*Genet. Res., Camb.* (2009), **91**, *pp.* 101–109. © Cambridge University Press 2009
doi:10.1017/S0016672309000032 Printed in the United Kingdom


# Molecular cloning and evolutionary analysis of the *GJA1* (*connexin43*) gene from bats (Chiroptera)

LI WANG[1,2], GANG LI[1]*, JINHONG WANG[1], SHAOHUI YE[2], GARETH JONES[3] AND SHUYI ZHANG[1]

[1] *School of Life Sciences, East China Normal University, Shanghai 200062, China*
[2] *College of Animal Science, Yunnan Agricultural University, Kunming 650201, China*
[3] *School of Biological Science, University of Bristol, Woodland Road, Bristol BS8 IUG, UK*

(*Received 18 June 2008 and in revised form 5 January 2009*)

**Summary**

Gap junction protein connexin43 (Cx43), encoded by the *GJA1* gene, is the most abundant connexin in the cardiovascular system and was reported as a crucial factor maintaining cardiac electrical conduction, as well as having a very important function in facilitating the recycling of potassium ions from hair cells in the cochlea back into the cochlear endolymph during auditory transduction processes. In mammals, bats are the only taxon possessing powered flight, placing exceptional demand on many organismal processes. To meet the demands of flying, the hearts of bats show many specialties. Moreover, ultrasonic echolocation allows bat species to orientate and often detect and locate food in darkness. In this study, we cloned the full-length coding region of *GJA1* gene from 12 different species of bats and obtained orthologous sequences from other mammals. We used the maximum likelihood method to analyse the evolution of *GJA1* gene in mammals and the lineage of bats. Our results showed this gene is much conserved in mammals, as well as in bats' lineage. Compared with other mammals, we found one private amino acid substitution shared by bats, which is located on the inner loop domain, as well as some species-specific amino acid substitutions. The evolution rate analyses showed the signature of purifying selection on not only different classification level lineages but also the different domains and amino acid residue sites of this gene. Also, we suggested that *GJA1* gene could be used as a good molecular marker to do the phylogenetic reconstruction.

## 1. Introduction

In vertebrates, six connexin molecules assemble to form a hemi-channel (connexon), and subsequent docking of two connexons in adjacent cell membranes result in the formation of a complete intercellular channel or gap junction. Gap junctions facilitate the exchange of nutritive material, ions, secondary messengers and small molecules of up to 1 kDa in size directly between adjacent cells (Goodenough *et al.*, 1996; Alexander & Goldberg, 2003). Gap junctions were thought to have a crucial role in the synchronized contraction of the heart and in embryonic development (Britz-Cunningham *et al.*, 1995). Also in the inner ear, gap junctions are divided into the epithelial cell gap junction system and the connective tissue gap junction system among the supporting cells of the organ of Corti. These gap junctions were considered to have a very important function, facilitating the recycling of potassium ions from the hair cells back into the cochlear endolymph during auditory transduction processes (Kikuchi *et al.*, 1995).

As a member of the connexin family, gap junction protein connexin 43 (Cx43) was encoded by the *GJA1* gene that is located on 6q21-q23.2 (Fishman *et al.*, 1990; Corcos *et al.*, 1993). The gene is broadly expressed in many different tissues and organs (Beyer *et al.*, 1989; Musil *et al.*, 1990; Fishman *et al.*, 1991; van der Heyden *et al.*, 2001; Willingham-Rocky *et al.*, 2007). Interestingly, for the working myocardium, *GJA1* gene expression was only found in mammalian

* Corresponding author. e-mail: li_gang1978@yahoo.com.cn

species (Becker *et al*., 1998). Moreover, connexin43 is the most abundant connexin in the cardiovascular system.

Many mutations occur in the different domains of connexin, most of them occurring in highly conserved sites across various species as well as in the human connexin isoforms, and have been linked to specific diseases. Homozygotes for a Leu→Phe substitution in the highly conserved codon 11 and one homozygote for a Val→Ala transversion at the highly conserved codon 24 of *GJA1* gene caused recessive, pre-lingual, profound deafness in four patients, suggesting these mutations could be a common cause of deafness in African–Americans, presumably by disrupting the recycling of potassium to the cochlear endolymph (Liu *et al*., 2001). Several novel polymorphisms, although no disease-associated mutations, were identified in the *GJA1* gene in Turkish families with autosomal-recessive non-syndromic hearing loss (Uyguner *et al*., 2003). Substitution of proline for serine at position 364 of *GJA1* gene in heart disease patients led to abnormally regulated cell–cell communication, associated with visceroatrial heterotaxia (Britz-Cunningham *et al*., 1995). Mutations in the arginine–serine–serine fragment concentrating on the carboxy terminus altered potential phosphorylation sites of connexin43, subsequently changing the permeability of gap junctions and resulting in heart malformation (Duncan & Fletcher, 2002). Transgenic animal models also show cardiac malformations (Reaume *et al*., 1995) and significant slowing of conduction velocity because of the absence of connexin43 (Beauchamp *et al*., 2004). There are also research attempts to reveal the relation between the *GJA1* gene and hibernation (Beauchamp *et al*., 2004; van der Heyden *et al*., 2004; Yan *et al*., 2006).

