Identification of a Novel Gig2 Gene Family Specific to Non-Amniote Vertebrates

Yi-Bing Zhang*, Ting-Kai Liu, Jun Jiang, Jun Shi, Ying Liu, Shun Li, Jian-Fang Gui*

State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China

Abstract

Gig2 (grass carp reovirus (GCRV)-induced gene 2) is first identified as a novel fish interferon (IFN)-stimulated gene (ISG). Overexpression of a zebrafish Gig2 gene can protect cultured fish cells from virus infection. In the present study, we identify a novel gene family that is comprised of genes homologous to the previously characterized Gig2. EST/GSS search and in silico cloning identify 190 Gig2 homologous genes in 51 vertebrate species ranged from lampreys to amphibians. Further large-scale search of vertebrate and invertebrate genome databases indicate that Gig2 gene family is specific to non-amniotes including lampreys, sharks/rays, ray-finned fishes and amphibians. Phylogenetic analysis and synteny analysis reveal lineage-specific expansion of Gig2 gene family and also provide valuable evidence for the fish-specific genome duplication (FSGD) hypothesis. Although Gig2 family proteins exhibit no significant sequence similarity to any known proteins, a typical Gig2 protein appears to consist of two conserved parts: an N-terminus that bears very low homology to the catalytic domains of poly(ADP-ribose) polymerases (PARPs), and a novel C-terminal domain that is unique to this gene family. Expression profiling of zebrafish Gig2 gene family shows that some duplicate pairs have diverged in function via acquisition of novel spatial and/or temporal expression under stresses. The specificity of this gene family to non-amniotes might contribute to a large extent to distinct physiology in non-amniote vertebrates.

Introduction

A gene family is composed of homologous genes that are formed by duplication of a single original gene. Duplication events can be classified into two types according to the mechanisms of their origins: whole genome duplication (WGD) and segmental duplication (SD) [1], both of which provide primary resources for the origin of gene families. Two rounds of WGDs are hypothesized to have occurred on the jawed vertebrate stem, after the divergence of urochordates but before the split between cartilaginous fish and bony vertebrates [2]; teleost fish is suggested to have specifically undergone a third round of WGD termed as fish-specific genome duplication (FSGD) [3]. SD is believed to contribute a lot to lineage-specific expansion of gene families [4], while WGD is important for shaping diversity and complexity of modern vertebrates, because reciprocal gene loss and re-diploidization after WGDs might contribute to the radiation of vertebrates [5]. The members of a gene family are likely to have highly similar functions, such as vertebrate interferon (IFN) family genes, all of which share a common ability to block cellular replication of different viruses [6]. In mammals, viral infection results in activation of an innate immune response for establishment of a host antiviral state, which is characterized by the production of IFN family proteins and the subsequent transcriptional upregulation of IFN-stimulated genes (ISGs) [7]. Dozens of gene families are involved in IFN antiviral immune response. For example, Toll-like receptor (TLR) family members (such as TLR3/4/7/8/9) and retinoic acid-inducible gene I (RIG-I)-like receptor (RLR) family members recognize different pathogen-associated molecular patterns (PAMPs) to trigger type I IFN signalling [6,7]; IFN-regulatory factor (IRF) 3 and IRF7, two IRF family members, directly control expression of IFN family genes [7,8,9]; IFIT (IFN-induced proteins with Tetratricopeptide Repeat (TPR) motifs) family members mediate IFN antiviral effects by selectively restricting the replication of virus lacking 2’-O methylation mRNA or with a 5’-triphosphate RNA [10,11]. Recently a new strategy has been proposed, in which cells reduce sensitivity to viruses through modification of the viral proteins by poly(ADP-ribose)ylation (PARylation), a post-translational modification of proteins that is catalyzed by poly(ADP-ribose) polymerases (PARPs) [12].

Similar to mammals, fish possess conserved IFN antiviral response by the TLR pathway and the RLR pathway [6,13]. Consistently, fish contain almost all subsets of TLRs corresponding to human TLRs [14]; the same is true for RLR family members [15]. Fish IRF family members show a clear orthologous relationship with mammalian counterparts [16]. In addition, fish IFNs exert antiviral function by induction of an array of ISGs that are conserved in fish [8,17,18]. However, an attempt to screen fish IFN responsive genes has identified some novel ISGs including Gig2 (grass carp reovirus (GCRV)-induced gene 2) [19,20]. Interestingly, Gig2 is not significantly homologous to any known genes and seems not to be found in mammalian genomes [19], which has sparked great interests to investigate its evolutionary origin.
Although there is low or no constitutive expression of crucian carp *Carassius auratus* Gig2 in cultured fish cells [19], it can be detected in tissues of healthy grass carp (*Ctenopharyngodon idella*) at mRNA and protein levels [21]. Despite these differences, Gig2 mRNA and protein are highly up-regulated by viral infection, IFN and poly(I:C) treatment [19,20,21]. Actually, Gig2 is a typical ISG [21] and is activated by fish IFN signalling [8,22,23]. Overexpression of a zebrafish Gig2 member (*DreGig2*) prevents spring virema of carp virus (SVCV) and *Rana grylio* virus (RGV) infection in EPC (Epithelioma papulosum cyprinid) cells [23]. This result confirms a potential antiviral role of fish Gig2, although the detailed mechanisms remain elusive. Strikingly, at least two proteins homologous to crucian carp Gig2 are found in IFN-produced fish cells [21], giving a clue that there may be a gene family homologous to Gig2 gene.

