A Genome-wide Survey, Model Selection, Phylogenetic Analysis and Protein-protein Interactional Network Profile of the Metazoan PRKAG Genes from 22 Vertebrate Genomes

Wuyi Liu1*, Huifang Lv1 and Alireza Seidavi2*
1Department of Biology Sciences, Biological and Food Engineering Faculty, Fuyang Normal University, China
2Department of Animal Science, Rasha Branch, Islamic Azad University, Rasht, Iran
*For correspondence: lwui@163.com; lwycu@163.com; alirezaseidavi@iaurasht.ac.ir

Abstract

AMP-activated protein kinase (AMPK) is a cellular energy sensor which regulates organizational energy and phosphorylates diverse enzymes, playing key roles in determining different metabolic phenotypes and the energy metabolism equilibrium. This genome-wide survey identified and analyzed 266 unique AMPK gamma subunit (PRKAG) gene sequences from the recently updated genomes of human and 21 animals. After the model selection of phylogenetic analysis parameters, all the 266 protein sequences of these identified human and animal PRKAG genes were used to compute ML (Maximum Likelihood) phylogenetic tree under the evolutionary WAG+G and the protein-protein interactional network analysis using the STRING database. Subsequently, the phylogenetic analysis verified and found all these PRKAG gene sub-families from the amphibian, fish, avian, and mammal genomes formed large monotonous phylogenetic clusters except for three frog PRKAG sequences. Particularly, all the primate PRKAG genes formed the super-class of their sole phylogenetic clusters. Furthermore, the mapped network profile analysis revealed that these PRKAG genes were functionally assembled and/or enriched in the energy metabolism, immune responsive and/or inflammatory and myosin or myosin heavy chain interacting proteins related signaling pathways. In addition, most of these proteins are interrelated and formed a tight protein interactional network of energy metabolism. In conclusion, the observed data and information of protein-protein interaction will contribute to understanding and analyzing the possible protein functions and interacted pathways of animal PRKAG genes and/or other AMPK subunit genes. The metazoan PRKAG gene sequences and identified dataset may be beneficial for related developmental and physiological research topics. These findings may also be potentially applied to explore and resolve the critical issues of disease-induced loss that many medical researchers are taking into research. © 2019 Friends Science Publishers

Keywords: Genome-Wide Survey; Phylogenetic Analysis; Evolutionary Algorithm; Protein-Protein Interactional network

Introduction

AMP-activated protein kinase (AMPK) is implicated as a cellular energy sensor which regulates organizational energy and phosphorylates diverse enzymes, including protein effectors and metabolic enzymes. In many studies and reports, AMPK plays key roles in determining different metabolic phenotypes and the energy metabolism equilibrium of biological organisms as well as other ATP-binding cassette proteins (Li et al., 2019) or ADG related uncoupling proteins (Jiang et al., 2019). Particularly, it appears as the primary energy inductor and regulator of cellular energy homeostasis acting in the long term and/or the short term in eukaryotes and to ameliorate the metabolic distress by post-translational metabolic enzyme modifications (Hardie, 2007; Hardie et al., 2012; Ahmad et al., 2015; Garcia and Shaw, 2017; Neumann and Viollet, 2019; Rashtchizadeh et al., 2019). In fact, AMPK and its subunits are also regarded the potential sensor molecules concerned in human and animal diseases and biomedicine targets for treating metabolic decline in aging and metabolism-dependent disorders (Burkewitz et al., 2014; Ahmad et al., 2015; Li et al., 2015; Carling, 2017; Troncone et al., 2017; Willows et al., 2017; Neumann and Viollet, 2019; Rashtchizadeh et al., 2019).

