5′-RACE GSP) and 5′-TGATTCCTAGAACACGCAGGAGTGA -3′ (3′-RACE GSP).

Sequence analysis

The cDNA sequence prediction was conducted using GenScan software (http://genes.mit.edu/GENSCAN.html). The protein prediction and analysis were performed using the Conserved Domain Architecture Retrieval Tool of BLAST at the National Center for Biotechnology Information (NCBI) server (http://www.ncbi.nlm.nih.gov/BLAST) and the ClustalW software (http://www.ebi.ac.uk/clustalw).

Results and Discussion

mRNA differential display

From the mRNA differential display, one gene, nominated as Gene 28, was found to be overexpressed in the Longissimus dorsi muscle of LW and LDLW, but moderately expressed in the Longissimus dorsi muscle of LD and LWLD as shown in **Fig. 1**.

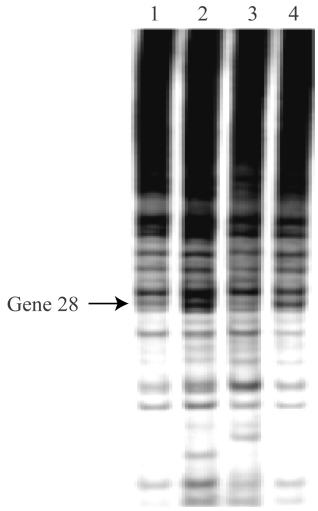

Fig. 1 The differential expression analysis of Gene 28
1, LD; 2, LDLW; 3, LWLD; 4, LW.

Semi-quantitative RT-PCR

The differentially expressed gene band was recovered from gel and used as the template for the re-amplification, which was performed with the corresponding oligo(dT) primer and the arbitrary primers used in the mRNA dif ferential display. The resulting PCR product was 550 bp. This was in agreement with the result of the mRNA differential display. The purified PCR product was then cloned into the T-vector and the recombinant plasmid was sequenced. Semi-quantitative RT-PCR was then conducted

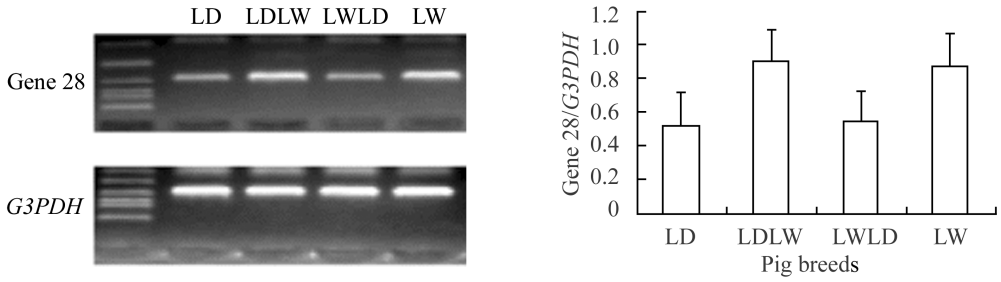

Fig. 2 Semi-quantitative RT-PCR of Gene 28

The Gene 28/*G3PDH* ratios are the averages of five semi-quantitative RT-PCRs using 1 µl, 2 µl, 3 µl, 4 µl and 5 µl cDNA as templates. The signals of the PCR product were measured by BandScan software version 4.50 (http://www.Glyco.com).

using the EST specific primers and the results are presented in **Fig. 2**.

The semi-quantitative RT-PCR results indicated that Gene 28 was overexpressed in the Longissimus dorsi muscle of LW and LDLW, but moderately expressed in the Longissimus dorsi muscle of LD and LWLD. This also coincided with the result of the mRNA differential display.

5′- and 3′-RACE

One PCR fragment of ~1.5 kb was amplified by 5′-RACE. The 3′-RACE product was too short to be displayed. These products were then cloned to T-vectors and sequenced. A 1716 bp complete cDNA sequence was finally obtained (**Fig. 3**).

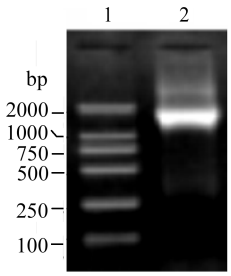

Fig. 3 5′-RACE for Gene 28

1, DL2000 marker; 2, 5′-RACE product.