In the present study, we reported identification of a novel gene family termed Gig2 gene family. Strikingly, Gig2 gene family is specific to non-amniotes including lampreys, sharks/rays, ray-finned fishes and amphibians, by large-scale search of EST, GSS and genome databases. Phylogenetic analyses revealed an extensive expansion of Gig2 gene family by lineage-specific and species-specific duplication. Expression characterization of zebrafish Gig2 family genes suggested that some duplicate pairs had diverged in function. In spite of no significant homology to known proteins, a typical Gig2 protein is composed of an N-terminus that exhibits very low homology to PARP catalytic domain and a C-terminus that is conserved in all Gig2 proteins but no homology to any a known domain.

**Results**

**Identification of Gig2 Genes in Crucian Carp**

Sequencing of 2035 ESTs from a subtractive cDNA library, which was made by UV-inactivated GCRV-infected *Carassius auratus* blastulae embryonic cells (CAB) against mock-infected CAB cells [24], retrieved 45 ESTs, which were homologous to a previously identified gene, Gig2 (termed *CauGig2A* in this paper) [19]. These ESTs were assembled to 4 contigs, according to which 4 full-length CDAs besides *CauGig2A* were obtained by RACE-PCR. These 4 genes were subsequently termed *CauGig2A, CauGig2D, CauGig2B* and *CauGig2O* with reference to the locus of zebrafish *Danio rerio* chromosome that was described thereafter. Multiple alignments reveal a high level of amino acid sequence identity between them, from 40% (*CauGig2A* versus *CauGig2O*) to 87% (*CauGig2A* versus *CauGig2B*) (Fig. 1A). Strikingly, *CauGig2A* contains 276 amino acids, almost twice in length than the other four proteins. Actually, *CauGig2A* is composed of two homologues units, since the 147-amino-acid-N terminus and the 129-amino-acid-C terminus of *CauGig2A* display a high level of sequence identity either to each other or to the other four proteins (Fig. 1A).

Similar to *CauGig2A* that was upregulated by GCRV infection and fish IFN [21], the other four genes were transcriptionally induced in CAB cells treated with UV-inactivated GCRV, exhibiting a similar expression pattern where the transcription expression gradually increased with stimulated time and reached a peak and then decreased, although the peak of *CauGig2O* was later than others (Fig. 1B). Additionally, these genes showed a ubiquitous and constitutive expression in tissues of healthy grass carps, with a relatively high level of transcription in thymus, spleen, head kidney, trunk kidney, gill, hindgut and brain, but a low level of transcription in liver, skin and heart (Fig. 1C), which was similar to *CauGig2A* [21].

**Identification of Gig2 Gene Clusters in Zebrafish**

Simultaneous responsiveness of multiple crucian carp Gig2 homologous genes to viral infection sparked our interests to investigate whether there was a Gig2 gene family in fish genomes. Initially, we searched zebrafish genome using crucian carp Gig2 protein sequences as queries by tBLASTn. 20 orthologous genes were found in zebrafish genome of the version 7 (Zv7), termed *DreGig2A* to *DreGig2T*. However, only 13 homologous genes were retained in zebrafish genome of Zv9, all of which were distributed in three gene loci: one locating on chromosome 21 with 10 members (*DreGig2D* to *DreGig2N* but lacking *DreGig2I*), one on chromosome 5 with 2 members (*DreGig2O* and *DreGig2P*), and one on chromosome 2 with only one member (*DreGig2Q*). *DreGig2M* was assumed to be a pseudogene according to genome annotation. Phylogenetic analysis revealed an orthologous relationship between zebrafish genes and crucian carp counterparts, indicating that Gig2 gene family in zebrafish and crucian carp had extensively expanded at least before the split of two fish species (Fig. 2).

Subsequently, the expression profiles of zebrafish Gig2 genes were investigated. Most genes were significantly upregulated in ZFL cells when transfected with poly(I:C), an effective IFN inducer [8], but four (*DreGig2N, O, P, Q*) not (Fig. 3A). *DreGig2M* had two different isoforms with the same ORF which might be caused by alternative splicing. Consistently, two PCR bands were detected by specific primers (Fig. 3A). In addition, zebrafish Gig2 genes were expressed ubiquitously in all tissues tested, although at differential expression levels (Fig. 3B). Notably, we failed to detect the expression of *DreGig2F* by RT-PCR. Although both *DreGig2A* and *DreGig2I* were not found in zebrafish genome of Zv9, their transcripts were detected in *in vitro* and *in vivo* (Fig. 3). These data indicated that there existed a Gig2 gene family in zebrafish, the members of which seemed to have diverged in biological function.