In structure, AMPK is made up of a catalytic alpha (α) subunit, a non-catalytic regulatory beta (β) subunit, and a connector gamma (γ) subunit. AMPK is actually a particularly economic three-dimensional complex with its effective α and γ core subunits and the additional β C-terminal domain, in which the β C-terminal domain is shuttled in the other two subunits. The α subunit has the main site threonine 172 (T172) of phospho-dependent disorders (Burkewitz et al., 2014; Ahmad et al., 2015; Li et al., 2015; Carling, 2017; Troncone et al., 2017; Willows et al., 2017; Neumann and Viollet, 2019; Rashtchizadeh et al., 2019).
Garcia and Shaw, 2017; Neumann and Viollet, 2019; Rashchizadeh et al., 2019). On the other hand, there are the cystathionine-β-synthase (CBS) motifs and other internal domains indwelling in the AMPK γ subunit (PRKAG) to form those two binding sites required for the enzymatic activities of specific allosteric activators of AMP (Burkewitz et al., 2014; Carling, 2017; Troncone et al., 2017; Neumann and Viollet, 2019; Rashchizadeh et al., 2019).

At present, many studies in the human and animal genetics and development research suggested that these AMPK subunit genes were functionally important in the nutrient intake and physiological and pathological changes in the energy metabolism equilibrium of skeletal and cardiac muscle tissues of human and pig with both physiological and genetic evidences (Milan et al., 2000; Andersson, 2003; Zhao et al., 2006; Scheffler et al., 2011; Yin et al., 2012; Scheffler et al., 2016; Yang et al., 2016). Overall, PRKAG2 mutations lead to a decrease in the ability of AMP to bind AMPK, resulting in excessive storage of glycogen and causing disease (Gollob et al., 2001; Burwinkel et al., 2005; Carling, 2017; Rashchizadeh et al., 2019). For example, in human genetics, previous studies have shown that the genetic mutated locus from Arg 302 to Gln 302 in the PRKAG2 gene usually rendered the human AMPK γ2 subunit to be a structurally inactive or deficient kinase, resulting in the pre-excitation syndrome (Gollob et al., 2001), i.e., Wolff-Parkinson-White syndrome. Furthermore, the fatal congenital glycogen hyperactivity was reported as mainly caused by the R531Q locus mutation of human PRKAG2 gene (Burwinkel et al., 2005). In the Hampshire pigs, Milan et al. (2000) found and reported a non-conservative single nucleotide polymorphism (i.e., a dominant RN- mutation) of R200Q locus of the swine PRKAG3 alleles (Milan et al., 2000; Andersson, 2003; Scheffler et al., 2011; Scheffler et al., 2016), which was identified as significantly associated with the excess glycogen content in pork. Later, many mutant substitutions like the V199I and AMPKγ3R200Q loci of the swine PRKAG3 alleles and mutant substitutions of the bovine PRKAG1 alleles were also identified and evaluated in different pig populations in recent years (Scheffler et al., 2011; Scheffler et al., 2016). Since then, similar mutant allele cases of PRKAG genes reported in animal muscle tissues were also characterized in the allele loci of chicken PRKAG3 gene (Zhao et al., 2006; Yin et al., 2012; Yang et al., 2016) and dairy cows’ PRKAG1 (Ahmad et al., 2015).

Therefore, the genetic mutations of many PRKAG alleles have been the hotspot subject of domestic animal researchers in the genomics era.

Previously, an initial study has been conducted to mainly identify avian PRKAG genes using the genome-wide BLAST searches (Liu, 2017). However, more and more sequences of PRKAG and related genes remain to be further explored and classified, whereas their functional interactional network and related signaling pathways remain to be characterized, especially these in domestic animal genomes. The present study conducted an updated genome-wide survey and the large-scale phylogenetic analysis to further analyze and characterize the human and animal PRKAG genes identified from 22 vertebrate genomes. Furthermore, the study also carried out the mapped interaction network profile analysis of PRKAG genes.