Owing to the extreme demands of echolocation and flight, bats are among the most unusual and specialized of all mammals. Flight requires three times as much energy output per unit time as does walking or running. To solve this problem, bats have the largest and most muscular heart of any mammal so that the heart can provide enough oxygen for flying (Neuweiler, 2000). The specialized hearing system of bats has an important role in receiving and processing echo information during orientation, and for many species during foraging (Siemers & Schnitzler, 2004; Jones & Teeling, 2006). The origination and evolution of echolocation in bats have been studied for long time. The traditional morphological tree and the advanced molecular phylogenetic tree, which is based on the nuclear and mitochondrial sequences, support the different hypotheses of gain or loss of echolocation of bats. The former indicated the independent origination of echolocation in microbats, but the latter suggested there is one gain of echolocation in the ancestor of all bats and one loss in the megabats (fruit bats) (Jones & Teeling, 2006). More evolution analyses about the functional gene sequences can benefit the better understanding of the natural history about the origination and evolution of the fascinating bats' ecological characters. We cloned the full-length coding region of *GJA1* gene from 12 bat species and obtained other full-length *GJA1* gene sequences of mammals from GenBank and try to test the evolution of this protein in bats' lineage. Phylogeny reconstruction and evolution rate estimation are employed to reconstruct the course of evolution for the *GJA1* gene in bats and other mammals. We also used the maximum likelihood method to identify signatures of natural selection.

## 2. Materials and methods

The cDNA of bats came from the former work about the *Foxp2* gene studies of bats (Li *et al*., 2007). Then, the coding region sequences of *GJA1* gene were amplified with the PCR conditions as 95 °C for 5 min, then 35–39 cycles of 95 °C for 35 s, 55 °C for 45 s and 72 °C for 2·5 min, with a final extension for 10 min at 72 °C. A pair of primers that were designed from conserved flanking untranslated region (UTR) conserved regions of connexin43 gene of other mammals is forward primer 5′-CGAGGTATCAGCACTTTTCTTTCATTAGG-3′ and reverse primer 5′-GGCTGTTGAGTACCACCTCCAC-3′.

### *Cloning and sequencing*

PCR products were analysed and isolated from 1% agarose gels and purified using the TaKaRa Agarose Gel DNA Purification Kit Ver.2.0 (TaKaRa, Japan), followed by ligation with the pGEM-T-easy vector (Promega, USA) and transformation into the DH5$\alpha$ competence cell (TaKaRa, Japan). The clones' identity and orientation were verified by the universal Sp6/T7 primer. The clones were cycle sequenced from both directions using BigDye sequencing kits (Applied Biosystems) on an ABI 3730A automated DNA sequencer. To avoid artefacts, multiple clones were sequenced for every specimen.

### *Method for the analyses*

We amplified and sequenced the complete coding sequences of *GJA1* gene from 12 species of bats and obtained 13 sequences of other mammals from GenBank (the species name and GenBank numbers are human, NM_000165; cattle, NM_174068; dog, NM_001002951; mouse, NM_010288; rat, NM_012567; African green monkey, AY382588; Chinese dwarf hamster, AY206456; European hedgehog, AY382589; Syrian hamster, AY206455; rabbit,

https://doi.org/10.1017/S0016672309000032 Published online by Cambridge University Press

AY382590; Russian dwarf hamster, AY382591; European ground squirrel, AY382592; pig, AY382593) for analysis. We used the software CLUSTALX 1.81 (Thompson *et al*., 1997) to align the nucleotide and amino acid sequences. MEGA3 (Kumar *et al*., 2004) was used to analyse the divergence of nucleotides and amino acids among the different mammalian lineages with the Kimura 2-parameter model to calculate pairwise comparisons of genetic distances and the Poisson correction model for amino acid data. The Nei–Gojobori method (Nei & Gojobori, 1986) was employed to calculate the synonymous and non-synonymous substitutions per synonymous and non-synonymous sites (d*S* and d*N*) among different lineages using MEGA3 and 500 replications for bootstrap tests. We also used MEGA3 to estimate the ratio of transition and transversion, as well as the saturation of the transition and transversion on three codon sites and the third codon sites, respectively. Before phylogenetic reconstruction, the program MODELTEST 3.6 (Posada & Crandall, 1998) was used to estimate the most appropriate nucleotide substitution model and parameters for the maximum likelihood method. We used MrBayes 3.1 (Huelsenbeck & Ronquist, 2001) to reconstruct the phylogenetic tree of *GJA1* gene from 25 species of mammals. In the control block of the input file for MrBayes 3.1, we set the generations equal to 1 000 000 and six Markov chains for simulation. A total of 500 000 generations were discarded before the simulation was predicted to reach to a steady condition.

After obtaining the topologies of the *GJA1* gene, which were supported in a robust statistical manner, we used the branch-specific models to test for different selection pressures among the different lineages of bats and other mammals using the program CODEML of the PAML package (Yang, 1997). Secondly, we calculated and compared the site-specific models to test the selection pressure on the different codon sites of the *GJA1* gene during the divergence and evolution of the different lineages of mammals. The site-specific models involved model M0 (One Ratio), which assumed the equal d*S*/d*N* ratio among all the amino acid sites. This model can be compared with M3 (discrete), which assumed the variable d*S*/d*N* ratio among different parts of amino acid sites, to test the possibility of positive natural selection affecting the protein evolution. Moreover, the comparison between the model M7 (beta) and model M8 (beta and $\omega$) was used to identify the sites, which were selected positively. The likelihood ratio tests (LRTs) were used to appraise the null hypothesis and alternative hypothesis. The tests were done using the $\chi^2$ distribution and the degrees of freedom were the parameter differences between two models tested.