**Specificity of Gig2 Gene Family to Non-amniotes**

To investigate the distribution of Gig2 gene family in different species, we searched the genome sequences in the databases of esembl and NCBI. 23 sequences orthologous to Gig2 were found from stickleback *Gasterosteus aculeatus*, medaka *Oryzias latipes*, green spotted pufferfish *Tetraodon flavilus*, takifugu *Takifugu rubripes* and frog *Xenopus laevis* in addition to zebrafish (Table 1). Then we searched non-redundant nucleotide databases of NCBI, and 9 Gig2 cDNAs were specifically hit besides crucian carp and zebrafish (Table S1). No additional protein sequences other than the sequences encoded by the identified cDNAs were found. To further investigate the species distribution of Gig2 gene family, we screened EST and GSS databases of NCBI by BLASTN using all identified Gig2 cDNA sequences as probes. 870 ESTs and 22 GSS were hit in 51 vertebrate species ranged from lampreys to amphibia (Table 1, Table S2). Unexpectedly, 3 ESTs in two algae and 5 GSSs in marine metagenomes were found encoding two Gig2-like genes in algae and 3 genes in marine water samples (Table S2), but we think that they might be caused by horizontal gene transfer [25]. We then assembled the Gig2 ESTs and GSSs using the CAP3 online service, and 140 cDNAs with putative full ORFs were in silico cloned and identified (Table S1). Together with the cDNAs and sequences from genomes, a total of 190 Gig2 genes were identified and distributed in 33 species including 2 algae (Table 1). Consistent with a taxonomy common tree constructed with the 53 species and other 10 representative species (Fig. S1), the identified Gig2 genes were found to be specific to non-amniote vertebrates, widely in ray-finned fishes (Table 1, Fig. S1). To exclude the possibility of eluding from the lower or upper species, we searched all the available databases of *Caenorhabditis elegans* (roundworm), *Drosophila melanogaster* (fruit fly), *Strongylocentrotus*
purpuratus (sea urchin), Branchiostoma lanceolatum (lancelet), Anolis carolinensis (lizard), Gallus gallus (chicken), Mus musculus (mouse) and Homo sapiens (human), but no Gig2 homologous sequences were found. Therefore, Gig2 gene family is specific to non-amniotes including lampreys, sharks/rays, ray-finned fish and amphibians regardless of the two algae.

**Domain Analyses of the Putative Gig2 Proteins**

To determine whether there existed any known protein domains, we searched NCBI’s conserved domain databases using each putative Gig2 protein sequence. 165 out of 190 Gig2 protein sequences hit a domain named Pssmid 30070 (TCCD_inducible_PARP_like) with a very low bitscore (generally less than 50,
Further searching Pfam-A (http://pfam.sanger.ac.uk/) found that Pssmid 30070 (TCCD_inducible_PARP_like) is a PARP catalytic domain with about 230 residues in length, the active site of which is composed of a block of 50 amino acids, known as the ‘PARP signature’ [26,27]. Multiple alignments showed that a partial N-terminal stretch of all Gig2 proteins is weakly homologous to the region equivalent to the PARP signature (Table S1), even in the Gig2 members that were not hit to the Pssmid 30070 by NCBI batch CD search (data not shown). In addition, there is high sequence homology between the C-terminus of Gig2 proteins although no known domains were predicted. Accordingly, a typical Gig2 protein might contain two conserved parts, one N-terminus with a stretch weakly homologous to the PARP signature, and one C-terminus with no significant homology to any a known domain (Fig. 4A).

Interestingly, some Gig2 proteins appear to have two stretches that weakly hit Pssmid 30070 (PARP-like domain), including CauGig2A, ManGig2I, ManGig2J, OmyGig2I, SauGig2E, SsaGig2M, SsaGig2N and SsaGig2O (Table S1). Similar to CauGig2A in Fig. 1A, they are composed of 2 tandem basic units with a high level of homology (data not shown), each of which has one conserved N-terminus and one conserved C-terminus with reference to the typical Gig2 protein (containing 1 basic unit) (Fig. 4B). The former part was named by appending "_i" to the
Table 1. Summary of Gig2 genes identified in this study.