Materials and Methods

Data Retrieval and Sequence Alignment of Human and Animal PRKAG Genes from the Genomic Databases

The genome-wide survey was carried out using human PRKAG genes as input query sequence of BLAST searches. That is, with a few pair nucleotide and protein sequences of human PRKAG genes, we systematically performed previously used PSI-BLAST searches (Luo et al., 2019) to retrieve all the putative PRKAG genes against the latest released genomes of 22 vertebrates in the NCBI genome resources (URL: http://www.ncbi.nlm.nih.gov/genome/guide/build.shtml), i.e., Actinopterygii, Bovidae, Bos taurus, Bos mutus, Bubalus bubalis, Canis lupus, Canis lupus familiaris, Capra hircus, Gorilla gorilla, Homo sapiens, Mus musculus, Ovis aries, Pongo abelii, Rana esculenta, Sus scrofa, Panthera tigris, Gallus gallus, Danio rerio, Meleagris gallopavo, Anas platyrhynchos, Taeniopygia guttata, Xenopus tropicalis, Xenopus laevis. In total, we had retrieved more than 600 sequences of the nucleotide and protein sequences of the putative PRKAG genes (i.e., PRKAG1, PRKAG2, and PRKAG3) and their isoform proteins. All of the retrieved nucleotide and protein sequences of vertebrate PRKAG genes were subsequently used in the sequence alignments that were transformed into FASTA sequence files with the Windows software package of ClustalX version 2.0 (Larkin et al., 2007). Meanwhile, the incomplete sequences and highly divergent sequences were excluded from the further analysis, because the incomplete sequences and sequence indels and/or gaps those sequences of would result in uncertain alignments.

Model Selection of Phylogenetic Tree Parameters of Vertebrate PRKAG Genes

MEGA version 6.06 (Tamura et al., 2013) was used to make model selection analysis of phylogenetic tree parameters for nucleotide and amino acid sequences. By default, MEGA presents for each model the estimated values of Bayesian information criterion (BIC), AICc (Akaike Information Criterion, corrected), Maximum Likelihood ratio (lnL), the estimated values of shape parameter of the discrete Gamma distribution (+G) and other parameters (including invariant sites and tree branch lengths) as applicable (Nei and Kumar, 2000; Tamura et al., 2013). Whenever applicable, the
software package MEGA shows the estimates of gamma shape parameter and/or the estimated fraction of invariant sites (Nei and Kumar, 2000; Tamura et al., 2013), whereas the non-uniformity of evolutionary rates among sites are computed under a discrete distribution of Gamma (+G) in the establishing of evolutionary phylogenetic trees (Tamura et al., 2013). Depending on the model, the goodness-of-fit of each model is mainly measured by the Bayesian information criterion (BIC) and the corrected Akaike information criterion (AICc) in calculation (Tamura et al., 2013). The simulated model parameters of phylogenetic analysis are briefly shown in Table 1 and the detailed phylogenetic model parameters provided in Supplementary Table 2.

**Phylogenetic Analyses of Vertebrate PRKAG Genes**

The evolutionary fit of major substitution models of phylogenetic tree parameters were compared and selected for nucleotide and amino acid sequences of vertebrate PRKAG genes (see the model selection part or simulated phylogenetic model parameters provided in Supplementary Table 2 for details). The finally identified putative protein sequences of vertebrate PRKAG genes were also used to compute the ML phylogenetic trees under the optimized evolutionary substitution model measured by the dual criterions of BIC and AICc (Nei and Kumar, 2000; Tamura et al., 2013), respectively. Among the subsequent analyses, ML phylogenetic trees were computed and reconstructed with the subtree-pruning-regrafting search heuristic method in all the established trees (Tamura et al., 2013).

**Protein Interactional Network Analysis using the STRING Database**

The Search Tool for the Retrieval of Interacting Genes (STRING; URL: https://string-db.org/) is a pre-built database of abundant known experimental and predicted protein interactions (Szklarczyk et al., 2015; Szklarczyk et al., 2019). To investigate all the possible protein–protein interactions between the vertebrate PRKAG genes retrieved and identified from the updated genome-wide survey in 22 vertebrates, the STRING database was used for the creation of protein–protein interactional network (PIN) files as previously described (Szklarczyk et al., 2015; Szklarczyk et al., 2019). In practice, the STRING database version 11 online was used to evaluate and analyze the protein–protein interactions of vertebrate PRKAG1 and PRKAG2 and PRKAG3 protein sequences (Szklarczyk et al., 2019). The finally created and extracted PIN network files were explored and compared based on the PRKAG gene expression products (protein sequences) and their PIN network interactions and/or topological characteristics reported by more than 20 reports or studies of vertebrate PRKAG genes.