## 3. Results

### *Sequence analyses*

To compare the *GJA1* gene sequences among lineages of mammals, we obtained the complete sequences of 12 different species of bats, involving three species of fruit bats (Family Pteropodidae: *Eonycteris spelaea*, *Cynopterus sphinx* and *Rousettus leschenaulti*), four species of horseshoe bats (Rhinolophidae: *Rhinolophus pearsoni*, *Rhinolophus rex*, *Rhinolophus macrotis* and *Rhinolophus ferrumequinum*) and each one species of Old World leaf nosed bats (Hipposideridae: *Hipposideros armiger*), a free-tailed bat (Molossidae: *Chaerephon plicata*), a false vampire (Megadermatidae: *Megaderma spasma*), a sac-winged bat (Emballonuridae: *Taphozous melanopogon*) and a vesper bat (Vespertilionidae: *Myotis ricketti*), respectively. The sequences are uploaded to GenBank with the accession numbers from EU195811 to EU195822.

After alignment of all the amino acid sequences of the *GJA1* gene from different mammals involved in this paper, the results show strong conservation of the *GJA1* gene. Except for an insertion amino acid in the cow and an amino acid absence in *T. melanopogon*, there are no variations in the length of the amino acid sequences among six orders of mammals analysed in this paper. Moreover, the summation of the amino acid sequence variations from eight orders of mammals showed that 348 out of 383 amino acid sites were conserved sites, which took more than 90·9% of whole gene. Among the sequences from bats, 366 amino acid sites were identified as conserved sites and less than 5% of amino acids were variable sites. The calculation of the pairwise distance of the amino acid distance using the Poisson correction model in software MEGA3 showed that the most amino acid difference value among bats were quite low (0·027). At the nucleotide level, 802 out of 1149 nucleotide sites were identical and 347 nucleotide sites were variable comprising 30·2% of complete sequences from all mammals discussed in this paper. A total of 204 nucleotide sites were identified as variable sites among bats. The overall average calculation of the transition/transversion based on the *GJA1* gene sequences from different lineages of mammals give the value of 2·258. The evaluation of the degree of saturation showed the linear increase of transition and transversion changes with increasing sequence divergence for complete sequences and the third codon sites data indicating the unsaturated of transition and transversion substitutions.

### *Results of the phylogenetic reconstruction*

MODELTEST selected the six parameters nucleotide substitution model (GTR) with a gamma distribution (G) and invariable sites involved (I), and the

https://doi.org/10.1017/S0016672309000032 Published online by Cambridge University Press

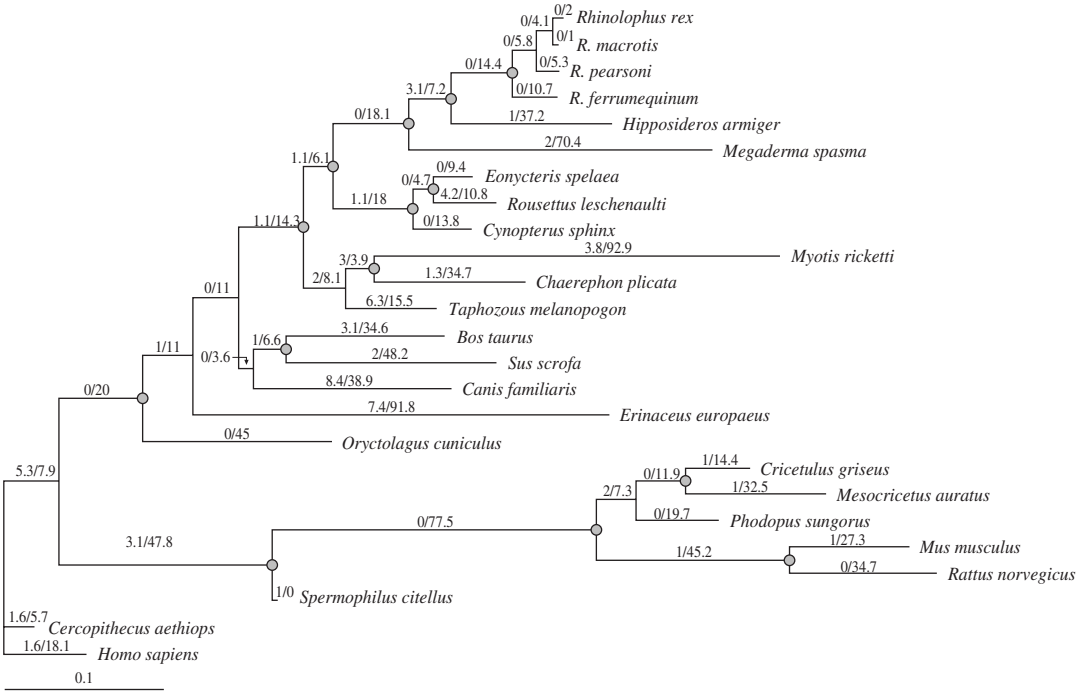


Fig. 1 . Unrooted maximum posterior probability (MPP) tree based on the complete sequences of *GJA1* gene from different mammalian lineages using software MrBayes3.1 with GTR + R + I model selected by MODELTEST 3.06. The cycles on the nodes indicate that the posterior probabilities value is higher than 0·95. The values on each branch are the estimated numbers of non-synonymous (before the slash) and synonymous (after the slash) substitutions for the relative branches (using Free Ratio Model in PAML).