| category          | organism                                  | Sequence sources          | No.  |
|-------------------|-------------------------------------------|---------------------------|------|
| ray-finned fish   | sablefish *Anoplopoma fimbria*            | 28 EST                    | 2    |
|                   | white sturgeon *Acipenser transmontanus*  | 8 EST/GSS                 | 1    |
|                   | Crucian carp *Carassius auratus*          | 5 complete mRNA           | 5    |
|                   | common carp *Cyprinus carpio*             | 10 EST                    | 3    |
|                   | lake whitefish *Coregonus clupeaformis*   | 4 EST                     | 1    |
|                   | European seabass *Dicentrarchus labrax*   | 8 EST/GSS                 | 6    |
|                   | Antarctic toothfish *Dissostichus mawsoni*| 2 GSS                     | 1    |
|                   | zebrafish *Danio rerio*                   | 13 EST, Genome (Chr.1,5,21)| 13   |
|                   | northern pike *Esox lucius*               | 15 EST                    | 3    |
|                   | mummichog *Fundulus heteroclitus*         | 4 EST                     | 2    |
|                   | stickleback *Gasterosteus aculeatus*      | 2 EST; Genome (Chr. XI, XIII)| 4   |
|                   | Atlantic cod *Gadus morhua*               | 20 EST                    | 6    |
|                   | Atlantic halibut *Hippoglossus hippoglossus*| 2 EST                   | 1    |
|                   | blue catfish *Ictalurus furcatus*         | 10 EST                    | 5    |
|                   | channel catfish *Ictalurus punctatus*     | 99 EST/GSS                | 10   |
|                   | Japanese seabass *Labeolabrax japonicus*  | 1 EST                     | 1    |
|                   | Barramundi perch *Lates calcarifer*       | 1 EST                     | 1    |
|                   | *Lipochromis* sp. ‘matumbi hunter’        | 5 EST                     | 2    |
|                   | Oriental weatherfish *Misgurnus anguillicaudatus* | 177 EST             | 10   |
|                   | Brown croaker *Mischthys miicy*           | 2 EST                     | 1    |
|                   | Japanese medaka *Oryzias latipes*         | 6 GSS; Genome (Chr.8, Chr.9)| 4   |
|                   | Rainbow smelt *Osmerus mordax*            | 1 EST                     | 1    |
|                   | Rainbow trout *Oncorhyncus mykiss*        | 35 EST                    | 9    |
|                   | Nile tilapia *Oreochromis niloticus*      | 13 EST/GSS, Genome        | 4    |
|                   | turbot *Psettia maxima*                   | 2 EST                     | 1    |
|                   | Japanese flounder *Paralichthys olivaceus*| 1 EST                   | 1    |
|                   | Fathead minnow *Pimephales promelas*      | 2 EST                     | 1    |
|                   | guppy *Poecilia reticulata*               | 1 EST                     | 1    |
|                   | *Pylochromis* sp. ‘redtail sheller’       | 3 EST                     | 2    |
|                   | Roach minnow *Rutilus rutilus*            | 1 EST                     | 1    |
|                   | gilthead seabream *Sparus aurata*         | 19 EST                    | 5    |
|                   | copper rockfish *Sebastes caurinus*       | 9 EST                     | 4    |
|                   | Chinese perch *Sipinperca chaussic*       | 1 complete mRNA           | 1    |
|                   | brook trout *Salvelinus fontinalis*        | 8 EST                     | 3    |
|                   | grass rockfish *Sebastes rastrelliger*    | 2 EST                     | 1    |
|                   | Atlantic salmon *Salmo salar*             | 167 EST                   | 16   |
|                   | green puffish fish *Tetraodon fluviatilis*| 1 EST                   | 1    |
|                   | green spotted puffishfish *Tetraodon nigriviridis* | 9 EST, Genome (Chr.3, 12) | 10   |
|                   | fugu *Takifugu rubripes*                  | 3 EST, Genome (Sf705, 101)| 4    |
|                   | Graying *Thymallus thymallus*             | 3 EST                     | 1    |
|                   | grass goby *Zosterisessor aphiopcephalus*  | 1 EST                     | 1    |
| Sharks/rays       | little skate *Leucoraia erinacea*         | 1 EST                     | 1    |
|                   | spiny dogfish *Squalus acanthius*         | 2 EST                     | 1    |
| Frogs/toads       | African clawed frog *Xenopus laevis*       | 16 EST/GSS                | 8    |
|                   | *Silurana Xenopus* (Silurana) tropicalis* | Genome                   | 5    |
| salamanders       | Axolotl *Ambystoma mexicanum*             | 12 EST                    | 5    |
|                   | Eastern tiger salamander *Ambystoma tigrinum tigrinum* | 18 EST               | 5    |
|                   | Japanese firebelly newt *Gynops pyrhogaster* | 8 EST                   | 3    |
|                   | Eastern newt *Notophthalmus viridescens*  | 12 EST                    | 2    |
| lampreys          | *Lethenteron japonicum*                   | 1 EST                     | 1    |
Table 1. Cont.