**Results**

**Genome-wide Survey and Identification of Vertebrate PRKAG Genes**

In total, more than 600 sequences were initially retrieved and screened from the NCBI genome resources (URL: http://www.ncbi.nlm.nih.gov/genome/guide/build.shtml). The nucleotide and protein sequences of vertebrate PRKAG genes were subsequently used and aligned by ClustalX 2.0 (Larkin et al., 2007). Those incomplete sequences and highly divergent sequences were excluded from the further analysis, since the incomplete sequences and sequence indels and/or gaps those sequences of would result in uncertain alignments. Furthermore, other removed records are predicted transcription factors or hypothetical proteins. After removing these incomplete sequences and uncertain and/or redundant sequence records, we then obtained 266 unique protein sequences and 266 unique nucleotide sequences of the putative PRKAG genes of human and animals for further analyses. It should be noted that all the 266 pairs of unique protein and nucleotide sequences retrieved from the genomes of 22 vertebrate species were carefully checked and identified and verified again. Finally, these 266 pairs of unique PRKAG protein and nucleotide sequences were further checked and examine by all the members of our group. These final used sequence dataset included a total of 266 unique protein sequences from the genomes of 22 different animal species (Supplementary Table 1), including 16 sequences from Acinonyx jubatus (Cheetah), 18 sequences from Bos taurus (cattle, bovine), 5 sequences from Bos mutus (wild yak), 18 sequences from Bubalus bubalis (water buffalo), 15 sequences from Canis lupus (gray wolf), 18 sequences from Canis lupus familiaris (dog), 18 sequences from Capra hircus (goat), 10 sequences from Gorilla gorilla (western gorilla), 22 sequences from Homo sapiens (human), 15 sequences from Mus musculus (mouse), 13 sequences from Ovis aries (sheep), 8 sequences from Pongo abelii (Sumatran orangutan), 13 sequences from Rattus norvegicus (rat), 14 sequences from Sus scrofa (pig), 4 sequences from Panthera tigris (tiger), 13 sequences from Gallus gallus (chicken), 13 sequences from Danio rerio (zebrafish), 3 sequences from Meleagris gallopavo (turkey), 11 sequences from Anas platyrhynchos (duck), 3 sequences from Taeniopygia guttata (zebra finch), 10 sequences from Xenopus tropicalis (Western clawed frog) and 6 sequences from Xenopus laevis (African clawed frog).

The corresponding information of these 266 pairs of unique protein and nucleotide sequences identified for further analyses of the putative PRKAG genes of human and animals are summarized in Supplementary Table 1.

**Model Selection and Phylogenetic Analyses of Vertebrate PRKAG Genes**

In the study, the analyzed ML phylogenetic trees were used to identify and elucidate the evolutionary and functional
sequences were finally identified and used in the phylogenetic analysis (Supplementary Table 1). In the ML phylogenetic tree of the 266 unique putative protein sequences of vertebrate PRKAG genes, almost all the protein sequences were clustered into four large classes of amphibian, fish, avian, and mammal PRKAG genes accordingly (Fig. 1). However, a few exceptions do exist in the ML phylogenetic tree (Fig. 1). From all the three subfamilies of PRKAG sequences (i.e., PRKAG1, PRKAG2, and PRKAG3), the PRKAG sequences identified from the amphibian, fish, avian, and mammal genomes formed large
The Metazoan PRKAG Genes Identified from 22 Vertebrate Genomes / Intl. J. Agric. Biol., Vol. 00, No. 0, 201x

monotonous phylogenetic clusters except for three frog PRKAG sequences (NP_001096506.1, NP_001085968.1, and XP_012811772.1) that were close to those of fish. Furthermore, all the primate PRKAG sequences formed a sole super-class of their own phylogenetic clusters in blue circles (Fig. 1). In brief, the ML phylogenetic analysis of putative vertebrate PRKAG protein sequences revealed that these PRKAG gene subfamilies (PRKAG1, PRKAG2 and PRKAG3) had their own independent common ancestor genes (Fig. 1).

