

Evolution of Hepatic Glucose Metabolism: Liver-Specific Glucokinase Deficiency Explained by Parallel Loss of the Gene for Glucokinase Regulatory Protein (GCKR)

Zhao Yang Wang¹, Ling Jin², Huanran Tan^{2*}, David M. Irwin^{1,2,3*}

1 Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada, **2** Department of Pharmacology, Peking University, Health Science Center, Beijing, China, **3** Banting and Best Diabetes Centre, University of Toronto, Toronto, Ontario, Canada

Abstract

Background: Glucokinase (GCK) plays an important role in the regulation of carbohydrate metabolism. In the liver, phosphorylation of glucose to glucose-6-phosphate by GCK is the first step for both glycolysis and glycogen synthesis. However, some vertebrate species are deficient in GCK activity in the liver, despite containing *GCK* genes that appear to be compatible with function in their genomes. Glucokinase regulatory protein (GCKR) is the most important post-transcriptional regulator of GCK in the liver; it participates in the modulation of GCK activity and location depending upon changes in glucose levels. In experimental models, loss of GCKR has been shown to associate with reduced hepatic GCK protein levels and activity.

Methodology/Principal Findings: *GCKR* genes and *GCKR*-like sequences were identified in the genomes of all vertebrate species with available genome sequences. The coding sequences of *GCKR* and *GCKR*-like genes were identified and aligned; base changes likely to disrupt coding potential or splicing were also identified.

Conclusions/Significance: *GCKR* genes could not be found in the genomes of 9 vertebrate species, including all birds. In addition, in multiple mammalian genomes, whereas *GCKR*-like gene sequences could be identified, these genes could not predict a functional protein. Vertebrate species that were previously reported to be deficient in hepatic GCK activity were found to have deleted (birds and lizard) or mutated (mammals) *GCKR* genes. Our results suggest that mutation of the *GCKR* gene leads to hepatic GCK deficiency due to the loss of the stabilizing effect of GCKR.

Citation: Wang ZY, Jin L, Tan H, Irwin DM (2013) Evolution of Hepatic Glucose Metabolism: Liver-Specific Glucokinase Deficiency Explained by Parallel Loss of the Gene for Glucokinase Regulatory Protein (GCKR). PLoS ONE 8(4): e60896. doi:10.1371/journal.pone.0060896

Editor: Klaus Brusgaard, Odense University Hospital, Denmark

Received: November 30, 2012; **Accepted:** March 4, 2013; **Published:** April 1, 2013

Copyright: © 2013 Wang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by a grant from the National Science Foundation of China (NSFC) -Canadian Institutes of Health Research (CIHR) China-Canada Joint Health Research Initiative to HT (81061120525) and DMI (CCI-109605). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: david.irwin@utoronto.ca (DI); tanlab@bjmu.edu.cn (HT)

Introduction

Glucose, a major source of energy for all tissues, is obtained from the diet and stored as glycogen in the liver and muscle when in excess. Storage and release of glucose is a tightly regulated process involving numerous enzymes and regulatory proteins, with glucokinase (GCK) being a key regulatory enzyme [1]. GCK is an isozyme of the hexokinases, which catalyzes the phosphorylation of six-carbon sugars. GCK differs from the other hexokinases in its affinity for glucose and end-product inhibition [2]. GCK is the major glucose-phosphorylating enzyme in the liver, pancreatic islet beta-cells, and a few other glucose-sensing cells of the gut and brain [1–3]. In the liver, GCK is the first, and the rate-limiting, step in glucose utilization leading to glucose storage as glycogen, while in pancreatic beta-cells GCK acts as a glucose sensor and controls the secretion of insulin [2,3]. GCK also appears to have similar glucose sensing functions in some cells of the gut and the brain [1,3]. Mutations that prevent *GCK* expression or function in liver and pancreatic beta-cells are known to result in the maturity onset diabetes of the young 2 (MODY2) form of diabetes [4].

Similarly, activating mutations of *GCK* causing elevated GCK activity in liver and pancreatic beta-cells are also known to result in persistent hyperinsulinemic hypoglycemia of infancy (HPPI), in which high insulin levels induce low blood glucose levels [5].

A pair of tissue-specific promoters drives the expression of the *GCK* gene: an upstream beta-cell specific promoter that is also used in gut and neuronal cells; and a downstream liver-specific promoter [3,6]. Liver-specific *GCK* expression is absolutely dependent upon the presence of insulin and is repressed by glucagon [3,6–8]. Consequently, *GCK* expression is observed to increase after feeding or with insulin treatment, and expression is repressed when starved or with insulin deficiency [7,8]. Hormonal regulation allows GCK to be active only when there is excess blood glucose, but the insulin effect on GCK is not potentiated by high glucose concentration [3,6–8], in fact, glucose represses *GCK* expression, potentially to protect phosphate homeostasis [9]. In contrast to the nutritional regulation of *GCK* gene expression in the liver, little change in *GCK* mRNA levels is seen with changes in blood glucose levels or insulin in pancreatic beta-cells, although some change in mRNA stability might occur [3,10]. The

differences in the regulation of *GCK* expression in these two major sites of expression are likely related to the tissue-specific functions: whereas the liver only needs *GCK* expression when there is excess blood glucose, pancreatic islet beta-cells requires constant expression of *GCK* as a sensor for measuring blood glucose levels [2,3].

In addition to regulation at the transcriptional level, GCK is also regulated in the liver at the post-transcriptional level by glucokinase regulatory protein (GCKR) [11,12]. GCKR is primarily expressed in the liver [13,14], and possibly the brain [15], with little if any expression in pancreatic beta-cells [16,17]. In the liver GCKR functions as both an inhibitor and a nuclear binding protein for GCK [11,18,19]. Under low glucose conditions, GCKR binds, inactivates and sequesters cytosolic GCK into the nucleus. With an influx of glucose, GCK within the nucleus is re-activated and returns to the cytoplasm [20–22]. Phosphorylated forms of fructose, whose levels reflect glucose metabolism, modulate the binding of GCK with GCKR – fructose-1-phosphate weakens GCK and GCKR interaction, while fructose-6-phosphate promotes GCKR-GCK binding and subsequent GCK inactivation [3,12,23]. GCKR also has a role in stabilizing GCK and preventing its degradation [20–22]. Mice deficient in GCKR have normal or raised *GCK* mRNA levels but decreased liver GCK protein concentration and activity [24,25]. Furthermore, overexpression of *GCKR* in hepatocytes by adenoviral vectors increases both GCK protein and enzymatic activity [26]. These observations suggest that the abundance of GCKR may be an important regulator of GCK protein levels specifically in the liver.

The livers of some mammalian (e.g., cow and cat) and non-mammalian (e.g., birds and lizards) vertebrate species have been reported to be deficient in GCK activity [27]. Hexokinases have been examined from the livers of numerous vertebrate species and their chromatographic profiles suggest that the livers of some mammals, including ruminants (e.g., cow and sheep), bats, and cats, as well as other vertebrate species such as birds and some reptiles, have little if any GCK activity [13,27–31]. In these surveys, species deficient in GCK were found to possess activities of the three other hexokinases (i.e., hexokinases I, II, or III) at levels similar to those of species that have GCK [27–31] suggesting that the specific depletion of GCK in these species was not compensated by an increase in the activity of another hexokinase. As these species do not demonstrate symptoms of diabetes or poor glucose metabolism, they most likely possess an intact glucose sensing mechanisms; thus, mutations that inactivate GCK function seem unlikely. This conclusion is supported by our recent characterization of *GCK* genes from diverse species, where intact *GCK* genes were found in the genomes of most vertebrates examined [32]. An exception was the genome sequences of two bat species (flying fox and little brown bat), where the *GCK* genes were found to be missing their liver-specific first exon, which might prevent expression of *GCK* in the liver but not other sites [32].

As indicated above, mice with a targeted disruption of the *GCKR* gene show a specific deficiency of GCK activity in their liver [24,25]. Low GCK activity is accompanied by an absence of GCKR in the liver of cats [33]. Since *GCKR* is the most important post-transcriptional regulator of GCK in most species and expressed almost exclusively in the liver of vertebrates that express GCK, including amphibians and reptiles [13], and not in the pancreatic beta-cells [16,17], the other major site of GCK function, these observations suggest that the loss of GCKR may explain a liver-specific GCK deficiency. *GCKR* genes, though, have only been identified or characterized in a few species [34–36]. A previous evolutionary analysis of *GCKR* sequences identified *GCKR*-like genes in mammals, amphibians, and fish, as well as

in non-vertebrate species, and concluded that *GCKR* evolved from an N-acetylmuramate 6-phosphate esterase by changing its binding specificity and losing its esterase activity [36]. To address the hypothesis that liver-specific GCK deficiency is caused by the loss of GCKR, we identified and characterized *GCKR* genes from the genomes of diverse vertebrate species. Our results show that species deficient in liver GCK activity have mutated or deleted *GCKR* genes, supporting the hypothesis that the loss of GCKR, and its ability to stabilize GCK, explains the loss of liver-specific GCK activity.