Sequence analysis

The nucleotide sequence analysis using the BLAST software at NCBI server (http://www.ncbi.nlm.nih.gov/BLAST) revealed that this gene was not homologous to any of the known porcine genes and it was then deposited into the GenBank database (Accession number: AY789645). The sequence prediction was carried out using the GenScan software. An open reading frame encoding 260 amino acids was found in the 1716 bp cDNA sequence.

In the predicted results, the coding region score was 753, probability of exon was 0.996 and the polyA signal was from 1331 bp to 1336 bp. The complete cDNA sequence of this gene (Gene 28) and the encoded amino acids are presented in **Fig. 4**.

The predicted protein was also blasted using the Conserved domain architecture retrieval tool of BLAST at the NCBI server (http://www.ncbi.nlm.nih.gov/BLAST) **(Fig. 5)**, and one putative domain was identified as the carbonic anhydrase (**Fig. 6**).

Further BLAST analysis of this protein revealed that it shared high homology with the carbonic anhydrase III (carbonate dehydratase III) (CA-III) of four species; namely, mouse (91%), horse (91%), rat (89%), human (86%). The alignment of the protein encoded by this new gene and four other kinds of carbonic anhydrase III is shown in **Fig. 7**.

From the results obtained above, it can be concluded that this protein is the swine carbonic anhydrase III, and the new gene is the swine carbonic anhydrase III gene. Based on the results of the alignment of five kinds of carbonic anhydrase III, a phylogenetic tree was constructed using the ClustalW software (http://www.ebi.ac.uk/clustalw), as shown in **Fig. 8**.

Tissue expression profile

The RT-PCR analysis of the tissue expression profile was carried out using the tissue cDNAs of one Large White pig as the templates, and the result revealed that the swine CA-III gene was not only expressed in Longissimus dorsi muscle, but also expressed in the heart, spleen, liver, kidney, small intestine, ovary and lung (**Fig. 9**).