Protein Interactional Network Analysis of Human and Animal PRKAG Protein Sequences

The protein-protein interaction of human and animal PRKAG protein sequences were analyzed and displayed using the STRING database (URL: http://string-db.org/) of experimental and predicted protein interactions. The STRING database is able to detect and find the molecular enrichment and functional correlation in a close protein interaction and experimentally validated contacted network among the genes implied the importance of any experimental subject pathways (Fig. 2). The picture was the evidence network view of key hub proteins (e.g., PRKAG1, PRKAG2 and PRKAG3) and their relevant partners (Figure 2). These interacted proteins mainly included GDE1 (Glycerophosphodiesphodiesterase 1), TCF3 (Transcription factor E2-alph, a basic helix-loop-helix transcription factor involved in the initiation of neuronal differentiation), TIP1 (Tight junction protein ZO-1 belonging to the MAGUK family whose N and C terminals may be involved in the cellular tight junction assembly), GYPA (Glycophorin-A, a receptor for influenza virus and a major intrinsic membrane protein critical for the function of SLC4A1), APRT (Adenine phosphoribosyltransferase catalyzing salvage reactions resulting in the formation of AMP), ATP8A2 (Phospholipid-transporting ATPase IB, a catalytic component of a P4-ATPase flippase complex), MLLT4 (Araf, belonging to an adhesion system playing a
role in the organization of homotypic, interneuronal and heterotypic cell-cell adherens junctions, ENSG00000160200 (Cystathionine-beta-synthase), MYH6 (Myosin-6, belonging to the Myosin family), MYH14 (Myosin-14, belonging to the Myosin family), ACACA (Acetyl-CoA carboxylase 1), PRKAA1 (AMP-activated protein kinase catalytic subunit alpha-1, a catalytic subunit of AMPK), PRKAA2 (AMP-activated protein kinase catalytic subunit alpha-2, a catalytic subunit of AMPK), PRKAG1 (AMP-activated protein kinase subunit gamma-1, the AMP/ATP-binding subunit of AMPK), PRKAG2 (AMP-activated protein kinase subunit gamma-2, the AMP/ATP-binding subunit of AMPK), and PRKAG3 (AMP-activated protein kinase subunit gamma-3, the AMP/ATP-binding subunit of AMPK). Besides GYP A (Glycophorin A), most of these protein partners (encoded by corresponding genes) are interrelated and formed a tight protein interactional network of energy metabolism (Fig. 2). In particular, the main hub proteins of the protein–protein interactional network are PRKAG1, PRKAG2, PRKAG3, PRKAA1, PRKAA2, APRT, ATP8A2, ENSG00000160200, MYH6, MYH14 and ACACA, interacting with no less than 3 protein partners or enzymes. Interestingly, they also linked to many immune responsive and/or inflammatory factors and myosin proteins (Fig. 2), such as MYH6 (myosin 6) and MYH14 (myosin 14).

Discussion

In this study, we identified and characterized 266 unique PRKAG genes from the updated genomes of human and 21 animals. In the present study, accurate ML phylogenetic analysis was used to identify putative homologous and/or orthologous relationships of these PRKAG genes in 22 different species with the presupposed corresponding model selection of phylogenetic trees. The phylogenetic clusters of PRKAG gene family analyzed here were in agreement much with the physiological and genetics classification of previous studies and reports with a more diversity observed among different vertebrate species. As mentioned in the corresponding part of materials and methods, these 266 finally identified putative protein sequences of human and animal PRKAG gene were also used to compute the ML phylogenetic trees under the optimized evolutionary WAG+G model measured by the dual criterions of BIC and AICc as shown in Table 1 and Supplementary Table 2 (Nei and Kumar, 2000; Tamura et al., 2013), respectively.