Materials and Methods

Genome sequence data

Genomic sequences encoding GCKR-like protein sequences were downloaded from release 69 of the Ensembl and PreEnsembl databases (www.ensembl.org and pre.ensembl.org) in August 2012. Genomes from all available vertebrate species maintained in these databases were searched by gene name, gene symbol (*GCKR*), or by similarity searches with the tblastn algorithm [37] using the human GCKR protein sequence [35]. Searches of the non-redundant and genome databases maintained at the National Center for Biotechnology Information (NCBI, www.ncbi.nlm.nih.gov) were used to complement the genomes gathered from the Ensembl database and aided in the identification of incomplete genes or coding region sequences. Sequences similar to *GCKR* were not found in several species. For these species, genomic sequences for genes in the predicted conserved genomic neighborhood were included in this analysis. Genomes were searched for orthologs of the genes that flank the *GCKR* genes in diverse species (see results for details) and the genomic sequences adjacent to these genes were searched for similarity to *GCKR* genes.

Alignment of sequences

Long genomic DNA sequences that included the *GCKR* gene were aligned with MultiPipMaker (pipmaker.bx.psu.edu/pipmaker/) [38,39]. Human and mouse *GCKR* genes were used as masters for these alignments, and the locations of exons and coding regions for these genes was obtained from the Ensembl or NCBI databases. The identity and locations of repetitive elements in the human and mouse genomic sequences were identified using RepeatMasker (www.repeatmasker.org). Genomic alignments were used to refine the predicted potential coding regions of the genes. Potential pseudogenes were identified as sequences that failed to predict open reading frames due to the presence of base changes that either introduced stop codons, created frame shifts that disrupted the coding sequence, or disrupted splicing consensus sequences. Predicted open reading frames and protein sequences were aligned with the ClustalW algorithm [40] as implemented in MEGA5 [41], with the protein sequences used as guides for the nucleotide sequence alignments.

Phylogenetic analysis

The origin of mutations that yielded the deletion of *GCKR* genes, or inactivation due to frame shift and/or splice site mutations were inferred by parsimony on the consensus vertebrate phylogenetic tree from Ensembl (www.ensembl.org). Phylogenetic trees were generated from the aligned GCKR protein sequences (see Figure S1 for alignment) using the neighbor joining method with Jones-Thornton-Thompson (JTT) or Dayhoff distances estimated using MEGA5 [41]. Phylogenetic trees of mammalian *GCKR* coding sequences were also generated from the DNA sequence data using either maximum likelihood composite

Table 1. Genomic location of vertebrate *GCKR* genes.

Species	Common name	Chromosome/Contig	Coding region	Ensembl Protein ID
<i>Homo sapiens</i>	Human	Chr 2	27,719,709-27,746,554	ENSP00000264717
<i>Pan troglodytes</i>	Chimpanzee	Chr 2A	27,882,887-27,909,760	ENSPTRP00000020223
<i>Gorilla gorilla</i>	Gorilla	Chr 2a	27,973,322-28,000,831	ENSGGOP00000012482
<i>Pongo abelii</i>	Orangutan	Chr 2a	84,605,694-84,632,388	ENSPPYP00000014009
<i>Nomascus leucogenys</i>	Gibbon	GL397331.1	6,798,770-6,825,699	ENSNLEP00000004669
<i>Papio hamadryas</i>	Baboon	Contig312330_Contig468503 ¹	252,057-274,688	ENSP00000264717_1
<i>Macaca mulatta</i>	Macaque	Chr 13	27,491,384-27,514,309	ENSETEP00000001256
<i>Saimiri boliviensis boliviensis</i>	Bolivian squirrel monkey	Scaffold JH378252.1 ¹	3,026,699-3,052,421	ENSP00000264717_1
<i>Callithrix jacchus</i>	Marmoset	Chr 14	80,920,115-80,945,184	ENSCJAP00000001541
		Ch 14	8,215,977-82,166,805	ENSCJAP000000051637
		GL285667.1	14,860-21,407	ENSCJAP000000046477
		GL289329.1	2,701-8,499	none
<i>Tarsius syrichta</i>	Tarsier	scaffold_142375	1-6,848	none
<i>Otolemur garnettii</i>	Bushbaby	GL873541.1	18,431,947-18,458,466	ENSOGAP00000008281
<i>Microcebus murinus</i>	Mouse Lemur	GeneScaffold_1280	170,174-180,857	ENSMICP000000014537
		scaffold_9408	44,808-49,358	
<i>Mus musculus</i>	Mouse	Chr 5	31,599,954-31,629,673	ENSMUSP000000072084
<i>Rattus norvegicus</i>	Rat	Chr 6 ¹	36,176,192-36,204,485	NCBI: NP_037252.1 ²
<i>Cricetulus griseus</i>	Chinese hamster	Scaffold JH000049 ¹	3,040,662-3,062,034	ENSMUST000000072228_1
<i>Dipodomys ordii</i>	Kangaroo rat	GeneScaffold_1944	64,619-93,530	ENDSORP000000004770
<i>Cavia porcellus</i>	Guinea pig	scaffold_18	9,315,790-9,341,258	ENSCPOP00000007381
<i>Ictidomys tridecemlineatus</i>	Squirrel	Scaffold JH393303	5,080,559-5,101,892	ENSSTOP000000004530
<i>Ochotona princeps</i>	Pika	GeneScaffold_1353	89,382-113,541	ENSOPRP000000011814
<i>Oryctolagus cuniculus</i>	Rabbit	Chr 2	158,578,950-158,598,892	ENSOCUP000000011833
<i>Bos taurus</i>	Cow	Chr 11	72,154,947-72,176,344	ENSBTAP000000041859
<i>Ovis aries</i>	Sheep	Chr 3 ¹	36,829,349-36,855,060	none
<i>Tursiops truncatus</i>	Dolphin	GeneScaffold_1519	243,720-266,976	ENSTTRP000000011887
<i>Sus scrofa</i>	Pig	Chr 3	118,796,212-118,823,097	ENSSSCP000000020241
<i>Vicugna pacos</i>	Alpaca	GeneScaffold_2392	230,396-239,542	ENSVPAP00000001840
<i>Equus caballus</i>	Horse	Chr 15	68,799,209-68,819,071	ENSECAP000000018966
<i>Canis familiaris</i>	Dog	Chr 17	21,434,251-21,459,913	ENSCAFP000000007696
<i>Felis scatus</i>	Cat	Chr A3 ¹	120,477,725-120,487,528	none
<i>Mustela putorius furo</i>	Ferret	Scaffold GL897010.1 ¹	4,462,918-4,658,018	ENSCAFP000000009165
<i>Ailuropoda melanoleuca</i>	Panda	GL192369.1	463,160-487,028	ENSAMEP00000001336
<i>Myotis lucifugus</i>	Little brown bat	GL430052	860,474-888,947	none
<i>Pteropus vampyrus</i>	Flying fox bat	GeneScaffold_1035	519,562-527,386	ENSPVAP000000014342
<i>Tupaia belangeri</i>	Tree shrew	GeneScaffold_1539	218,507-244,816	ENSTBEP000000003453
<i>Erinaceus europaeus</i>	Hedgehog	GeneScaffold_2521	113,709-138,867	ENSEEUP000000006890
<i>Sorex araneus</i>	Shrew	Not found		none
<i>Loxodonta africana</i>	Elephant	scaffold_20	41,160,463-41,182,623	ENSLAFP000000005668
<i>Procavia capensis</i>	Hyrax	GeneScaffold_2165	16,762-42,556	ENSPCAP000000015899
<i>Echinops telfairi</i>	Lesse hedgehog tenrec	GeneScaffold_9143	353,017-423,957	ENSETEP00000001256
<i>Dasypus novemcinctus</i>	Armadillo	GeneScaffold_3428	67,859-94,885	ENSDNOP000000014416
<i>Choloepus hoffmanni</i>	Sloth	GeneScaffold_3794	42,395-42,809	none
		scaffold_170360	376-835	
		scaffold_83977	28-3734	
<i>Monodelphi domestica</i>	Opossum	Chr 1	508,848,942-508,873,322	ENSMODP000000019482
<i>Macropus eugenii</i>	Wallaby	GeneScaffold_1196	275-7,845	ENSMEUP000000006501
<i>Sarcophilus harrisii</i>	Tasmanian devil	GL856719.1	1,420,689-1,458,490	ENSSHAP000000016711

Table 1. Cont.