| category            | organism           | Sequence sources         | No. |
|---------------------|--------------------|--------------------------|-----|
| sea lamprey         | Petromyzon marinus | 7 EST                    | 2   |
| dinoflagellates     | Karenia brevis     | 2 EST                    | 1   |
| haptophytes         | Emiliania huxleyi  | 1 EST                    | 1   |
| marine genome       | Marine metagenome water samples | 5 EST/GSS               | 5   |

Figure 4. Domain analyses of Gig2 family proteins. (A) Schematic diagram of the structure of a typical Gig2 protein (Basic unit), with an N-terminus containing a stretch weakly homologous to PARK-like domain and a conserved C-terminus. (B) Schematic diagram of Gig2 proteins with two tandem units or four tandem units. (C) Phylogenetic tree of separate units of Gig2 proteins with two or four tandem units by the software Geneious using the Neighbour-Joining method with a bootstrap test of 1000 replicates. Sequences containing 2 basic units or 4 basic units were cut into 2 parts or 4 parts, which was suffixed by "_i" and "_ii" or by "_i", "_ii", "_iii" and "_iv", respectively.

doi:10.1371/journal.pone.0060588.g004
gene name to differentiate the latter part appended by "_ii", and both parts exhibited highly homology (Fig. 4B). 2 tandem units were also seen in DreGig2E, CelGig2, OMyGig2S, StoGig2C, ThGig2, in which the N-terminus in the latter unit was weakly hit on Psmid 30070, and EtuGig2C with no hits on Psmid 30070. Strikingly, DreGig2E and OniGig2D were found to consist of 4 tandem units, and therefore cut into 4 parts: i, ii, iii and iv (Fig. 4B).

To further determine the origin of proteins with multiple units, the separated basic unit sequences were subjected to construct a NJ phylogenetic tree. Four major clades were formed. One was comprised of fish sequences from the order Salmoniformes (the red in Fig. 4C), two from the order Salmoniformes (the green in Fig. 4C), and one from the order Pseudocarchariniformes (the blue in Fig. 4C). This result indicated that domain duplications of these Gig2 proteins had occurred after the split of the ancestors of Salmoniformes, Cypriniformes, and Pseudocarchariniformes. In a specific order group, the former moieties of sequences with two tandem units were likely to cluster together, and the same was true to the latter parts (Fig. 4C). For DreGig2E, the unit_i was clustered with the unit_iii, and the unit_ii with the unit_iv; for OniGig2D, the unit_i was clustered with the unit_ii, and the unit_iii with the unit_iv (Fig. 4C), suggesting that both four-unit-containing proteins were likely to cluster together, and the same was true to the units in other fish species (Fig. 4B).

Lineage-specific Gene Duplication of Gig2 Gene Family

Phylogenetic trees were constructed to investigate the evolutionary history of 190 Gig2 family proteins identified above. Since Gig2 genes exhibited weak homology to PARP family members, three PARP 11 proteins from chicken, turkey and pig were included as an outgroup. As shown in Fig. 5A, Gig2 members were clustered into six major clades, except that 12 Gig2 members mainly from marine water samples and algae showed some distant relationship to these six groups (the dark brown in Fig. 5A). Most amphibian Gig2 family members were clustered into one clade (the blue in Fig. 5A), whereas fish Gig2 members were clustered into four major clades. Three lamprey Gig2 members were grouped with amphibian-related clade (the light brown in Fig. 5A), which was possibly ascribed to a rapid evolution during radiation of teleost species than was seen in lamprey and amphibian lineages [28]. Notably, Gig2 members from the close species were more likely to cluster together. This was consistent with the evolutionary relationship indicated by the taxonomy common tree (Fig. S1). For example, Gig2 members from two frogs were clustered in one subgroup, whereas those from four salamanders clustered in another (the blue in Fig. 5A). The same was true for four major fish-specific clades. These results suggested that Gig2 gene family had been extensively expanded by lineage-specific duplication.

In addition, the major four fish-specific clades were clustered dependently on chromosome locations. Specifically, while Gig2 members from zebrafish chromosome 21, stickleback chromosome group X, medaka chromosome 8 and green spotted pufferfish chromosome 3 were clustered in one clade (the red in Fig. 5A), the members from zebrafish chromosome 5, stickleback chromosome group X, medaka chromosome 9 and green spotted pufferfish chromosome 12 were clustered in another one (the pink in Fig. 5A). However, DreGig2Q on zebrafish chromosome 2 and the other four Cypriniformes members were grouped to form one different clade (the light green in Fig. 5A); there was an additional Salmoniformes-related clade that consisted of sequences without definite chromosome information (the dark green in Fig. 5A). Further, 36 Gig2 members with definite chromosome information were subjected to make an unrooted phylogenetic tree. Similar to the findings above, these genes located on two sets of fish chromosomes were clearly grouped into two major clades (the red and the pink in Fig. 5B).