Carbonic anhydrases (CAs) are a large family of zinc

Gene 28:

```
  GAGTACGCGGGGACGGCGTGGAAGAGAAAGCAGCAGCCGTCCAGTGCCCACGAAGACGACCATGGCCAAGGAGTGGGGCTACGCCGACCAC
                                                                M  A  K  E  W  G  Y  A  D  H
AATGGTCCTGACCACTGGCATGAACTTTACCCAATTGCCAAGGGAGACAACCAATCGCCCATTGAACTGCACACTAAGGACATCAAGCACGAC
 N  G  P  D  H  W  H  E  L  Y  P  I  A  K  G  D  N  Q  S  P  I  E  L  H  T  K  D  I  K  H  D
CCTTCTCTGCTGCCCTGGACAGCATCTTATGACCCTGGCTCTGCCAAGACCATCCTGAACAATGGGAAGACCTGCAGGGTTGTATTTGATGAT
 P  S  L  L  P  W  T  A  S  Y  D  P  G  S  A  K  T  I  L  N  N  G  K  T  C  R  V  V  F  D  D
ACTTATGACAGGTCAATGCTGAGAGGTGGTCCTCTCACTGCAGCCTACCGGCTTCGCCAGTTTCATCTTCACTGGGGCTCCTCAGATGATCAC
 T  Y  D  R  S  M  L  R  G  G  P  L  T  A  A  Y  R  L  R  Q  F  H  L  H  W  G  S  S  D  D  H
GGATCTGAGCACACTGTGGATGGAGTCAAGTATGCTGCGGAGCTCCATTTGGTTCATTGGAATTCAAAGTATAACAGTTTTGCAACTGCTCTG
 G  S  E  H  T  V  D  G  V  K  Y  A  A  E  L  H  L  V  H  W  N  S  K  Y  N  S  F  A  T  A  L
AAGCACCCTGATGGAGTGGCTGTAGTTGGCATTTTTCTGAAGATAGGACGTGAGAAAGGCGAGTTCCAACTAGTCCTTGATGCATTGGACAAA
 K  H  P  D  G  V  A  V  V  G  I  F  L  K  I  G  R  E  K  G  E  F  Q  L  V  L  D  A  L  D  K
ATTAAGACAAAGGGCAAGGAGGCACCCTTCACGAACTTCAACCCATCCTGCCTGTTCCCTGCCTGCCGGGACTACTGGACCTACCATGGCTCC
 I  K  T  K  G  K  E  A  P  F  T  N  F  N  P  S  C  L  F  P  A  C  R  D  Y  W  T  Y  H  G  S
TTCACCACGCCGCCCTGCGAGGAGTGCATTGTGTGGCTCCTGCTGAAGGAGCCCATCACCGTGAGCTCTGACCAGATGGCCAAGCTGCGAAGC
 F  T  T  P  P  C  E  E  C  I  V  W  L  L  L  K  E  P  I  T  V  S  S  D  Q  M  A  K  L  R  S
CTCTACTCCAGTGCGGAGAACGAGCCCCCTGTCCCCCTAGTGAGGAACTGGCGCCCCCCACAGCCTATCAAGGGCAGGATAGTGAAGGCCTCC
 L  Y  S  S  A  E  N  E  P  P  V  P  L  V  R  N  W  R  P  P  Q  P  I  K  G  R  I  V  K  A  S
TTCAAATGAGGCTGCTGGAGCTTGCCCTCTTCAGGAAAGGAAGCCTGCTACTGCAGAGCTTGGTTCCTTGCCTCCTTTTGGTGCTCCTTATTC
 F  K  *
CAAGTATTATTTCAGTTTTCCACACTGAGCAATGAAGGTGAGAGCTGCAATCACCAAGAACCAAAGTTACTATTGACAGATCTAATACGAAAG
CATTTGGCCTTTGTAAGAATCATCTTTCCTGTAAAAGAAAACTCTTACTAAGTTTCAAAGAAAAAGAAACAGAGATAGAAAAGATGGAGAAAA
ATGGCTGTTGGGCGCCATTTTGTGTCATCTTAAATTTCGCACAACTTCAGTTTTTACTCTTTTCATGTTACTAGTTATCGATCTTAAAGAAAT
AATGAGTAATTCTATATGAGGAGTAGAGGTATATGAAGATCATGTAGCAATTACACATAAGCCAGAAATTAAAATAATTGTGGACGTCAAGAA
TTGCAATGATGCTGTGTGTTTTAACCCTTAAATAAAACTAACTAGAGGTCCATCATTTATTAAAATTGGAGTTCTGTAAAAAAGGGCTGGTTT
TAATAATGAAAGTAACACATTTGACCGGGATTAAAAAAAAAAAAAAAGACCCATTTTTTGAATTTTTTCTTCATAAAGATAGTTGACTTTATCC
TATTTCTCTTTACCTGAAGGAGGGCCATTTATTTTTCTTTTTACTACTTTTATCTTTGCATGCTTATTAAAATAAAAACTGCCTCTGCTATCT
TTCCATTTTGAGGGTGGGATTAATGATTCCTAGAACACGCAGGAGTGAATGACTGCACTTGAAGATAAAGGCAGTTGTATTGATCATCAAAAC
TAATAAAGACGTGAATGCAAAAAAAAAAAAAAAAAAAAAAAAAA
```

Fig. 4 The complete cDNA sequence of Gene 28 (GenBank accession number: AY789645)

* indicates the stop codon.

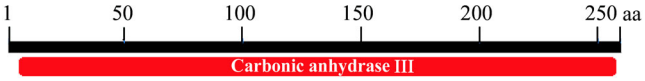

Fig. 5 The putative domain of the protein encoded by Gene 28

aa, amino acid residues. The 3-D structural evidence of the putative conserved domain is also presented in **Fig. 6**.