Species	Common name	Chromosome/Contig	Coding region	Ensembl Protein ID
<i>Ornithorhynchus anatinus</i>	Platypus	Not found		none
<i>Anolis carolinensis</i>	Anole Lizard	Not found		none
<i>Chrysemys pictabellii</i>	Painted turtle	JH585650.1 ¹	8,656-43,647	none
<i>Pelodiscus sinensis</i>	Chinese softshell turtle	Scaffold JH210441.1	6,448-7,388	none
		Scaffold JH211948.1	132,429-114,901	
<i>Gallus gallus</i>	Chicken	Not found ¹		none
<i>Meleagris gallopavo</i>	Turkey	Not found		none
<i>Taeniopygia guttata</i>	Zebra finch	Not found		none
<i>Melopsittacus undulatus</i>	Budgerigar	Not found ¹		none
<i>Anas platyrhynchos</i>	Duck	Not found ¹		none
<i>Xenopus tropicalis</i>	Western clawed frog	GL176436.1	2,467-2,564	ENSXETP00000060990
		GL175638.1	1,885-7,075	
		GL181463.1	68-7,426	
		GL175969.1	6,442-11,214	
<i>Xenopus laevis</i>	African clawed frog			NCBI: NP_001083276.1 ²
<i>Latimeria chalumnae</i>	Coelacanth	Scaffold JH127635.1	202,083-223,756	ENSLACP00000006889
<i>Gadus morhua</i>	Cod	Not found		none
<i>Takifugu rubripes</i>	Takifugu	scaffold_64	713,339-722,055	ENSTRUP000000032438
<i>Oryzias latipes</i>	Medaka	Chr 2	24,524,153-24,535,721	ENSORLP00000005539
<i>Gasterosteus aculeatus</i>	Stickleback	group1	23,443,090-23,447,491	ENSGACP000000019720
<i>Tetraodon nniroviridis</i>	Tetraodon	Chr 3	4,336,678-4,340,938	ENSTNIP000000017284
<i>Danio rerio</i>	Zebrafish	Chr 17	5,624,859-5,639,741	ENSDARP0000000124274
<i>Oreochromis niloticus</i>	Tilapia	Scaffold GL831144.1	2,399,129-2,414,317	ENSONIP000000015155
<i>Xiphophorus maculatus</i>	Platyfish	JH556962.1 ¹	106,171-111,589	none
<i>Lepisosteus oculatus</i>	Spotted Gar	LG1 ¹	45,522,798-45,550,720	none
<i>Petromyzon marinus</i>	Lamprey	GL476557	253,805-279,336	ENSPMAP000000010620

¹Sequences identified/searched for in the PreEnsembl database (some represent newer genome assemblies than those in Ensembl)

²Protein ID from the NCBI database

doi:10.1371/journal.pone.0060896.t001

distances or divergence at synonymous and nonsynonymous sites corrected by the Kimura 2-parameter method using MEGA5 [41]. The reliability of the trees was assessed by the bootstrap method, with the tree of vertebrate sequences rooted with the lamprey sequence and those of placental mammals with the sequence from a marsupial, the Tasmanian devil. Comparisons of the rates of sequence evolution on different species lineages were conducted using relative rate tests [42] as implemented in MEGA5 [41]. For relative rate tests, the sequence from the Tasmanian devil was used as the outgroup, and the numbers of unique amino acid substitutions was counted for each lineage. The probability that numbers of unique amino acid substitutions were equal was tested by a chi square test [42].

Results

Identification of GCKR-like genomic sequences

The human *GCKR* gene is composed of 19 exons distributed over roughly 27 kb of genomic DNA [43]. Previously, cDNAs and predicted genes for *GCKR* have been identified or characterized in a few other species, such as the rat, dog, zebrafish, tetraodon, and *Xenopus laevis* [14,34,36], and evidence for GCKR protein activity was identified in a few additional species, such as chicken, trout, carp, and goldfish [44]. To better characterize *GCKR* genes in

vertebrates we searched the available genomes in the Ensembl (or preEnsembl) and NCBI databases for sequences that potentially encode GCKR. Our searches resulted in the identification of *GCKR*-like sequences in most species (Table 1) including all species for which a *GCKR* cDNA/gene had previously been identified. In the *Xenopus tropicalis* genome, a genome that is still not completely assembled, we identified four short genomic scaffolds containing sequences highly similar to the previously characterized *Xenopus laevis* *GCKR* cDNA and could be combined to predict a near complete *GCKR* gene sequence with only a single exon (exon 15) missing (Table 1). In a few other species (i.e., mouse lemur, sloth, and Chinese softshell turtle, Table 1) multiple genomic sequences were also identified that were consistent with a single gene because no part of the *GCKR* gene was found twice. An exception was the marmoset, where as many as four copies of the *GCKR* gene may exist (Table 1).

Missing GCKR-like genes

While genomic sequences similar to *GCKR* were found in most vertebrates, sequences similar to *GCKR* were not found in the genome sequences of two mammals (shrew and platypus), a reptile (anole lizard), all 5 available birds (chicken, turkey, duck, zebra finch, and budgerigar), and a fish (cod) (Table 1). Since genome assemblies for some species are still at a draft stage it might be

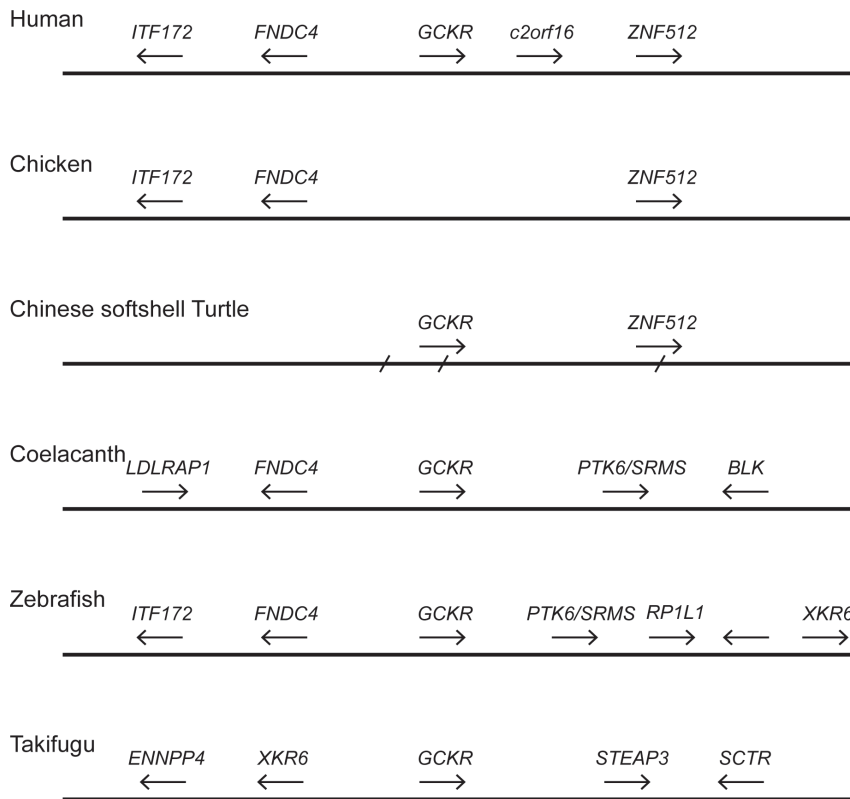


Figure 1. Genomic neighborhoods around vertebrate *GCKR* genes. Relative order and orientation of genes near the *GCKR* genes in human, chicken, Chinese softshell turtle, coelacanth, zebrafish and takifugu genomes. Gene names, as annotated in the human genome, are shown above the arrows, with the arrowhead indicating direction of transcription. Gene sizes and distance between genes are not to scale. Human is representative of the gene organization in mammals, while chicken is representative of birds. *GCKR* gene was not found in bird genomes. In fish, two distinct gene organizations were found – one found in zebrafish, and a second found in other fish genomes and represented by takifugu. The Chinese softshell turtle *GCKR* gene was distributed over two genomic contigs, with the slashes (/) indicating the ends of contigs. doi:10.1371/journal.pone.0060896.g001

expected that the failure to identify a *GCKR*-like sequence may simply be due to gaps in the available genomes. In contrast, our searches of these same genomes for *GCK* gene-like sequences resulted in the identification of complete or partial genes in all of these mammalian [32] and non-mammalian vertebrate species (Table S1), suggesting that the absence of sequences similar to *GCKR* in at least some (but possibly not all) of the genomes is due to the loss of this gene. Gene loss can be due to either deletion of a sequence from the genome, or failure to find using the blast similarity search algorithm as the sequence became non-functional and degenerated into a pseudogene.