To further understand the evolution of Gig2 family, we analysed the synteny of chromosomes currently available that contain Gig2 genes including zebrafish, stickleback, medaka, green spotted pufferfish and fugu (Fig. 6). First, the gene synteny of chromosomes within species was compared. The common genes were found in the flanks of Gig2 gene loci, such as 3HT7, AccQ96N67 and PDE32 in stickleback group XI and XIII, ATPase116Iso1, ATPase116Iso2, GRBLG and RARS in medaka chromosome 6 and chromosome 9, TA02 and AccQ96273 in green spotted pufferfish chromosome 3 and chromosome 12. Subsequently, the gene synteny of chromosomes among species was compared. It showed that stickleback group XI, medaka chromosome 8, green spotted puffer fish chromosome 3 and fugu scaffold 705 & 2110 shared high co-localization of Gig2 gene loci, with some common surrounding genes (shown in red and blue in the upper panel in Fig. 6). In addition, a better gene synteny was shown in stickleback group XIII, medaka chromosome 9, green spotted pufferfish chromosome 12 and fugu scaffold 101, since a stretch of common genes were found in the flank of Gig2 gene loci, with the same gene arrangement and the same transcriptional orientation (shown in brown in the lower panel in Fig. 6). Notably, the Gig2 gene members in the upper set of chromosomes in Fig. 6 were grouped together in phylogenetic tree analysis (shown in red in Fig. 5A), while the genes from the lower set were clustered in another clade (shown in pink in Fig. 5A).

Discussion

In the present study, a novel gene family was identified. Since the founding gene is named Gig2 [19] and this name is widely used in other literatures [22,29,30,31], this gene family is termed Gig2 gene family. Five crucian carp Gig2 genes were identified by expression analyses, and most of Gig2 family genes were in silico cloned by analyses of EST/GSS sequences derived from experimental data or NCBI databases, suggesting that these genes identified here are indeed active genes. Consistently, a proportion of family genes were further verified by analyses of the available genomes from fish species and amphibians. Large-scale search of the genome databases from invertebrates, reptiles, birds and mammals did not find any Gig2 homologous genes. Although 5 GSSs from marine water species and 3 ESTs from algae were homologous to Gig2 genes, it is reasonable that they might be caused by horizontal gene transfer [25]. Therefore, the data in the present study support a conclusion that Gig2 gene family is specific to non-annamites including lampreys, sharks, rays, ray-finned fishes and amphibians.

The results presented here also provide strong evidence for lineage-specific and species-specific expansion of Gig2 gene family. The direct evidence is derived from phylogenetic analyses, where fish and amphibian Gig2 members are grouped into separated clades (Fig5). Subgroups of Gig2 members are formed dependently on different orders they belong to, even on different species, suggesting involvement of extensive segmental duplication in evolution of Gig2 gene family. This notion is further verified by gene synteny analyses that there are varied gene copies in Gig2 gene loci in different fish species (Fig. 6). In addition, domain analyses revealed that some Gig2 proteins harbour 2 or 4 basic units, each of which share a conserved N-terminus and a conserved C-terminus (Fig. 4A). These basic units exhibit a close evolutionary relationship (Fig. 4B), supporting that the multiple-
Figure 5. Phylogenetic analyses of Gig2 family proteins. (A) Phylogenetic analysis of all Gig2 proteins identified in this study. A neighbour-joining tree was constructed based on analysis of 190 Gig2 family protein sequences using Geneious, with bootstrap values for 1000 replicates. Three PARP11 proteins were included as an outgroup. The evolutionary distances were computed using the Poisson correction method with the pairwise deletion option. (B) Phylogenetic analyses of Gig2 proteins with definite chromosomal information. The indication by different colours was seen in the text.

doi:10.1371/journal.pone.0060588.g005
Given the essential function of Gig2 genes under stresses, it is intriguing why this gene family is found exclusively in lower vertebrates from lampreys to amphibians. Considering that these lower organisms live in a similar environment, such as in water, half in water, or at least in very humid environments, they suffer from some common threats such as hypoxia and common pathogens, or they have some common developmental events or morphological structures. The origin of Gig2 gene family might enable non-ammioites to adapt to the aquatic environments by evolution of lineage- or species-specific phenotypic traits. For example, extensive expansion of the Claudin gene family has resulted in a subgroup of fish-specific Claudin genes, which might be involved in regulating the exchange of specific solutes with the aqueous environment in fishes [41]. However, answering the issue of what is the physiological significance of Gig2 gene family in non-ammioites should wait for clarity of these gene function in the future studies.

Methods

Ethics Statement
This study was performed in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the Chinese Academy of Sciences. The protocol was approved by the Committee on the Ethics of Animal Experiments of the Institute of Hydrobiology ( Permit Number: Y213191301). All surgery was performed under sodium pentobarbital anesthesia, and every effort was made to minimize suffering.