estimated values of other relative parameters were base frequencies: $A = 0{\cdot}2302$, $C = 0{\cdot}2895$, $G = 0{\cdot}2609$, $T = 0{\cdot}2194$; rate matrix of substitution model: $A$–$C = 1{\cdot}9176$, $A$–$G = 4{\cdot}9434$, $A$–$T = 0{\cdot}4246$, $C$–$G = 0{\cdot}7914$, $C$–$T = 7{\cdot}2726$; proportion of invariable sites (I) = 0·5707; gamma distribution shape parameter = 1·0709. According to the topologies of the tree (Fig. 1), the evolution of the *GJA1* gene in mammals corresponds with the advanced molecular phylogenetic reconstruction of the different lineages of mammals (Murphy *et al*., 2001). Moreover, the monophyly of the *GJA1* gene in bats was highly supported. Within the lineage of Chiroptera, two main clades diverged from each other. One is composed of bats, which were assigned to the Pteropodidae, Rhinolophoidea, Hipposideridae and Megadermatidae, and the other clade included all the other bats sequenced. This topology corresponds with the results of the most advanced molecular phylogenetic reconstruction of bats with multiple-molecular markers (Teeling *et al*., 2000), which showed divergence between two major lineages of bats, the Yinpterchiroptera (Pteropodidae, Rhinolophoidea, Hipposideridae and Megadermatidae covered in this study) and the Yangochiroptera (Emaballonuridae, Molossidae and Vespertilionidae sampled here). The concordance between the gene tree of *GJA1* gene from bats and the species tree of Teeling *et al*. (2000) suggests that the divergence of the *GJA1* gene among the bats reflects the evolutionary history of this clade.

### *Results of the selection signature test of* GJA1 *gene*

The Nei–Gojobori method (Nei & Gojobori, 1986) was used to calculate the average rate of evolution in bats, rodents, artiodactyls, and in all non-bat species involved in this paper and in all species samples (Table 1). The results showed significantly higher d$S$ than d$N$ rates in each lineage and the ratio of d$N$/d$S$ was significantly less than 1 for all analyses. Except for the lower ratio of rodents, the ratio of bats was quite similar to that of other mammals at about 0·02. Furthermore, we used the maximum likelihood method to estimate the $\omega$ ratio of the *GJA1* gene among the different lineages of mammals, which was regarded as the branch-specific model. Table 2 presents the log likelihood values and the relative parameters estimated by maximum likelihood under different models. The One Ratio Model assumed the equal d$N$/d$S$ ratio of each branch to calculate the average d$N$/d$S$ ratio for each branch and site. The One Ratio Model estimated the d$N$/d$S$ ratio at 0·024 showing that the average synonymous substitution rate is much higher than the rate of non-synonymous substitution and that the *GJA1* gene has undergone elementary purifying selection in general during

https://doi.org/10.1017/S0016672309000032 Published online by Cambridge University Press

Table 1. *Calculation of average values of non-synonymous* (*d*N) *and synonymous* (*d*S) *substitutions within different mammal taxa*

| | Bats | Rodents | Artiodactyls | Non-bats | All |
|---|---|---|---|---|---|
| No. species | 12 | 6 | 2 | 13 | 25 |
| d$N \pm$SE | $0{\cdot}0061 \pm 0{\cdot}0017$ | $0{\cdot}0039 \pm 0{\cdot}0015$ | $0{\cdot}0057 \pm 0{\cdot}0024$ | $0{\cdot}0096 \pm 0{\cdot}002$ | $0{\cdot}0099 \pm 0{\cdot}002$ |
| d$S \pm$SE | $0{\cdot}3016 \pm 0{\cdot}0232$ | $0{\cdot}3587 \pm 0{\cdot}0306$ | $0{\cdot}3034 \pm 0{\cdot}0413$ | $0{\cdot}5387 \pm 0{\cdot}0377$ | $0{\cdot}5021 \pm 0{\cdot}0323$ |
| d$N$/d$S$ | 0·0202 | 0·0109 | 0·0188 | 0·0178 | 0·0197 |

Table 2. *Estimated parameters of the site-specific models*

| Model | $\ell$ | Parameters | Positively selected sites |
|---|---|---|---|
| Free-ratio | −5996·85 | $\kappa = 2{\cdot}473$ | – |
| One Ratio (M0) | −6049·30 | $\kappa = 2{\cdot}459$, $\omega = 0{\cdot}024$ | – |
| M3: discrete ($k = 3$) | −5985·38 | $\kappa = 2{\cdot}443$<br>$P_0 = 0{\cdot}904$, $P_1 = 0{\cdot}084$ ($P_2 = 0{\cdot}012$)<br>$\omega_0 = 0{\cdot}004$, $\omega_1 = 0{\cdot}180$, $\omega_2 = 0{\cdot}568$ | Null |
| M7: beta | −5987·69 | $\kappa = 2{\cdot}441$<br>$p = 0{\cdot}064$, $q = 1{\cdot}774$ | – |
| M8: beta and $\omega$ | −5986·55 | $\kappa = 2{\cdot}450$<br>$P_0 = 0{\cdot}996$ ($P_1 = 0{\cdot}004$)<br>$p = 0{\cdot}068$, $q = 2{\cdot}245$, $\omega = 1{\cdot}00$ | Null |

mammal evolution. This model gave the log maximum likelihood value as −6049·30 (Table 2). The Free Ratio Model assumed an independent $\omega$ ratio for each branch in the tree. In the tree of Fig. 1, 54 branches were assumed one $\omega$ ratio under the free ratio estimation, which leads to a log maximum likelihood value of −5996·85. In comparison with the One Ratio Model, the Free Ratio Model involved 53 extra parameters. Therefore, we used LRT to compare these two models with a $\chi^2$ distribution with 53 degrees of freedom. The result showed that the Free Ratio Model fits significantly better than the One Ratio Model ($2\Delta\ell = 104{\cdot}9$, df = 53, $P < 0{\cdot}001$), which indicated the indeed inconsistent d$N$/d$S$ ratios among different lineages of mammals. But the branch-specific model does not support the hypothesis that there is positive selection to accelerate the evolutionary rate of the *GJA1* gene in specific mammal lineages.

