Molecular evolution of the ATP-binding cassette subfamily G member 2 gene subfamily and its paralogs in birds

Shengchao Ma[1], Hehe Liu[1*], Wenqiang Sun[1], Ahsan Mustafa[2], Yang Xi[1], Fajun Pu[1], Yanying Li[1], Chunchun Han[1], Lili Bai[1] and He Hua[1]

Abstract

Background: ATP-binding cassette (ABC) transporters are involved in the active transportation of various endogenous or exogenous substances. Two *ABCG2* gene subfamily members have been identified in birds. A detailed comparative study of the *ABCG2* and *ABCG2-like* genes aid our understanding of their evolutionary history at the molecular level and provide a theoretical reference for studying the specific functions of *ABCG2* and *ABCG2-like* genes in birds.

Results: We first identified 77 *ABCG2/ABCG2-like* gene sequences in the genomes of 41 birds. Further analysis showed that both the nucleic acid and amino acid sequences of *ABCG2* and *ABCG2-like* genes were highly conserved and exhibited high homology in birds. However, significant differences in the N-terminal structure were found between the ABCG2 and ABCG2-like amino acid sequences. A selective pressure analysis showed that the *ABCG2* and *ABCG2-like* genes were affected by purifying selection during the process of bird evolution.

Conclusions: We believe that multiple members of the *ABCG2* gene subfamily exist on chromosome 4 in the ancestors of birds. Over the long course of evolution, only the *ABCG2* gene was retained on chromosome 4 in birds. The *ABCG2-like* gene on chromosome 6 might have originated from chromosome replication or fusion. The structural differences between the N terminus of ABCG2 protein and those of ABCG2-like proteins might lead to functional differences between the corresponding genes.

Keywords: Birds, *ABCG2*, *ABCG2-like*, Gene duplication, Gene family evolution, Phylogeny, Selection pressures, Chromosomal synteny

Background

The ATP-binding cassette (ABC) subfamily G member 2 (junior blood group) (*ABCG2*) gene is the second member of the G subfamily of ABC transporters and is also considered the breast cancer resistance protein (*BCRP*) gene [1, 2]. The ABCG2/BCRP protein is mainly distributed in tissues with secretory and excretory functions,

such as placental/synovial trophoblasts, small and large intestinal epithelia, liver tubule membrane, canaliculi, mammary lobule and vascular endothelial cells [3, 4]. The human ABCG2 protein contains a nucleotide-binding domain (NBD) and six transmembrane domains (TMDs) [5]. The ABCG2 protein can transport substrates from intracellular fluids to extracellular interstitial fluids and is reportedly involved in various other functions [6, 7], such as the stability of stem cells [8], the steady state of tissues cells [9–11], maintaining the blood-brain barrier and the fetal blood barrier [1,

* Correspondence: liuee1985@sicau.edu.cn
[1]Farm Animal Genetic Resources Exploration and Innovation Key Laboratory of Sichuan Province, Sichuan Agricultural University, Chengdu, Sichuan 611130, P.R. China
Full list of author information is available at the end of the article

Ma *et al. BMC Evolutionary Biology*   (2020) 20:85

Page 2 of 14

12] and reducing drug absorption, distribution and excretion [13].

The members of the ABC transporter superfamily in most mammals can be classified into seven subfamilies (from A to G) [14], and each of these subfamilies might have undergone a long evolutionary process, from single structures to half structures or ABC2 structures, and then from half structures to full structures (simple to complex structures). Xiong et al. [15] found that NBD and TMD domain fusion events might have occurred during the above process and that these fusion events occurred at least four times during the transformation from the half-structured transporters to full structures. Some ABC proteins have lost their TMD, which leads to changes in their basic functions (e.g., ABCE and ABCF). During the evolution of the seven full-structure ABC transporters, ABCA, ABCB, ABCC and ABCG originated before the last eukaryotic common ancestor (LECA), whereas the ABCD, ABCE and ABCF families originated before terrestrial plants, archaea, and the differentiation between bacteria and archaea, respectively.

A large number of gene family duplications have occurred via whole-genome duplication (WGD) events [16–18]. Seret et al. [19] found that members of the ATP-binding cassette superfamily, namely, *Pdr5p* and *Snq2p*, derived from a common ancestor gene before WGD. In contrast, both *Pdr10p* (*Pdr5p* paralog) and *YNR070wp* (*Snq2p* paralog) originated from independent duplicating events after WGD. Duplication events of ABC transporter genes have occurred in both fish and mammalian genomes, but ABC transporter gene loss events have also occurred due to duplicating events in a large number of genes. Annilo et al. [20] found that both gene transformation and coevolution occurred during the introduction and loss of ABC transporter genes. Moreover, human ABC transporter genes show 94, 85 and 77% homology to those in mammals, chickens and zebrafish, respectively. However, only 41 ABC transporter genes can be found in chickens, which is fewest number of those found in any higher vertebrate, and no specific genome-duplication events have been detected in birds.

The *ABCG2* gene is located in a quantitative trait locus (QTL) in some livestock species, and mutations in *ABCG2* are associated with performance and disease traits in livestock and humans [21]. For example, *ABCG2* variants are likely to affect the milk yield and composition in Holstein cattle [21–23] and the development of gout [24] and drug resistance in human cancer cells [1, 25–28] (e.g., breast [1, 27], colon [25] and liver [28] cancer cells). However, the evolutionary process and functions of the *ABCG2-like* gene (*ABCG2* paralog and a member of the *ABCG2* gene subfamily in birds) remain unknown.