Cells, Virus, Fish and Induction
Carassius auratus blastulae embryonic cells (CAB) and zebrasfish liver cells (ZFL) were maintained in medium 199 supplemented with 10% fetal calf serum (FCS) at 28°C. GCRV was proliferated in CAB cells and UV inactivation of GCRV was carried out as described previously [20]. For induction, cells were seeded in 25 cm² flasks one day in advance, and washed with PBS once and then treated with 0.5 ml UV-inactivated GCHV containing medium the titers of which was 1×10⁷ TCID₅₀/ml before exposing to UV irradiation. After incubation at 28°C for 1 h, the medium of treated cells were replaced with appropriate FCS-free 199 medium. The cells were collected at 2, 6, 12, 24, 48, 72, 96 and 120 h after treatment. Control cells were treated with FCS-free 199 medium alone. RNA samples of the cultured cells and fish tissues were prepared according to a previous report [20,21]. Transfection assays were performed according to previous reports [8,21]. Zebrafish (Danio rerio) were maintained at 28.5°C on a 14 h light/10 h dark cycle. All embryos used were collected by natural spawning and staged according to standard procedures [20,21].

Sequence Cloning and RT-PCR
Primers were designed against the previously screened Gig2 ESTs to PCR amplify both 5’ and 3’ termini of GadGig2 genes by RACE from the SMART cDNA library we previously constructed [24]. The PCR products were ligated into the pMD18-T vector (Takara) and transformed into competent E. coli cells. The total RNA extraction and Semi-quantitative RT-PCR and Real-time PCR were performed as the previous description [20,21].

Analysis of Gig2 Genes in Genome Databases
The 5 crucian carp Gig2 sequences were first subjected to search the zebrafish genome database using BLAST search against NCBI zebrafish genome database (Zv9.0). The annotated Gig2 homologues retrieved by first search were further used as sequence baits for TBLASTN. Mapview of Gig2 genes in zebrafish genomes was revised from the mapview provided by NCBI. The same analyses were performed in genome databases from torafugu (Takifugu rubripes, FUGU4), spotted green pufferfish (Tetraodon nigroviridis, TETRAODON8), three-spinned stickleback (Gasterosteus aculeatus, BROADS1), medaka (Oryzias latipes, MEDAKA1), and western clawed frog (Xenopus tropicalis, JGI 4.2).

Sequences Survey and in Silico Cloning
Crucian carp and zebrafish Gig2 sequences were used as query to BLAST search all the sequence databases on NCBI and ensembl. The hit sequences were downloaded and organised in a spreadsheet. The ESTs were assembled by the CAP3 online service at http://pbil.univ-lyon1.fr/cap3.php. Protein sequences of the assembled EST sequences were deduced by DNAMAN for Linux (Lynnon Corporation).

Multiple Sequence Alignment and Phylogenetic Analysis
Multiple sequences alignments were performed by ClustalW. The alignment sequences were visualized by boxshade (http://www.ebi.ac.uk/Tools/BOX_form.html). The evolutionary history of all the studied Gig2 protein sequences was inferred using the Neighbour-Joining method with the bootstrap test 1000 replicates in the software Geneious (http://www.geneious.com/). The evolutionary distances were computed using the Poisson correction method. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option).

Synteny Analysis
Synteny analysis was performed by the pipeline of SynBlast which uses the genomic region around a focal reference gene to retrieve candidates for homologous regions from a collection of target genomes and ranks them in accord with the available evidence for homology [44]. Briefly, here we used Gig2 as a focal reference gene to retrieve candidate mRNA sequence from 5 Mbp...
regions at both of its flanks in each chromosome sequence. All the retrieved mRNA sequences were in turn used as query to BLAST search the chromosome sequences. The mRNA sequences which hit more than two chromosome sequences were collected in Table S4 and figured out.

Supporting Information

Figure S1 The common taxonomy tree was constructed using NCBI taxbrowser with 10 representative species and 52 species (in blue) where using NCBI taxbrowser with 10 representative species.

References

The detailed information for all the identified Gig2 family genes, including in silico information, cDNA sequences, ORF sequences, putative protein sequences, and the locations corresponding to putative PARP-like domains.

(XLS)

Table S2 The detailed information for 990 EST sequences homologous to Gig2 in the study.

(XLS)

Table S3 Primers used for RACE-PCR and expression analyses in the present studies.

(DOC)

Table S4 Genes used in gene synergy analysis and corresponding accession numbers.

(DOC)

Author Contributions

Conceived and designed the experiments: YBZ TKL. Performed the experiments: YBZ TKL JJ JS SL YL. Analyzed the data: YBZ TKL. Contributed reagents/materials/analysis tools: YBZ JFG. Wrote the paper: YBZ TKL JFG.

Phylogeny of Gig2 Gene Family

Table S1 The common taxonomy tree was constructed using NCBI taxbrowser with 10 representative species and 52 species (in blue) where Gig2 sequences were found. The representative amniotes were coloured in light blue to emphasize the lack of Gig2 genes. The taxonomy IDs used in this tree are listed in the Table 1 or Table S1. The bracket immediately after the name of each species indicates the abbreviation of the species.