We also used the site-specific models tests to show that the LRT between model M3 and model M0 is significant ($2\Delta\ell = 127{\cdot}84$, df = 4, $P < 0{\cdot}001$), suggesting that discrete selection pressures act on the different sites of the *GJA1* gene. The estimation of model M3 showed that 90·4% (Table 2) of sites were under purifying selection having an extremely low evolution ratio of 0·004. Moreover, 8·4 and 1·2% of sites also had more synonymous substitutions than non-synonymous substitutions with the d$N$/d$S$ ratios as 0·18 and 0·57, respectively. The LRT between model M8 and M7 did not support the hypothesis that model M8 was better than model M7 ($2\Delta\ell = 2{\cdot}28$, df = 3, $P > 0{\cdot}05$) and none of the sites were identified being under positive selection.

## 4. Discussion

The members of gap junction gene families have been widely found in vertebrate species, as well as being well studied in human and mouse (Willecke *et al.*, 2002; Desplantez *et al.*, 2003). The phylogenetic reconstruction of the gap junction gene super families shows a monophyletic clade for the *GJA1* (*connexin43*) gene, which included human, mouse and zebra fish (Eastman *et al.*, 2006). Other researchers identified conserved characters in amphibians (van der Heyden *et al.*, 2001) and birds (Musil *et al.*, 1990), suggesting that *GJA1* diverged from other isogenous genes an extraordinarily long time ago, at least before the radiation of vertebrate species. Previous work compared differences among vertebrate lineages at nucleotide and amino acid levels and showed the highest divergences of 31·3 and 28·4% at nucleotides and amino acids, respectively (van der Heyden *et al.*, 2004). The saturation tests give the results that the transition and transversion of whole data and third codon sites data do not reach saturation, which could bring the mis-estimation in phylogenetic reconstruction. In this paper, we sequenced the *GJA1* gene from 12 bat species from seven families and obtained orthologous sequences of mammals from GenBank to

https://doi.org/10.1017/S0016672309000032 Published online by Cambridge University Press

investigate *GJA1* gene evolution. In a manner similar to other mammals, the bat sequences showed high conservation. The phylogenetic reconstruction results showed high concordance between the *GJA1* gene tree and the species tree based on the combined multiple-markers data on the higher-level classification of mammals. In the Bayesian tree of the *GJA1* gene, high statistical support (Bayesian posterior probabilities values higher than 0·95) was obtained for the monophyly of the major clades of the mammals as suggested by the advanced molecular phylogenetic work on Placentalia (Madsen *et al*., 2001; Murphy *et al*., 2001). Chiroptera, Carnivora, Artiodactyla and Insectivora were suggested as orders belonging to Laurasiatheria (Murphy *et al*., 2001), which were highly supported as a monophyletic group in our data. Moreover, the monophyly of Chiroptera was highly supported and matched the results of molecular phylogenetic and morphological research (Simmons *et al*., 1991; Simmons, 1994; Teeling *et al*., 2000). Our results support recognition of the Yinpterchiroptera and Yangochiroptera clades based on combined nuclear and mitochondrial markers (Teeling *et al*., 2000). Gene duplication and extinction can create discrepancies between gene trees and species trees (Page, 1998). The high comparability of the *GJA1* gene tree and the published species tree gives information about the evolutionary stability of this gap junction protein in mammals and the divergence patterns of the *GJA1* gene correspond closely with the species' origins and divergence patterns. According to all above, we suggested that *GJA1* gene should be a good molecular marker for phylogenetic reconstruction, especially for investigation of the high-level relationships in mammals.

The estimated evolution ratios of all the branches in the phylogenetic tree show that strong purifying selection is the predominant factor in the evolution of the *GJA1* gene in mammals. Although the ancestral sequence reconstruction gave different evolutionary rates among different lineages of mammals and significant results of different selection powers among the different locations of this protein, the average synonymous mutation rate is much higher than the non-synonymous mutation rate, meaning that natural selection tends to stabilize this protein and eliminate nucleotide mutations caused by the deleterious changes to the protein's functions.

The *GJA1* gene codes for the protein connexin43, which is a critical factor necessary to maintain cardiac electrical conduction. In the cardiac working myocardium, three connexin proteins have been identified (connexin40, connexin43 and connexin45). The expression of each gene is in different locations in the heart. With a few exceptions, such as nodal tissues and the sinuatrial node, connexin43 is the most abundant connexin in the heart. Connexin40 and connexin45 were expressed in atrial tissue, nodal tissues and in the conduction system (Gros & Jongsma, 1996). Connexin43 is the exclusive connexin protein in the intercalated disk (ID) region in adult ventricles (van Kempen *et al*., 1995). According to the disease-related research, changes in connexin43 can disrupt the normal distribution of the gap junction in the ID of patients, and the amount of connexin43 protein was reduced by almost 40 % in the heart of the people when the left ventricle is hypertrophied (Peters *et al*., 1993). All these results suggest that connexin43 has a vital function in the heart involving conduction and ventricular action potentiality. The importance of the *GJA1* gene in maintaining heart function was also supported by knockout work on the mouse. Knockout of the *GJA1* gene causes death because of malignant ventricular arrhythmias (Danik *et al*., 2004). Among the mammals, bats are faced with extreme demands imposed by flight and echolocation. Flying requires more than three times the energy cost of running (Neuweiler, 2000), bringing major challenges to the circulatory and respiratory systems. The heart is the engine of the circulatory system, and shows extreme adaptation in bats compared with other mammals. Given similarities in size and body weight, the heart proportion of bats is two to three times bigger than that of a mouse (Wachtlova *et al*., 1970). Moreover, bats possess an extreme range of heart rates. The average heart beat frequencies of bats can be 500 beats per minute, rising to more than 1000 beats per minute in flight, as well as the lowest 4 beats per minute in hibernation (Neuweiler, 2000). This variation is probably the largest among the mammals and brings large stress on cardiac electrical conduction.