Missing genes can sometimes be found if the gene order surrounding a gene is conserved, as the adjacent genes can then be used to search for orthologous genomic regions. This approach has successfully been used to identify genes from divergent genomes showing minimal or no similarity in blast searches, such as a homolog of the mammalian leptin gene in fish [45]. To determine whether this approach might be useful, we examined the gene neighborhoods flanking the *GCKR* genes to determine whether gene order is conserved in vertebrate genomes. The genes *FNDC4* (fibronectin type III domain containing 4) and *ITF172* (intraflagellar transport 172 homolog) are located 5' to the human *GCKR* gene, while *ZNF512* (zinc finger gene 512) gene and the predicted gene *C2orf16* (chromosome 2 open reading frame 16) are located 3' (Figure 1). Genomes of other mammals possessing the *GCKR* gene have the same gene order – for some species, the genomic fragment containing *GCKR* is shorter and do not predict

all four genes. While a *GCKR* gene could not be found in any of the bird genomes, it was found in two of the three reptile genomes, albeit they were on short genomic fragments (Table 1). Searches using the Painted turtle genomic sequence failed to identify any additional gene; however, a sequence downstream of the 3' end of the Chinese softshell turtle *GCKR* gene is similar to the *ZNF512* gene (Figure 1). The *ZNF512* gene is located 3' and in the same transcriptional orientation as *GCKR* in both mammals and a turtle (Figure 1), suggesting that this order existed in their common ancestor.

The *Xenopus tropicalis GCKR* gene is distributed over four small genomic fragments (Table 1) that do not show similarity to any other gene. The coelacanth is a lobe-finned fish that is more closely related to tetrapods (mammals, birds, reptiles and amphibians) than to other fish [46]. Genes similar to *FNDC4* and *LDLRAP1* (low density lipoprotein receptor adaptor protein 1) are found 5' to the coelacanth *GCKR* gene, while genes similar to *PTK6/SRMS* (protein tyrosine kinase 6/src-related kinase lacking C-terminal regulatory tyrosine and N-terminal myristylation sites) and *BLK* (B lymphoid tyrosine kinase) were found 3' (Figure 1). The gene neighborhood near the coelacanth *GCKR* gene differs considerably from that of the human *GCKR* gene, this suggests that recombination has occurred; however, the *FNDC4* is located in a similar position with the same transcriptional orientation, in both species, suggesting that the genomic linkage of these two genes was present in their common ancestor.

The gene neighborhood of the zebrafish *GCKR* gene strengthens the conclusion of an ancestral *GCKR* – *FNDC4* genomic linkage, as the zebrafish also has an *FNDC4*-like gene 5' to its *GCKR* gene (Figure 1). The zebrafish genome also predicts an *IFT172*-like gene in a position orthologous to the human *IFT172* gene, and a *PTK6/SRMS6*-like gene in a position orthologous to that found in the coelacanth genome (Figure 1). These observations suggest that the *FNDC4* and *IFT172* genes were 5' to the *GCKR* gene (as seen in the human and zebrafish genomes) and the *PTK6/SRMS6* gene was located 3' to *GCKR* (as seen in the zebrafish and coelacanth) in the common ancestor of fish and tetrapods. The gene order near the *GCKR* gene in the other fish genomes examined differed from that of zebrafish, and they shared an order that was consistent with that shown for the takifugu (Figure 1). All of the fish genomes have an *XKR6* (XK, Kell blood group complex subunit-related family, member 6) gene near the *GCKR* gene, albeit in different relative positions and in a different transcriptional orientation, suggesting that this genomic region has been reorganized within fish.

Based on these gene maps (Figure 1), we used the conserved flanking genes to identify orthologous genomic regions in species where Blast failed to identify a *GCKR*-like sequence. Since the ancestral tetrapod is predicted to have an *FNDC4*-like and a *ζNF512*-like gene flanking the *GCKR* gene, we used these sequences to search the mammalian genomes (shrew and platypus), reptilian (anole lizard) and avian (chicken, turkey, duck, zebra finch and budgerigar) genomes where we failed to identify a *GCKR* gene with our similarity searches (see Table 1). Searches of the mammalian (shrew and platypus) and the reptilian (anole lizard) genomes resulted in identifying only short genomic fragments encoding sequences similar to *FNDC4* and *ζNF512*, with none of these genomic fragments encoding any additional genes (results not shown). This result is consistent with the low coverage and fragmented nature of these genome sequences, and

thus we cannot determine whether the gene neighborhood is conserved, or if the *GCKR* gene was lost or is present in an unsequenced gap of these genome assemblies. In contrast, large genomic fragments were typically identified from the avian genomes, where the *FNDC4* and *ζNF512* genes were linked on the same genomic fragment, as shown for the chicken in Figure 1. Searches of the genomic sequences between the *FNDC4* and *ζNF512*-like genes from the bird genomes failed to identify any sequence with similarity to *GCKR*. These results strongly suggest that the *GCKR* gene was deleted from the avian genomes, and that this deletion event occurred before the divergence of the avian lineages represented by the genome sequences. The cod genome was also searched with genes that flank the *GCKR* gene in different fish genomes, and a genomic fragment with some similarity to that seen in coelacanth was found (containing *FNDC*, *LDLRAP1*, and *BLK*-like genes), but different from other ray-finned fish. This suggests that further reorganization has occurred (results not shown). Due to recombination, it is difficult to determine whether the cod *GCKR* gene was lost or is at a different location, such as an unsequenced gap, in this genome.

GCKR pseudogenes

Surprisingly, *GCKR* genes from many species, especially mammals, are poorly annotated (i.e., truncated or missing exons) or not annotated at all in the Ensembl database. To determine why some *GCKR* genes are poorly annotated we used MultiPipMaker [38,39] to align the genomic sequences of mammalian *GCKR* genes to the well-characterized human *GCKR* gene [43]. With the human exon sequences as guides, coding exons in the other mammalian *GCKR* genes could be identified. Similar results were obtained when the mouse *GCKR* gene was used as the master sequence for the MultiPipMaker analysis (results not shown). Alignments of the *GCKR* gene exons allowed us to identify changes that have occurred in the *GCKR* gene in each species. Many of the exons that had not been annotated as exons of the *GCKR* genes in species with poorly annotated *GCKR* genes in the Ensembl database were found to have mutations that either introduced frame shifts (due to insertion or deletion of sequences) into the coding sequence, or would be predicted to prevent mRNA splicing (Table 2). The effect of these mutations likely explains why gene prediction programs that annotated these genome assemblies failed to identify these sequences as coding exons (since splicing and coding potential are properties that should be retained by coding exons). The *GCKR* genes of many species were found to contain multiple disrupting mutations. In some cases, more than half of the 19 coding exons (e.g., cow with 11 exons) harbored mutations predicted to disrupt splicing or translation (Table 2). Only a few species (squirrel, tree shrew, and opossum) have a single disrupting mutation that potentially could be explained as a sequencing error or the sequence of a rare mutated allele present in the population of alleles for those species. A total of 14 mammalian species were identified to have *GCKR*-like genes in their genomes that likely do not encode a functional *GCKR* protein as they have two or more coding sequence disrupting mutations (Table 3).

The spotted gar was the only non-mammalian genome that contained a genomic sequence where all 19 *GCKR* coding exons could be identified and could be used to predict an intact *GCKR* protein (Table 1 and Figure S1). For all other non-mammalian species, the *GCKR*-like genomic sequence was missing at least one exon, most often exon 1. Unfortunately the missing exons cannot be identified because the non-mammalian (turtle, frog, and fish) genomic sequences do not align with the human, or any other mammalian, *GCKR* gene sequence used as a master sequence in

Table 2. Exons bearing inactivating mutation in mammalian *GCKR* genes.