In practice, the protein–protein interaction of human and animal PRKAG protein sequences were analyzed using the STRING database of experimental and predicted protein interactions. The mapped integrative profile of STRING protein network was built to probe into the functional protein interactions of vertebrate PRKAG genes (e.g., PRKAG1, PRKAG2 and PRKAG3) and their relevant partners. Interestingly, these interacted proteins were linked to many immune responsive and/or inflammatory factors and myosin proteins (Fig. 2), such as MYH6 (myosin 6) and MYH14 (myosin 14), whereas the previous reports indicated that PRKAG results in phosphorylation of the myosin heavy chains (Williams and Coluccio, 1995; Rosenberg and Ravid, 2006; Yan et al., 2018) and there was a cross talk between circadian rhythm and coronary heart disease identified by multiple correlation analysis (Yan et al., 2018). This phenomenon is in accord with previous studies that both the muscle and non-muscle myosin proteins are involved in the phosphorylation of ATP (Adenosine Triphosphate) and responses in energy metabolism and inflammatory (Williams and Coluccio, 1995; Musi et al., 2003; Rosenberg and Ravid, 2006; Yan et al. 2018; Vilchinskaya et al., 2018), which is evident to be associated with the function and activities of PRKAG genes. A recent review regarded the AMP-activated protein kinase genes (especially the PRKAG genes) as the key triggers for the disuse-induced skeletal muscle remodeling on some molecular targets (Vilchinskaya et al., 2018). In addition, there was a significant association of AMPK subunit gamma subunit (PRKAG) gene polymorphisms with growth, feed intake, and feed efficiency in meat-type chickens (Liu et al., 2016).

Meanwhile, the online resources of STRING database provide both these experimental data as well as the bioinformatics data and those predicted interaction information with confidence scores. The PIN profile of PRKAG kinases and their protein partners suggest that these proteins associated with short distances to each other in the functionally interacted network are more likely to share the common biological functions (Szklarczyk et al., 2015; Yan et al., 2018; Szklarczyk et al., 2019), whereas those interactive neighbors are more likely to have identical biological function than non-interactive ones (Szklarczyk et al., 2015; Yan et al., 2018; Szklarczyk et al., 2019). It is inferred that the query kinases and their interactive protein partners may form a complex to perform particular functions or involved in the same functional pathways. In brief, these STRING experimental and predicted protein interaction data and corresponding resulted protein interactional network will contribute to understanding and analyzing the possible protein functions and interacted pathways of the PRKAG genes and/or other AMPK subunit genes.

Conclusion

In the present study, phylogenetic analyses indicated that all the identified gene sequences belong to three sub-families of PRKAG genes forming four large super-classes of monotonous phylogenetic clusters (i.e., the amphibian, fish, avian, and mammal monotonous clusters). In particular, those PRKAG gene sequences identified from the amphibian, fish, avian and mammal genomes formed large monotonous phylogenetic clusters except for three frog PRKAG sequences that were close to those of fish.
Furthermore, the mapped network profile of PIN analysis revealed that the PRKAG genes were functionally assembled and/or enriched in energy metabolism, immune responsive and/or inflammatory, and myosin interacting proteins related signaling pathways. These observed data and information of protein-protein interaction will contribute to understanding and analyzing the possible protein functions and interacted pathways of vertebrate PRKAG genes and/or other AMPK subunit genes. The resultant findings will also be potentially applied to explore and resolve the critical issues of disease-induced loss that many medical researchers are taking into research.

Acknowledgments

This work was supported by the Major Projects of Horizontal Cooperation between Fuyang Municipal Government and Fuyang Normal University (No. XDHX201725) and Anhui Provincial Educational Commission National Science Foundation (No. KJ2016SD47) and the Project of National Natural Science Foundation of China (No.31301965).

References


(Received 07 May 2019; Accepted 29 July 2019)