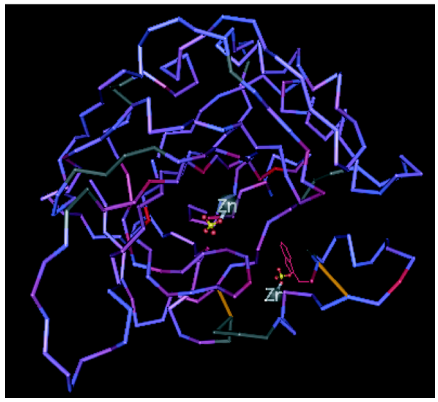

Fig. 6 The 3-D structural evidence of the putative conserved domain

metalloenzymes that catalyze the reversible hydration of carbon dioxide. From the 3-D structure of the putative conserved domain, we can also find two zinc ion binding sites, and this corresponds with the character of the CAs. The CAs participate in a variety of biological processes, including respiration, calcification, acid-base balance, bone resorption and the formation of aqueous humor, cerebrospinal fluid, saliva, and gastric acid. There are three evolutionary unrelated CA gene families entitled α, β and γ. In the animal kingdom, eleven active isoenzymes have been identified and they are all of the α type, and some isoenzymes of the α-enzyme family are expressed in most tissues in the mammalian body [7–10]. The tissue expression profile analysis of the swine CA-III gene revealed that this gene was also expressed in most pig tissues. This implies that our study results are reliable.

Up until now, there have been no reports about the swine gene of carbonic anhydrase. We first obtained the full-length cDNA sequence of the swine CA-III gene and found that it was overexpressed in the Longissimus dorsi muscle

```
Pig     MAKEWGYADHNGPDHWHELYPIAKGDNQSPIELHTKDIKHDPSLLPWTASYDPGSAKTIL   60
Mouse   MAKEWGYASHNGPDHWHELYPIAKGDNQSPIELHTKDIKHDPSLQPWSASYDPGSAKTIL   60
Horse   -AKEWGYADHNGPDHWHEFYPIAKGDNQSPIELHTKDINHDPSLKAWTASYDPGSAKTIL   59
Rat     MAKEWGYASHNGPEHWHELYPIAKGDNQSPIELHTKDIRHDPSLQPWSVSYDPGSAKTIL   60
Human   MAKEWGYASHNGPDHWHELFPNAKGENQSPIELHTKDIRHDPSLQPWSVSYDGGSAKTIL   60

Pig     NNGKTCRVVFDDTYDRSMLRGGPLTAAYRLRQFHLHWGSSDDHGSEHTVDGVKYAAELHL  120
Mouse   NNGKTCRVVFDDTYDRSMLRGGPLSGPYRLRQFHLHWGSSDDHGSEHTVDGVKYAAELHL  120
Horse   NNGRTCRVVFDDTYDRSMLRGGPLTAPYRLRQFHLHWGSSDDHGSEHTVDGVKYAAELHL  119
Rat     NNGKTCRVVFDDTFDRSMLRGGPLSGPYRLRQFHLHWGSSDDHGSEHTVDGVKYAAELHL  120
Human   NNGKTCRVVFDDTYDRSMLRGGPLPGPYRLRQFHLHWGSSDDHGSEHTVDGVKYAAELHL  120

Pig     VHWNSKYNSFATALKHPDGVAVVGIFLKIGREKGEFQLVLDALDKIKTKGKEAPFTNFNP  180
Mouse   VHWNPKYNTFGEALKQPDGIAVVGIFLKIGREKGEFQILLDALDKIKTKGKEAPFTHFDP  180
Horse   VHWNPKYNTYGGALKQPDGIAVVGVFLKIGREKGEFQLFLDALDKIKTKGKEAPFTNFDP  179
Rat     VHWNPKYNTFGEALKQPDGIAVVGIFLKIGREKGEFQILLDALDKIKTKGKEAPFNHFDP  180
Human   VHWNPKYNTFKEALKQRDGIAVIGIFLKIGHENGEFQIFLDALDKIKTKGKEAPFTKFDP  180

Pig     SCLFPACRDYWTYHGSFTTPPCEECIVWLLLKEPITVSSDQMAKLRSLYSSAENEPPVPL  240
Mouse   SCLFPACRDYWTYHGSFTTPPCEECIVWLLLKEPMTVSSDQMAKLRSLFSSAENEPPVPL  240
Horse   SCLFPTCRDYWTYRGSFTTPPCEECIVWLLLKEPITVSSDQVAKLRSLFSSAENEPPVPL  239
Rat     SCLFPACRDYWTYHGSFTTPPCEECIVWLLLKEPMTVSSDQMAKLRSLFASAENEPPVPL  240
Human   SCLFPACRDYWTYQGSFTTPPCEECIVWLLLKEPMTVSSDQMAKLRSLLSSAENEPPVPL  240

Pig     VRNWRPPQPIKGRIVKASFK      260
Mouse   VGNWRPPQPVKGRVVRASFK      260
Horse   VRNWRPPQPLKGRVVRASFK      259
Rat     VGNWRPPQPIKGRVVRASFK      260
Human   VSNWRPPQPINNRVVRASFK      260
```