Species	Exons with mutations	
	Frame shift mutations	Splice junction mutations
Tarsier	10, 15, 16	16
Squirrel	8	–
Cow	2, 5, 7, 8, 15, 17, 18, 19	2, 9, 10, 12, 15, 18
Sheep	5, 6, 8, 12, 18	9, 10, 12
Dolphin	1, 3, 4, 6, 8, 9, 15, 19	1
Alpaca	5, 7	1
Cat	2, 4, 10	2
Ferret	–	9, 10, 12
Little brown bat	2	3
Flying fox bat	5, 6, 8, 10, 12	5, 8, 10, 11
Hedgehog	1, 3, 4, 17	2
Tree shrew	4	–
Hyrax	1, 3, 5, 10, 11, 17	1, 3, 4, 5, 7, 11, 19
Lesser hedgehog tenrec	5, 7, 8, 9, 13, 17, 18	5, 6, 9, 13, 14, 15
Armadillo	8, 13, 14, 15, 19	13, 15, 16, 18, 19
Sloth	7, 13, 19	2, 4, 5, 6, 7, 18
Opossum	12	–

doi:10.1371/journal.pone.0060896.t002

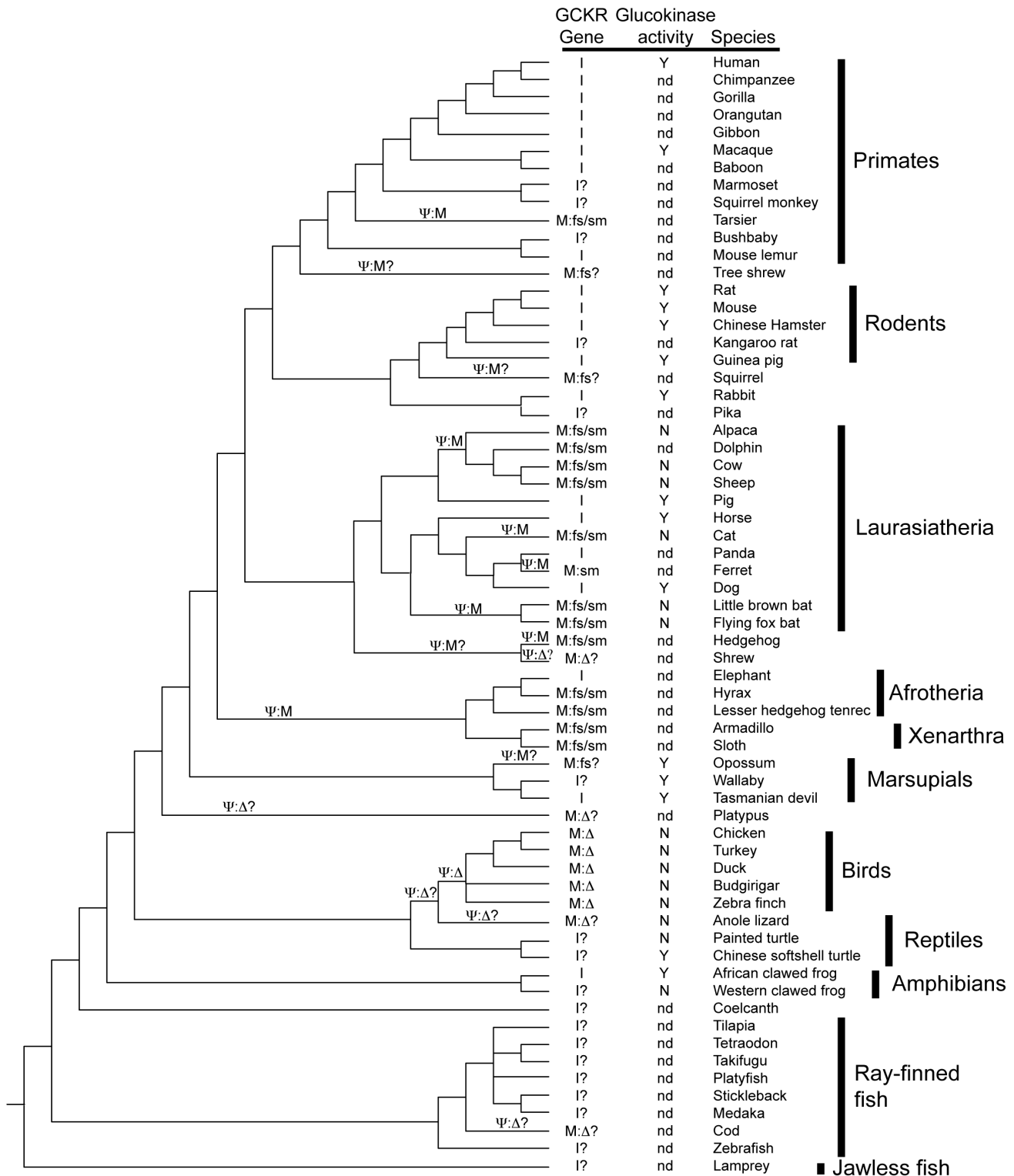


Figure 2. Evolution of GCKR genes and hepatic GCK activity in vertebrates. Summary of information on the structure of GCKR genes and hepatic GCK activity are placed on a phylogeny of vertebrates with available genome sequences (with the common names of the species shown on the right). Higher-level taxonomic groups of species are indicated to the right, with their composition indicated by the vertical bars. The phylogenetic relationship is from Ensembl (www.ensembl.org). *Xenopus laevis* is added to the tree as it has an intact GCKR cDNA. Genes are labeled as intact (I), likely intact (I?) or mutated (M), with the types of mutation indicated: fs = frame shift, sm = splice mutation, Δ = deletion, and Δ? = likely deletion (see Tables 2 and 3). The phylogenetic placement of gene inactivation events (Ψ, with type of inactivation indicated) was determined by parsimony. Possible inactivations, or events with unresolved locations (i.e., on the bird lineage), are indicated by the ? symbol. Hepatic GCK activity is from references 27-31, with Y = activity found, N = no or very low activity, and nd = not determined. doi:10.1371/journal.pone.0060896.g002

Table 3. Vertebrate *GCKR* genes.

<u>Intact coding regions</u>			
Human	Chimpanzee	Gorilla	Orangutan
Gibbon	Macaque	Baboon	Mouse lemur
Mouse	Rat	Chinese hamster	Guinea pig
Rabbit	Pig	Horse	Dog
Panda	Elephant	Tasmanian devil	African clawed frog
Spotted gar			
<u>Potentially intact genes</u>			
Squirrel monkey	Marmoset	Bushbaby	Kangaroo rat
Squirrel*	Pika	Tree shrew*	Opossum*
Wallaby	Painted turtle	Chinese softshell turtle	Western clawed frog
Coelacanth	Takifugu	Medaka	Stickleback
Tetraodon	Zebrafish	Tilapia	Platyfish
Lamprey			
<u>Genes with inactivating mutations</u>			
Tarsier	Cow	Sheep	Dolphin
Alpaca	Cat	Ferret	Little brown bat
Flying fox bat	Hedgehog	Hyrax	Lesser hedgehog tenrec
Sloth	Armadillo		
<u>Deleted genes</u>			
Shrew	Platypus	Anole lizard	Chicken
Turkey	Zebra finch	Budgerigar	Duck
Cod			

*Predicted genes have a single frame shift mutation that may be a sequencing error.

doi:10.1371/journal.pone.0060896.t003

MultiPipMaker (results not shown), due to the large amount of sequence divergence that has occurred since their common ancestor that existed at least 300 million years ago [47]. Similar results were obtained when the spotted gar sequence was used as master sequence, as it also is distantly related to all of the other species (including ray-finned fish, whom they diverged from more than 300 million years ago [48]). However, the coding exons that were identified for the non-mammalian *GCKR* genes (see Table 1) maintained open reading frames and possessed intact splice consensus sequences that should allow the generation of an intact *GCKR* proteins (see Figure S1). While it remains possible that inactivating mutations exist in some of the uncharacterized exons, our current genomic evidence suggests that functional *GCKR* proteins exist for all of the non-mammalian species where a *GCKR* gene was identified.

Parallel inactivation of mammalian *GCKR* genes

Our characterization of vertebrate *GCKR* genes indicates that functional versions of this gene has been lost by several mechanisms, including point mutations that disrupt splicing or cause translational frame shifts or deletion mutations which have removed the gene from the genome (Tables 2 and 3). Evidence for deletion of the gene is strongest in birds, while the failure to find a *GCKR* gene in some of the other species could be due to gaps in the current genome assemblies. All birds that have an assembled genome lack a *GCKR*-like gene sequence, suggesting that the gene was deleted before the common ancestor of birds (see Figure 2). The anole lizard is closely related to birds and lacks a *GCKR*-like gene (Table 1) suggesting that the deletion of the *GCKR* gene may have occurred in a reptilian ancestor of the anole lizard and birds

(see Figure 2). The *GCKR* gene was not deleted in the common ancestor of all reptiles as the gene was found in turtle genomes (Table 1). Within mammals, inactivating mutations appears to be the dominant mechanism for the loss of *GCKR* gene function – 14 species showed inactivated genes due to mutations, while only 2 species may have deleted the gene (Table 3). A single mutational event cannot explain the origin of the inactivated mammalian *GCKR* genes. First, no single mutation is shared among the inactivated *GCKR* genes, even among the mutationally inactivated genes (see Table 2). Secondly, species with inactivated *GCKR* genes are not monophyletic (Figure 2). Parsimony was used to infer the minimum number of gene inactivation events on the phylogeny of vertebrates. At least six gene inactivation events are required to explain the diversity of species that have multiple mutations (i.e., at least one frame shift and at least one splice site mutation) (Figure 2). Inactivating mutations must have occurred in the *GCKR* gene on the lineages leading to: (1) tarsier, (2) ruminant artiodactyls, (3) cat, (4) bats, (5) hedgehog, and (6) common ancestor of Afrotheria and Xenarthra (see Figure 2).