On the other hand, research shows that the *GJA1* gene impacts upon the growth and maturation of follicles in the ovary (Teilmann, 2005). Moreover, communication changes induced by expression changes of *GJA1* mRNA and connexin43 protein levels influence the meiotic maturation and the change of atretic follicles (Nuttinck *et al*., 2000). Bats are among the few mammals that show delayed ovulation and fertilization (Oxberry, 1979, and the unpublished data of Zhe Wang *et al*.). Although we found the amino acid Arg109 as the bats-specific amino acid substitution, which is caused by a nucleotide mutation (A to G) at the second codon position in the ancestral bats, more experimental work is needed to test the relationship between the substitution of this amino acid and bats-specific ecological characters. Like other connexins, connexin43 consists of the following topological domains: an amino terminus, two extracellular loops, four membrane-spanning domains, one cytoplasmic loop and a carboxy terminus. The parts of cytoplasmic loop and carboxy terminus are highly variable among the different connexins and were

thought to be important for regulation (Kumar & Gilula, 1996). Except for this, every domain has special amino acid sites and specific functions (Krutovskikh & Yamasaki, 2000). The 109 site is located on the inner loop between the second and third transform membrane domains, an area suggested as having an important influence on pH and voltage gating sensitivity in cells (Wang *et al*., 1996). Previous work compared the sequences of the *GJA1* gene from some hibernating and non-hibernating mammals, and tried to investigate possible contribution of this gene to the cardiovascular physiological adaptation to hibernation (van der Heyden *et al*., 2004). The research of van der Heyden *et al*. (2004) identified six amino acid sites (A116, T118, S244, H248, L254 and A349), where mammals differed from other non-mammalian vertebrates. However, in our alignment of amino acid sequences, we find that *M. spasma* did not fit with this pattern. *M. spasma* has a substitution at M254 that is shared with the chicken. This site is in the carboxy terminus of connexin43. Although there are some phosphorylation sites (Swenson *et al*., 1990; Kanemitsu & Lau, 1993; Saez *et al*., 1997; Lampe *et al*., 2000) in this domain, the potential effect of this amino acid substitution is still not clear.

In the alignment of the amino acid sequences presented here, we also found several bats having distinct amino acids. Most of these amino acids are located in carboxy terminus, which is the most variable region among the different connexins and was also thought to be important for regulation (Kumar & Gilula, 1996). Although carboxy terminus of connexin43 is longer than other connexins, only two functional roles have been suggested for this domain so far, one of which is that it contains the phosphorylation sites for different kinases (Krutovskikh & Yamasaki, 2000). The other suggested function is that the last six residues of connexin43 directly interact with the ZO-1 protein of tight junctions (Giepmans & Moolenaar, 1998; Toyofuku *et al*., 1998). Moreover, the rather short size of the carboxy terminus of connexin26 suggests that it might not be essential for normal function of connexin (Krutovskikh & Yamasaki, 2000). Several mutations in the carboxy terminus sites of connexin43 can cause heart disease, such as malformation and defects of laterality in patients (Britz-Cunningham *et al*., 1995, but see Casey & Ballabio, 1995; Penman Splitt *et al*., 1997).

In conclusion, we firstly studied the *GJA1* gene of bats. Using the molecular experiment method, we cloned and sequenced the *GJA1* gene from 12 bats. We found the novel amino acid substitutions on the specific branches of bats which have the potentially affection of function. Using phylogenetic methods, the evolutionary analyses showed the high power of purifying selection in mammals, as well as in bats. The gene tree of *GJA1* gene from mammals gives the high accordant topologies with good statistic support to that of the species tree based on the multiple molecular markers. This suggested the high evolutionary stability of *GJA1* gene, which should be a good molecular marker to study the phylogenetic relationships among the mammals.

This work was funded by grants awarded to S. Zhang under the Key Construction Program of the National "985" Project and "211" Project. We thank Z. Wang for her data on procreation of bats.

## References

Alexander, D. B. & Goldberg, G. S. (2003). Transfer of biologically important molecules between cells through gap junction channels. *Current Medicinal Chemistry* **10**, 2045–2058.

Beauchamp, P., Choby, C., Desplantez, T., de Peyer, K., Green, K., Yamada, K. A., Weingart, R., Saffitz, J. E. & Kleber, A. G. (2004). Electrical propagation in synthetic ventricular myocyte strands from germline connexin43 knockout mice. *Circulation Research* **95**, 170–178.

Becker, D. L., Cook, J. E., Davies, C. S., Evans, W. H. & Gourdie, R. G. (1998). Expression of major gap junction connexin types in the working myocardium of eight chordates. *Cell Biology International* **22**, 527–543.

Beyer, E. C., Kistler, J., Paul, D. L. & Goodenough, D. A. (1989). Antisera directed against connexin43 peptides react with a 43-kD protein localized to gap junctions in myocardium and other tissues. *The Journal of Cell Biology* **108**, 595–605.

Britz-Cunningham, S. H., Shah, M. M., Zuppan, C. W. & Fletcher, W. H. (1995). Mutations of the Connexin43 gap-junction gene in patients with heart malformations and defects of laterality. *The New England Journal of Medicine* **332**, 1323–1329.

Casey, B. & Ballabio, A. (1995). Connexin43 mutations in sporadic and familial defects of laterality. *The New England Journal of Medicine* **333**, 941; author reply 941–942.