Fig. 7 The alignment of the protein encoded by Gene 28 from pig and four other kinds of carbonic anhydrase III

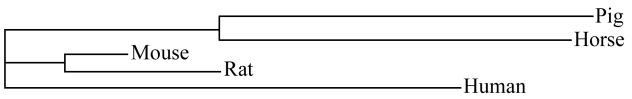

Fig. 8 The phylogenetic tree for five kinds of carbonic anhydrase III

The swine CA-III gene shows a closer genetic relationship with the horse CA-III gene than with those of mouse, rat and human.

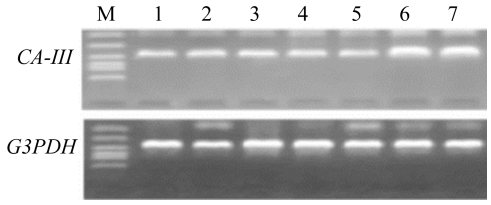

Fig. 9 Tissue expression profile analysis of the swine CA-III gene

M, DL2000 marker; 1, heart; 2, spleen; 3, liver; 4, kidney; 5, small intestine; 6, ovary; 7, lung.

of LW and LDLW, but moderately expressed in the Longissimus dorsi muscle of LD and LWLD. We can also infer that this gene may be affected by the maternal effect.

We have only obtained the cDNA sequence of the novel gene, but the genomic sequence and function of this gene remain largely unknown. To further understand the function of the novel gene, more research based on these primary results is needed.

References

1 Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science, 1992, 257(5072): 967–971
2 Liang P, Averboukh L, Pardee AB. Distribution and cloning of eukaryotic mRNAs by means of differential display: Refinements and optimization. Nucleic Acids Res, 1993, 21(14): 3269–3275
3 Yamazaki M, Saito K. Differential display analysis of gene expression in plants. Cell Mol Life Sci, 2002, 59(8): 1246–1255
4 Liu GY, Xiong YZ, Deng CY, Zuo B, Zhang JH. Comparison of gene expression patterns in Longissimus dorsi of pigs between the high-parent heterosis cross combination Landrace×Large White and the mid-parent heterosis cross combination Large White×Meishan. Asian-Aust J Anim Sci, 2004, 17(9): 1192–1196

5 Daigo Y, Takayama I, Ponder BA, Caldas C, Ward SM, Sanders KM, Fujino MA. Differential gene expression in the murine gastric fundus lacking interstitial cells of Cajal. BMC Gastroenterol, 2003, 3(1): 14
6 Fehr JE, Trotter GW, Oxford JT, Hart DA. Comparison of Northern blot hybridization and a reverse transcriptase-polymerase chain reaction technique for measurement of mRNA expression of metalloproteinases and matrix components in articular cartilage and synovial membrane from horses with osteoarthritis. Am J Vet Res, 2000, 61(8): 900–905
7 Maren TH. Carbonic anhydrase: Chemistry, physiology, and inhibition. Physiol Rev, 1967, 47(4): 595–781
8 Hewett-Emmet D, Tashian RE. Functional diversity, conservation, and convergence in the evolution of the α-, β-, and γ-carbonic anhydrase gene families. Mol Phylogenet Evol, 1996, 5(1): 50–77
9 Tashian RE. The carbonic anhydrases: Widening perspectives on their evolution, expression and function. Bioessays, 1989, 10(6): 186–192
10 Tweedie S, Edwards Y. Mouse carbonic anhydrase III: Nucleotide sequence and expression studies. Biochem Genet, 1989, 27(1-2): 17–30

Edited by
Rong-Jia ZHOU