Are some intact *GCKR* genes pseudogenes?

Many of the inactivated *GCKR* genes are found in Laurasiatheria (mammalian orders: Artiodactyls, Perisodactyls, Carnivores, Chiroptera, Eulipotyphla, and Pholidota) where 4 parallel gene loss events are inferred to have occurred (Figure 2). An alternative explanation for this distribution of inactivated genes is that all of the Laurasiatherian *GCKR* genes are pseudogenes, which were generated by a single mutational event that did not disrupt the open reading frame but did disrupt *GCKR* protein function. Subsequent to the ancestral inactivating mutational event,

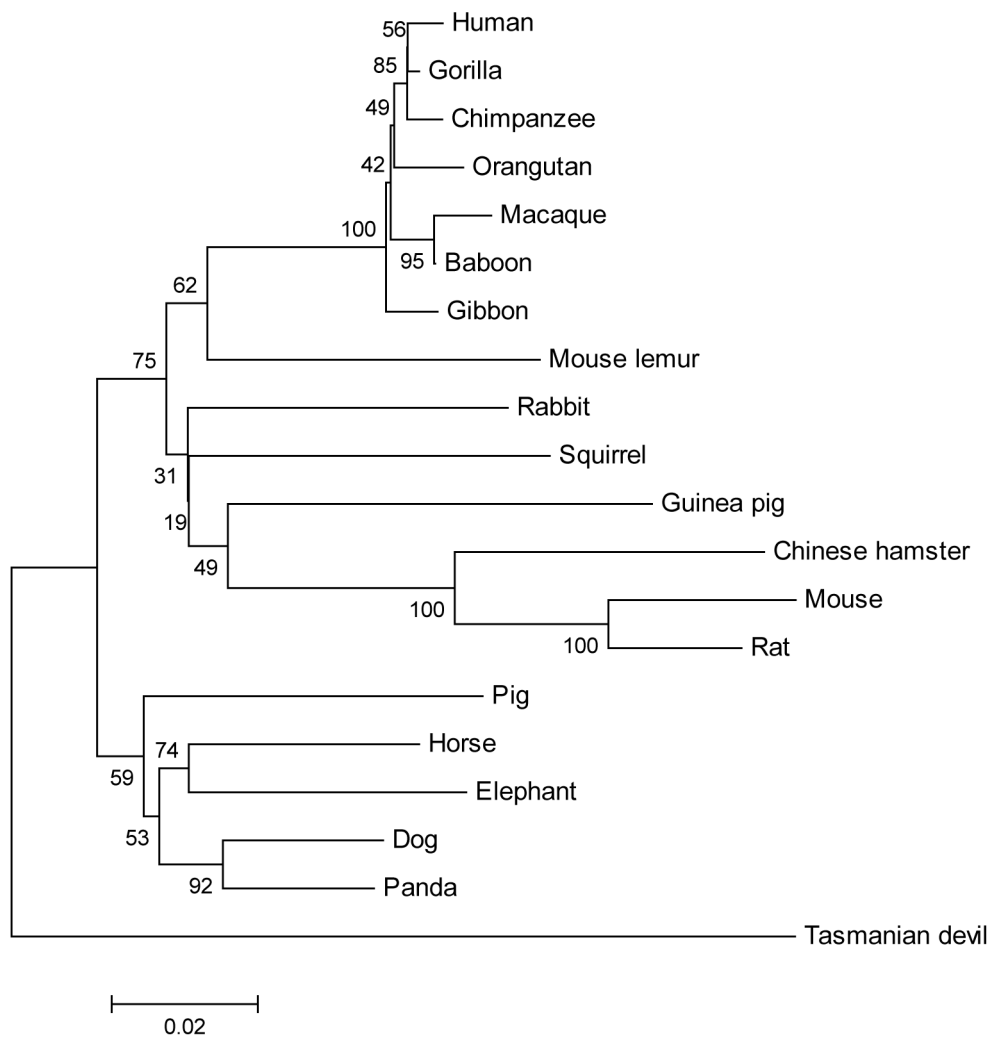


Figure 3. Phylogeny of mammalian GCKR protein sequences. Phylogeny of mammalian GCKR protein sequences generated from full length GCKR coding sequences (see Figure S1). The bootstrapped (1000 replications) neighbor-joining distance tree was generated using JTT protein distances. The sequence from the Tasmanian devil was used as the outgroup. Similar trees were generated when different protein distance measures, or distance measures based on nonsynonymous distances calculated from aligned DNA sequences, were used or if trees were built by other methods, such as parsimony or maximum likelihood.
doi:10.1371/journal.pone.0060896.g003

additional mutations introduced frame shifts and splice site mutation into the *GCKR* genes of most of the species representing Laurasiatheria (e.g., cow, cat, and bats). However, the gene sequences from dog, panda, horse, and pig did not acquire these types of mutations, and thus retained intact open reading frames. Pseudogenes should acquire greater numbers of mutations compared to functional genes. Thus if all of the *GCKR* genes in Laurasiatheria are pseudogenes, then the intact *GCKR* sequences from dog, panda, horse and pig should have accumulated more mutations than those of other mammals that have functional *GCKR* genes, such as rodents and humans [11,12,19–23]. To test this possibility we first constructed a phylogenetic tree (Figure 3) of the mammalian *GCKR* protein sequences from species that have full-length sequences, rooted with the sequence from the Tasmanian devil, a marsupial outgroup for placental mammals [47]. As shown in Figure 3, the dog, panda, horse and pig *GCKR* sequences may actually show less protein sequence evolution than other mammals, especially rodents, which is certainly not a pattern consistent with accelerated evolution that would be expected of

pseudogenes. Similar conclusions are drawn from phylogenies generated using more distantly related outgroups, including fish, amphibians and reptiles, where the dog and elephant *GCKR* proteins sequences do not accumulate greater numbers of amino acid substitutions compared to human, mouse or Tasmanian devil (Figure 4). To confirm that there was no acceleration in the rate of amino acid substitution in the *GCKR* sequences of species within Laurasiatheria compared to other placental mammals we conducted relative rate tests [42]. As shown in Table 4, when the dog, panda, horse, and pig sequences were tested against the human sequence, no significant difference in the number of amino acid substitutions was seen on either lineage. When the mouse sequence was used, a significant difference in the number of substitutions was seen, but it was the mouse sequence, not the potential pseudogene sequences, which had accumulated the greater number of amino acid substitutions (Table 4). If the rat sequence was used instead of the mouse, similar results were observed (results not shown). These results indicate that the intact *GCKR* coding sequence from species within Laurasiatheria are evolving as

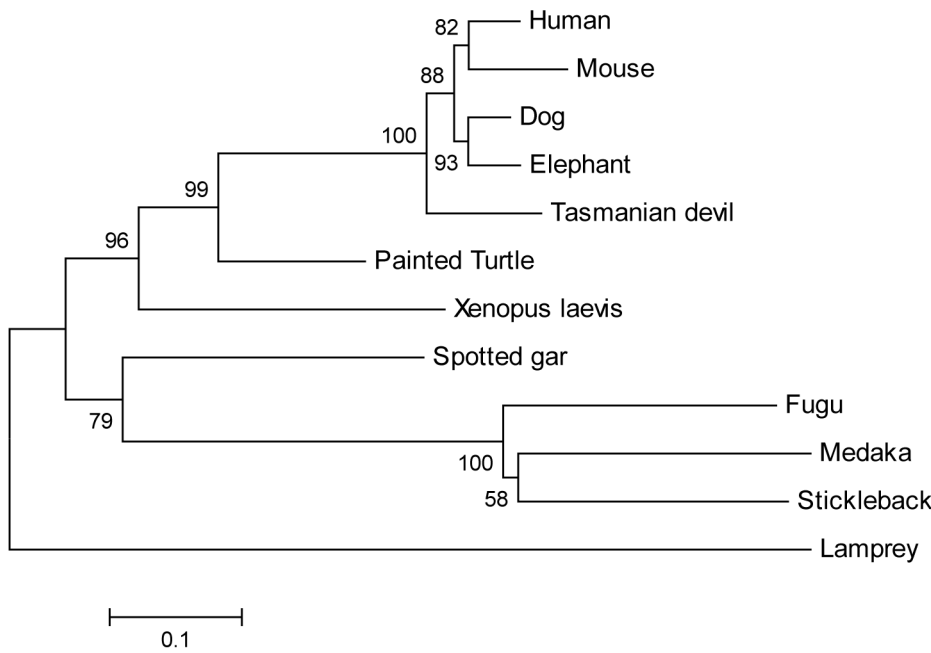


Figure 4. Phylogeny of vertebrate GCKR protein sequences. Phylogeny of vertebrate GCKR protein sequences generated from intact and near-full length *GCKR* coding sequences (see Figure S1). Only select mammals were included in the analysis. The bootstrapped (1000 replications) neighbor-joining distance tree was generated using JTT protein distances. The Lamprey (jawless fish) sequence was used as the outgroup. Similar trees were generated when different protein distance measures were used or if other tree building methods, such as parsimony or maximum likelihood, were used.
doi:10.1371/journal.pone.0060896.g004

slow, or slower, than those of other mammals thus it is extremely unlikely that they are non-functional.