Corcos, I. A., Meese, E. U. & Loch-Caruso, R. (1993). Human connexin43 gene locus, *GJA1*, sublocalized to band 6q21–>q23.2. *Cytogenetic Cell Genetics* **64**, 31–32.

Danik, S. B., Liu, F., Zhang, J., Suk, H. J., Morley, G. E., Fishman, G. I. & Gutstein, D. E. (2004). Modulation of cardiac gap junction expression and arrhythmic susceptibility. *Circulation Research* **95**, 1035–1041.

Desplantez, T., Marics, I., Jarry-Guichard, T., Veteikis, R., Briand, J. P., Weingart, R. & Gros, D. (2003). Characterization of zebrafish Cx43.4 connexin and its channels. *Experimental Physiology* **88**, 681–690.

Duncan, J. C. & Fletcher, W. H. (2002). alpha 1 Connexin (connexin43) gap junctions and activities of cAMP-dependent protein kinase and protein kinase C in developing mouse heart. *Developmental Dynamics* **223**, 96–107.

Eastman, S. D., Chen, T. H., Falk, M. M., Mendelson, T. C. & Iovine, M. K. (2006). Phylogenetic analysis of three complete gap junction gene families reveals lineage-specific duplications and highly supported gene classes. *Genomics* **87**, 265–274.

Fishman, G. I., Eddy, R. L., Shows, T. B., Rosenthal, L. & Leinwand, L. A. (1991). The human connexin gene family of gap junction proteins: distinct chromosomal locations but similar structures. *Genomics* **10**, 250–256.

Fishman, G. I., Spray, D. C. & Leinwand, L. A. (1990). Molecular characterization and functional expression of the human cardiac gap junction channel. *Journal of Cell Biology* **111**, 589–598.

Giepmans, B. N. & Moolenaar, W. H. (1998). The gap junction protein connexin43 interacts with the second PDZ domain of the zona occludens-1 protein. *Current Biology* **8**, 931–934.

Goodenough, D. A., Goliger, J. A. & Paul, D. L. (1996). Connexins, connexons, and intercellular communication. *Annual Review of Biochemistry* **65**, 475–502.

Gros, D. B. & Jongsma, H. J. (1996). Connexins in mammalian heart function. *Bioessays* **18**, 719–730.

Huelsenbeck, J. P. & Ronquist, F. (2001). MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755.

Jones, G. & Teeling, E. C. (2006). The evolution of echolocation in bats. *Trends in Ecology and Evolution* **21**, 149–156.

Kanemitsu, M. Y. & Lau, A. F. (1993). Epidermal growth factor stimulates the disruption of gap junctional communication and connexin43 phosphorylation independent of 12–0-tetradecanoylphorbol 13-acetate-sensitive protein kinase C: the possible involvement of mitogen-activated protein kinase. *Molecular Biology of Cell* **4**, 837–848.

Kikuchi, T., Kimura, R. S., Paul, D. L. & Adams, J. C. (1995). Gap junctions in the rat cochlea: immunohistochemical and ultrastructural analysis. *Anatomy and Embryology* (*Berlin*) **191**, 101–118.

Krutovskikh, V. & Yamasaki, H. (2000). Connexin gene mutations in human genetic diseases. *Mutation Research* **462**, 197–207.

Kumar, N. M. & Gilula, N. B. (1996). The gap junction communication channel. *Cell* **84**, 381–388.

Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* **5**, 150–163.

Lampe, P. D., TenBroek, E. M., Burt, J. M., Kurata, W. E., Johnson, R. G. & Lau, A. F. (2000). Phosphorylation of connexin43 on serine368 by protein kinase C regulates gap junctional communication. *The Journal of Cell Biology* **149**, 1503–1512.

Li, G., Wang, J., Rossiter, S. J., Jones, G. & Zhang, S. (2007). Accelerated FoxP2 evolution in echolocating bats. *PLoS ONE* **2**, e900.

Liu, X. Z., Xia, X. J., Adams, J., Chen, Z. Y., Welch, K. O., Tekin, M., Ouyang, X. M., Kristiansen, A., Pandya, A., Balkany, T., Arnos, K. S. & Nance, W. E. (2001). Mutations in GJA1 (connexin 43) are associated with non-syndromic autosomal recessive deafness. *Human Molecular Genetics* **10**, 2945–2951.

Madsen, O., Scally, M., Douady, C. J., Kao, D. J., DeBry, R. W., Adkins, R., Amrine, H. M., Stanhope, M. J., de Jong, W. W. & Springer, M. S. (2001). Parallel adaptive radiations in two major clades of placental mammals. *Nature* **409**, 610–614.

Murphy, W. J., Eizirik, E., O'Brien, S. J., Madsen, O., Scally, M., Douady, C. J., Teeling, E., Ryder, O. A., Stanhope, M. J., de Jong, W. W. & Springer, M. S. (2001). Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* **294**, 2348–2351.

Musil, L. S., Beyer, E. C. & Goodenough, D. A. (1990). Expression of the gap junction protein connexin43 in embryonic chick lens: molecular cloning, ultrastructural localization, and post-translational phosphorylation. The *Journal of Membrane Biology* **116**, 163–175.

Nei, M. & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* **3**, 418–426.

Neuweiler, G. (2000). The Biology of Bats. Oxford: Oxford University Press.

Nuttinck, F., Peynot, N., Humblot, P., Massip, A., Dessy, F. & Flechon, J. E. (2000). Comparative immunohistochemical distribution of connexin 37 and connexin 43 throughout folliculogenesis in the bovine ovary. *Molecular Reproduction and Development* **57**, 60–66.