(PDF)

Table S1 The detailed information for all the identified Gig2 family genes, including in silico information, cDNA sequences, ORF sequences, putative protein sequences, and the locations corresponding to putative PARP-like domains.

(XLS)

Table S2 The detailed information for 990 EST sequences homologous to Gig2 in the study.

(XLS)

Table S3 Primers used for RACE-PCR and expression analyses in the present studies.

(DOC)

Table S4 Genes used in gene synergy analysis and corresponding accession numbers.

(DOC)

Author Contributions

Conceived and designed the experiments: YBZ TKL. Performed the experiments: YBZ TKL JJ JS SL YL. Analyzed the data: YBZ TKL. Contributed reagents/materials/analysis tools: YBZ JFG. Wrote the paper: YBZ TKL JFG.

Phylogeny of Gig2 Gene Family

Table S1 The common taxonomy tree was constructed using NCBI taxbrowser with 10 representative species and 52 species (in blue) where Gig2 sequences were found. The representative amniotes were coloured in light blue to emphasize the lack of Gig2 genes. The taxonomy IDs used in this tree are listed in the Table 1 or Table S1. The bracket immediately after the name of each species indicates the abbreviation of the species.

(PDF)

Table S1 The detailed information for all the identified Gig2 family genes, including in silico information, cDNA sequences, ORF sequences, putative protein sequences, and the locations corresponding to putative PARP-like domains.

(XLS)

Table S2 The detailed information for 990 EST sequences homologous to Gig2 in the study.

(XLS)

Table S3 Primers used for RACE-PCR and expression analyses in the present studies.

(DOC)

Table S4 Genes used in gene synergy analysis and corresponding accession numbers.

(DOC)

Author Contributions

Conceived and designed the experiments: YBZ TKL. Performed the experiments: YBZ TKL JJ JS SL YL. Analyzed the data: YBZ TKL. Contributed reagents/materials/analysis tools: YBZ JFG. Wrote the paper: YBZ TKL JFG.

Phylogeny of Gig2 Gene Family

Table S1 The common taxonomy tree was constructed using NCBI taxbrowser with 10 representative species and 52 species (in blue) where Gig2 sequences were found. The representative amniotes were coloured in light blue to emphasize the lack of Gig2 genes. The taxonomy IDs used in this tree are listed in the Table 1 or Table S1. The bracket immediately after the name of each species indicates the abbreviation of the species.

(PDF)

Table S1 The detailed information for all the identified Gig2 family genes, including in silico information, cDNA sequences, ORF sequences, putative protein sequences, and the locations corresponding to putative PARP-like domains.

(XLS)

Table S2 The detailed information for 990 EST sequences homologous to Gig2 in the study.

(XLS)

Table S3 Primers used for RACE-PCR and expression analyses in the present studies.

(DOC)

Table S4 Genes used in gene synergy analysis and corresponding accession numbers.

(DOC)

Author Contributions

Conceived and designed the experiments: YBZ TKL. Performed the experiments: YBZ TKL JJ JS SL YL. Analyzed the data: YBZ TKL. Contributed reagents/materials/analysis tools: YBZ JFG. Wrote the paper: YBZ TKL JFG.

Phylogeny of Gig2 Gene Family

Table S1 The common taxonomy tree was constructed using NCBI taxbrowser with 10 representative species and 52 species (in blue) where Gig2 sequences were found. The representative amniotes were coloured in light blue to emphasize the lack of Gig2 genes. The taxonomy IDs used in this tree are listed in the Table 1 or Table S1. The bracket immediately after the name of each species indicates the abbreviation of the species.

(PDF)

Table S1 The detailed information for all the identified Gig2 family genes, including in silico information, cDNA sequences, ORF sequences, putative protein sequences, and the locations corresponding to putative PARP-like domains.

(XLS)

Table S2 The detailed information for 990 EST sequences homologous to Gig2 in the study.

(XLS)

Table S3 Primers used for RACE-PCR and expression analyses in the present studies.

(DOC)

Table S4 Genes used in gene synergy analysis and corresponding accession numbers.

(DOC)

Author Contributions

Conceived and designed the experiments: YBZ TKL. Performed the experiments: YBZ TKL JJ JS SL YL. Analyzed the data: YBZ TKL. Contributed reagents/materials/analysis tools: YBZ JFG. Wrote the paper: YBZ TKL JFG.
42. Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF (1995) Stages of embryonic development of the zebrafish. Dev Dyn 203: 253–310.
43. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(ΔΔCT) Method. Methods 25: 402–408.
44. Lehmann J, Stadler PF, Prohaska SJ (2008) SynBlast: assisting the analysis of conserved synteny information. BMC Bioinformatics 9: 351.