Discussion

In humans, GCK has essential functions in both the liver and pancreatic beta cells, and defects at either site contribute to diabetes [3,49]. Knockout mice have been used to define the tissue-specific roles of *GCK* [50]. Homozygous pancreatic beta cell-specific deletion of the *GCK* leads to death due to severe diabetes, while homozygous mice survive but are moderately hyperglycemia [50]. Liver-specific deletion of *GCK* leads to hyperglycemia due to defects in glucose metabolism and impacts on insulin secretion in response to glucose [50]. In the liver, GCK acts as a gatekeeper for glucose utilization, as phosphorylation of glucose by GCK drives the storage of glucose as glycogen [49]. Despite the importance of GCK function in the liver of humans, a number of vertebrate species have been reported to be deficient in GCK activity, such as cat, bat, ruminants, and birds [27,28]. Deficiency of hepatic GCK

activity is unlikely to be caused by mutations that inactivate the entire *GCK* gene as insulin secretion is not lost in these species and they do not exhibit the symptoms of diabetes that are seen when GCK is depleted from pancreatic islets of mice [50]. Liver and beta-cell expression of *GCK* is driven by two different promoters [3,6], thus, mutations potentially could specifically inactivate the liver-specific isoform of GCK if they occur in the liver-specific first exon or prevent function of the liver-specific promoter. Indeed, this mechanism may explain GCK deficiency in bats, as the liver-specific 1st exon of the *GCK* gene in bats appears to be deleted [32]. All other mammals examined, though, possess a liver-specific 1st exon [32] and the sequences of this exon and liver-specific promoter do not display characteristics of being non-functional. GCKR regulates the function of GCK in the liver [15–17], and knockout of the mouse *GCKR* gene results in deficiency of liver GCK protein and activity [24,25], while overexpression of GCKR in hepatocytes leads to increase in both GCK protein and enzymatic activity levels [26]. These results indicate that in liver,

Table 4. Relative rate tests.

Species A	Human				Mouse			
	Dog	Panda	Horse	Pig	Dog	Panda	Horse	Pig
Differences in Sequence A ¹	22	23	21	18	48	52	47	44
Differences in Sequence B ¹	15	12	19	25	10	11	15	21
Chi square ²	1.32	3.46	0.1	1.14	24.9	26.68	16.52	8.14
P	0.25	0.06	0.75	0.29	<0.01	<0.01	<0.01	<0.01

¹– Number of unique amino acid substitutions on the lineages to species A and B when the Tasmanian devil sequence was used as the outgroup.

²– Chi square value for the expectation that an equal number of substitutions occurred on each lineage.

doi:10.1371/journal.pone.0060896.t004

GCK requires GCKR for stability, and the absence of GCKR leads to GCK degradation. *GCKR* does not appear to be essential for GCK function in pancreatic beta-cells, as *GCKR* knockout mice did not show impaired insulin secretion [24,25] and the mRNA expression ratio of *GCK:GCKR* is much higher in human pancreatic islets when compared with human liver [17]. As GCKR is the most important post-transcriptional regulator of GCK levels in the liver [11,12], and vertebrate species that are deficient in hepatic GCK appear to possess a functional *GCK* gene, we raised the idea that a lack of GCKR impacts liver-specific GCK levels and activity. To test this hypothesis, we have identified and characterized *GCKR* genes from the genomes of diverse vertebrate species. We found that species that lack GCK activity had deleted or mutated *GCKR* genes, while species that have reported liver GCK activity have intact *GCKR* genes. Our results are in accord with results seen in *GCKR* knockout mice [24,25] and with over-expression of *GCKR* [26], but further study is still needed to understand the mechanisms by which GCKR specifically regulates GCK protein levels in liver cells.

Loss of GCK activity from the liver of humans contributes to mild diabetes [4]; however, many species with limited hepatic GCK activity do not have diabetes [27–31]. Changes in the requirement for hepatic GCK activity may be associated with changes in diet. Ruminant animals, such as cow and sheep, acquire most of their energy from volatile fatty acids generated by fermentation in their foregut; thus, limited amounts of glucose are acquired from their diets [51]. Glucose that is needed by other tissues is produced in the liver, and its production is regulated by insulin and glucagon [51]. Therefore, ruminants do not need GCK to remove excess glucose from the circulation. Cats, like other carnivores, consume a diet containing large amounts of protein and relatively little carbohydrate [52]. As with the ruminants, glucose in carnivores is generated in the liver by a regulated process from other food sources; thus GCK may not be needed to remove excess glucose. Birds, species that have low hepatic GCK activity levels, have the highest blood glucose levels among vertebrates. With an average of 15.6 mM, or two fold higher than mammals [53], the relatively high glucose levels in

birds suggests that the loss of GCKR has prevented hepatic GCK from efficiently removing excess blood glucose and may have driven additional unknown changes in glucose metabolism.

Here we have shown that the loss of hepatic GCK activity in many species is likely due to mutations in the *GCKR* gene, rather than mutations in *GCK*. Mutation of the liver-specific *GCKR* gene may have allowed these species to specifically lose GCK activity from the liver without affecting GCK activity in other glucose-sensing tissues such as pancreatic beta-cells [2,3]. How can hepatic tissue function with the loss of GCK activity? Glucose is essential as an energy molecule by many cells in the body such as the neurons [1]; however, import of glucose is not essential for liver cells. GCK phosphorylates glucose as it enters hepatic cells, but this function is only essential if blood glucose levels vary due to diet. If glucose is not directly obtained from the diet, but rather synthesized in the liver, then blood glucose levels should be regulated by the production, rather than uptake, of glucose. As diets change, the requirements for hepatic GCK activity change, and this appears to have been achieved by the loss of *GCKR* gene function by multiple mechanisms on multiple lineages within vertebrates.

Supporting Information

Figure S1 Alignment of GCKR protein sequences. Full-length and near full-length GCKR protein sequence predicted from *GCKR* genes listed in Table 1 aligned with ClustalW [40]. (DOCX)

Table S1 Genomic locations of non-mammalian *GCK* genes. (DOC)

Author Contributions

Conceived and designed the experiments: ZYW LJ HT DMI. Performed the experiments: ZYW DMI. Analyzed the data: ZYW LJ HT DMI. Contributed reagents/materials/analysis tools: HT DMI. Wrote the paper: ZYW LJ HT DMI.