Oxberry, B. A. (1979). Female reproductive patterns in hibernating bats. *Journal of Reproduction and Fertility* **56**, 359–367.

Page, R. D. (1998). GeneTree: comparing gene and species phylogenies using reconciled trees. *Bioinformatics* **14**, 819–820.

Penman Splitt, M., Tsai, M. Y., Burn, J. & Goodship, J. A. (1997). Absence of mutations in the regulatory domain of the gap junction protein connexin 43 in patients with visceroatrial heterotaxy. *Heart* **77**, 369–370.

Peters, N. S., Green, C. R., Poole-Wilson, P. A. & Severs, N. J. (1993). Reduced content of connexin43 gap junctions in ventricular myocardium from hypertrophied and ischemic human hearts. *Circulation* **88**, 864–875.

Posada, D. & Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.

Reaume, A. G., de Sousa, P. A., Kulkarni, S., Langille, B. L., Zhu, D., Davies, T. C., Juneja, S. C., Kidder, G. M. & Rossant, J. (1995). Cardiac malformation in neonatal mice lacking connexin43. *Science* **267**, 1831–1834.

Saez, J. C., Nairn, A. C., Czernik, A. J., Fishman, G. I., Spray, D. C. & Hertzberg, E. L. (1997). Phosphorylation of connexin43 and the regulation of neonatal rat cardiac myocyte gap junctions. *Journal of Molecular and Cellular Cardiology* **29**, 2131–2145.

Siemers, B. M. & Schnitzler, H. U. (2004). Echolocation signals reflect niche differentiation in five sympatric congeneric bat species. *Nature* **429**, 657–661.

Simmons, N. B. (1994). The case for chiropteran monophyly. *American Museum Novitates* **3103**, 1–54.

Simmons, N. B., Novacek, M. J. & Baker, R. J. (1991). Approaches, methods, and the future of the chiropteran monophyly controversy: a reply to J. D. Pettigrew. *Systems Zoology* **40**, 239–243.

Swenson, K. I., Piwnica-Worms, H., McNamee, H. & Paul, D. L. (1990). Tyrosine phosphorylation of the gap junction protein connexin43 is required for the pp60v-src-induced inhibition of communication. *Cellular Regulation* **1**, 989–1002.

Teeling, E. C., Scally, M., Kao, D. J., Romagnoli, M. L., Springer, M. S. & Stanhope, M. J. (2000). Molecular evidence regarding the origin of echolocation and flight in bats. *Nature* **403**, 188–192.

Teilmann, S. C. (2005). Differential expression and localisation of connexin-37 and connexin-43 in follicles of different stages in the 4-week-old mouse ovary. *Molecular and Cellular Endocrinology* **234**, 27–35.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**, 4876–4882.

Toyofuku, T., Yabuki, M., Otsu, K., Kuzuya, T., Hori, M. & Tada, M. (1998). Direct association of the gap junction protein connexin-43 with ZO-1 in cardiac

myocytes. *The Journal of Biological Chemistry* **273**, 12725–12731.

Uyguner, O., Emiroglu, M., Uzumcu, A., Hafiz, G., Ghanbari, A., Baserer, N., Yuksel-Apak, M. & Wollnik, B. (2003). Frequencies of gap- and tight-junction mutations in Turkish families with autosomal-recessive non-syndromic hearing loss. *Clinical Genetics* **64**, 65–69.

van der Heyden, M. A., Roeleveld, L., Peterson, J. & Destree, O. H. (2001). Connexin43 expression during Xenopus development. *Mechanisms of Development* **108**, 217–220.

van der Heyden, M. A., van Eijk, M., Wilders, R., de Bakker, J. M. & Opthof, T. (2004). Connexin43 orthologues in vertebrates: phylogeny from fish to man. *Development Genes and Evolution* **214**, 261–266.

van Kempen, M. J., ten Velde, I., Wessels, A., Oosthoek, P. W., Gros, D., Jongsma, H. J., Moorman, A. F. & Lamers, W. H. (1995). Differential connexin distribution accommodates cardiac function in different species. *Microscopy Research and Technique* **31**, 420–436.

Wachtlova, M., Poupa, O. & Rakusan, K. (1970). Quantitative differences in the terminal vascular bed of the myocardium in the brown bat (*Myotis myotis*) and the laboratory mouse (*Mus musculus*). *Physiologia Bohemoslovaca* **19**, 491–495.

Wang, X., Li, L., Peracchia, L. L. & Peracchia, C. (1996). Chimeric evidence for a role of the connexin cytoplasmic loop in gap junction channel gating. *Pflugers Archive* **431**, 844–852.

Willecke, K., Eiberger, J., Degen, J., Eckardt, D., Romualdi, A., Guldenagel, M., Deutsch, U. & Sohl, G. (2002). Structural and functional diversity of connexin genes in the mouse and human genome. *Biological Chemistry* **383**, 725–737.

Willingham-Rocky, L. A., Golding, M. C., Wright, J. M., Kraemer, D. C., Westhusin, M. E. & Burghardt, R. C. (2007). Cloning of GJA1 (connexin43) and its expression in canine ovarian follicles throughout the estrous cycle. *Gene Expression Patterns* **7**, 66–71.

Yan, J., Burman, A., Nichols, C., Alila, L., Showe, L. C., Showe, M. K., Boyer, B. B., Barnes, B. M. & Marr, T. G. (2006). Detection of differential gene expression in brown adipose tissue of hibernating arctic ground squirrels with mouse microarrays. *Physiological Genomics* **25**, 346–353.

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Computational and Applied Bioscience* **13**, 555–556.