References

- Lin HV, Accili D (2011) Hormonal regulation of hepatic glucose production in health and disease. *Cell Metab* 14: 9-19.
- Wilson JE (2003) Isozymes of mammalian hexokinase: structure, subcellular localization and metabolic function. *J Exp Biol* 206: 2049-2057.
- Iynedjian PB (2009) Molecular physiology of mammalian glucokinase. *Cell Mol Life Sci* 66: 27-42.
- Malecki MT (2005) Genetics of type 2 diabetes mellitus. *Diabetes Res Clin Pract* 68: S10-21.
- Hussain K (2010) Mutations in pancreatic β -cell Glucokinase as a cause of hyperinsulinaemic hypoglycaemia and neonatal diabetes mellitus. *Rev Endocr Metab Disord* 11: 179-183.
- Iynedjian PB (1993) Mammalian glucokinase and its gene. *Biochem J* 293: 1-13.
- Iynedjian PB, Gjinovci A, Renold AE (1988) Stimulation by insulin of glucokinase gene transcription in liver of diabetic rats. *J Biol Chem* 263: 740-744.
- Iynedjian PB, Jotterand D, Nospikel T, Asfari M, Pilot PR (1989) Transcriptional induction of glucokinase gene by insulin in cultured liver cells and its repression by the glucagon-cAMP system. *J Biol Chem* 264: 21824-21829.
- Arden C, Petrie JL, Tudhope SJ, Al-Oanzi Z, Claydon AJ, et al. (2011) Elevated glucose represses liver glucokinase and induces its regulatory protein to safeguard hepatic phosphate homeostasis. *Diabetes* 60: 3110-3120.
- da Silva Xavier G, Varadi A, Ainscow EK, Rutter GA (2000) Regulation of gene expression by glucose in pancreatic beta-cells (MIN6) via insulin secretion and activation of phosphatidylinositol 3'-kinase. *J Biol Chem* 275: 36269-36277.
- van Schaftingen E, Dethoux M, Veiga da Cunha M (1994) Short-term control of glucokinase activity: role of a regulatory protein. *Faseb J* 8: 414-419.
- van Schaftingen E, Veiga da Cunha M, Niculescu L (1997) The regulatory protein of glucokinase. *Biochem Soc Trans* 15: 136-140.
- Vandercammen A, Van Schaftingen E (1993) Species and tissue distribution of the regulatory protein of glucokinase. *Biochem J* 294: 551-556.
- Dethoux M, Vandekerckhove J, Van Schaftingen E (1993) Cloning and sequencing of rat liver cDNAs encoding the regulatory protein of glucokinase. *FEBS Lett* 321: 111-115.
- Roncero I, Sanz C, Alvarez E, Vázquez P, Barrio PA, Blázquez E (2009) Glucokinase and glucokinase regulatory proteins are functionally coexpressed before birth in the rat brain. *J Neuroendocrinol* 21: 973-981.
- Tiedge M, Steffek H, Elsner M, Lenzen S (1999) Metabolic regulation, activity state, and intracellular binding of glucokinase in insulin-secreting cells. *Diabetes* 48: 514-523.
- Beer NL, Tribble ND, McCulloch IJ, Roos C, Johnson PR, et al. (2009) The P446L variant in GCKR associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. *Hum Mol Genet* 18: 4081-4088.
- Agius L (1994) Control of glucokinase translocation in rat hepatocytes by sorbitol and the cytosolic redox state. *Biochem J* 298: 237-243.
- Brown KS, Kalinowski SS, Megill JR, Durham SK, Mookhtiar KA (1997) Glucokinase regulatory protein may interact with glucokinase in the hepatocyte nucleus. *Diabetes* 46: 179-186.
- Toyoda Y, Miwa I, Satake S, Anai M, Oka Y (1995) Nuclear location of the regulatory protein of glucokinase in rat liver and translocation of the regulator to the cytoplasm in response to high glucose. *Biochem Biophys Res Commun* 215: 467-473.
- de la Iglesia N, Veiga-da-Cunha M, Van Schaftingen E, Guinovart JJ, Ferrer JC (1999) Glucokinase regulatory protein is essential for the proper subcellular localisation of liver glucokinase. *FEBS Lett* 456: 332-338.
- Shiota C, Coffey J, Grimsby J, Grippo JF, Magnuson MA (1999) Nuclear import of hepatic glucokinase depends upon glucokinase regulatory protein, whereas export is due to a nuclear export signal sequence in glucokinase. *J Biol Chem* 274: 37125-37130.

23. Veiga-da-Cunha M, Van Schaftingen E (2002) Identification of fructose 6-phosphate- and fructose 1-phosphate-binding residues in the regulatory protein of glucokinase. *J Biol Chem* 277: 8466-8473.
24. Farrelly D, Brown KS, Tieman A, Ren J, Lira SA, et al. (1999) Mice mutant for glucokinase regulatory protein exhibit decreased liver glucokinase: a sequestration mechanism in metabolic regulation. *Proc Natl Acad Sci USA* 96: 14511-14516.
25. Grimsby J, Coffey JW, Dvorozniak MT, Magram J, Li G, et al. (2000) Characterization of glucokinase regulatory protein-deficient mice. *J Biol Chem* 275: 7826-7831.
26. Slosberg ED, Desai UJ, Fanelli B, St Denny I, Connelly S, et al. (2001) Treatment of type 2 diabetes by adenoviral-mediated overexpression of the glucokinase regulatory protein. *Diabetes* 50: 1813-1820.
27. Ureta T (1982) The comparative isozymology of vertebrate hexokinases. *Comp Biochem Physiol* 71B: 549-555.
28. Ureta T, Reichberg SB, Radojković J, Slebe JC (1973) Comparative studies on glucose phosphorylating isoenzymes of vertebrates. IV. Chromatographic profiles of hexokinases from the liver of several avian species. *Comp Biochem Physiol* 45B: 445-461.
29. Ureta T, Slebe JC, Radojković J, Lozano C (1975) Comparative studies on glucose phosphorylating isoenzymes of vertebrates. V. Glucose phosphotransferases in the liver of reptiles. *Comp Biochem Physiol* 50B: 515-524.
30. Ureta T, Radojković J, Zepeda S, Guixé V (1981) Comparative studies on glucose phosphorylating isoenzymes of vertebrates – VII. Mammalian heokinases. *Comp Biochem Physiol* 70B: 225-236.
31. Ureta T, Radojković J, Diaz N, Slebe JC, Lozano C (1978) Comparative studies on glucose phosphorylating isoenzymes of vertebrates. Identification and characterization of amphibian liver hexokinases. *Arch Biochem Biophys* 186: 235-247.
32. Wang Y, Guo T, Zhao S, Li Z, Mao Y, et al. (2012) Expression of the human glucokinase gene: important roles of the 5' flanking and intron 1 sequences. *PLoS One* 7: e45824.
33. Hiskett EK, Suwitheechon OU, Lindbloom-Hawley S, Boyle DL, Schermerhorn T (2009) Lack of glucokinase regulatory protein expression may contribute to low glucokinase activity in feline liver. *Vet Res Commun* 33: 227-240.
34. Veiga-Da-Cunha M, Detheux M, Watelet N, Van Schaftingen E (1994) Cloning and expression of a *Xenopus* liver cDNA encoding a fructose-phosphate-insensitive regulatory protein of glucokinase. *Eur J Biochem* 225: 43-51.
35. Warner J P, Leek JP, Intody S, Markham AF, Bonthron DT (1995) Human glucokinase regulatory protein (GCKR): cDNA and genomic cloning, complete primary structure, and chromosomal localization. *Mamm Genome* 6: 532-536.
36. Veiga-da-Cunha M, Sokolova T, Opperdoes F, Van Schaftingen E (2009) Evolution of vertebrate glucokinase regulatory protein from a bacterial N-acetylmuramate 6-phosphate etherase. *Biochem J* 423: 323-332.
37. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389-3402.
38. Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, et al. (2000) PipMaker--a web server for aligning two genomic DNA sequences. *Genome Res* 10: 577-586.
39. Schwartz S, Elmitski L, Li M, Weirauch M, Riemer C, et al. (2003) MultiPipMaker and supporting tools: Alignments and analysis of multiple genomic DNA sequences. *Nucleic Acids Res* 31: 3518-3524.
40. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673-4680.
41. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731-2739.
42. Tajima F (1993) Simple methods for testing molecular clock hypothesis. *Genetics* 135: 599-607.
43. Hayward BE, Dunlop N, Intody S, Leek JP, Markham AF, et al. (1998) Organization of the human glucokinase regulator gene GCKR. *Genomics* 49: 137-142.
44. Polakof S, Miguez JM, Soengas JL (2009) A hepatic protein modulates glucokinase activity in fish and avian liver: a comparative study. *J Comp Physiol* 179B: 643-652.
45. Kurokawa T, Uji S, Suzuki T (2005) Identification of cDNA coding for a homologue to mammalian leptin from pufferfish, *Takifugu rubripes*. *Peptides* 26: 745-750.
46. Shan Y, Gras R (2011) 43 genes support the lungfish-coelacanth grouping related to the closest living relative of tetrapods with the Bayesian method under the coalescence model. *BMC Res Notes* 4: 49.
47. Kumar S, Hedges SB (1998) A molecular timescale for vertebrate evolution. *Nature* 392: 917-920.
48. Hurley IA, Mueller RL, Dunn KA, Schmidt EJ, Friedman M, et al. (2007) A new time-scale for ray-finned fish evolution. *Proc Biol Sci* 22: 489-498.
49. Osbak KK, Colclough K, Saint-Martin C, Beer NL, Bellanné-Chantelot C, et al. (2009) Update on mutations in glucokinase (GCK), which cause maturity-onset diabetes of the young, permanent neonatal diabetes, and hyperinsulinemic hypoglycemia. *Hum Mutat* 30: 1512-1526.
50. Postic C, Shiota M, Niswender KD, Jetton TL, Chen Y, et al. (1999) Dual roles for glucokinase in glucose homeostasis as determined by liver and pancreatic beta cell-specific gene knock-outs using Cre recombinase. *J Biol Chem* 274: 305-315.
51. Aschenbach JR, Kristensen NB, Donkin SS, Hammon HM, Penner GB (2010) Gluconeogenesis in dairy cows: the secret of making sweet milk from sour dough. *IUBMB Life* 62: 869-877.
52. Verbrugge A, Hesta M, Daminet S, Janssens GP (2012) Nutritional modulation of insulin resistance in the true carnivorous cat: a review. *Crit Rev Food Sci Nutr* 52: 172-182.
53. Polakof S, Mommsen TP, Soengas JL (2011) Glucosensing and glucose homeostasis: from fish to mammals. *Comp Biochem Physiol* 160B: 123-149.