Knowledge of the molecular evolution of gene families is an important prerequisite for understanding the functional differences among protein family members and predicting new functions for their paralogous and orthologous genes. Therefore, this study aimed to analyze the genetic structure, genome duplication characteristics, chromosome distribution, phylogeny and other aspects of the bird *ABCG2* and *ABCG2-like* genes and to therefore determine the potential changes in and connections between these genes and their functions in birds. Overall, the present study can provide theoretical references for studying not only the trait regulatory functions of *ABCG2* and *ABCG2-like* genes but also the evolution aspects of the ABC transporter superfamily in birds.
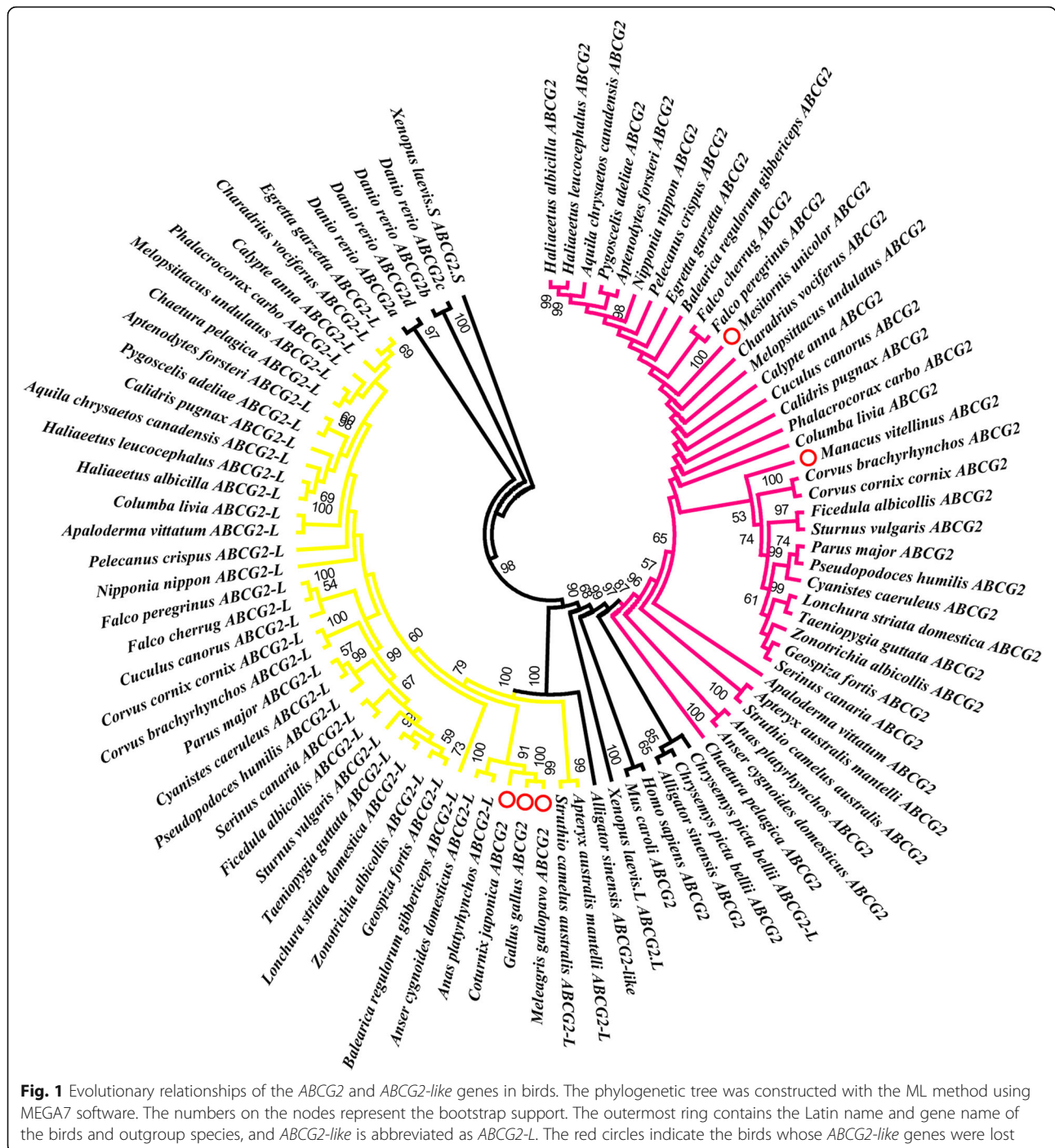
Results

ABCG2 gene subfamilies in different birds

Through BLAST, 77 *ABCG2* and *ABCG2-like* nucleic acid sequences, including 41 of the *ABCG2* gene and 36 of the *ABCG2-like* gene, were obtained from the genomes of 41 bird belonging to 33 families. Combined with the current NCBI nomenclature system, the BLAST results suggested that only two *ABCG2* gene subfamily members, i.e., the *ABCG2* and *ABCG2-like* genes, exist in birds. However, the *ABCG2-like* genes were lost in *Coturnix japonica* (Phasianidae), *Meleagris gallopavo* (Phasianidae), *Gallus gallus* (Phasianidae), *Mesitornis unicolor* (Mesitornithidae) and *Manacus vitellinus* (Pipridae) (Additional file 1, Tables S1 and S2).

Phylogenetic analysis

A maximum likelihood (ML) phylogenetic analysis (Fig. 1) was performed using the nucleic acid sequences (coding sequences) of the *ABCG2* and *ABCG2-like* genes, which were obtained from all birds (Additional file 1, Table S1) and outgroup species (Additional file 2, Table S3). The results supported the classification of the two gene subfamily members (*ABCG2* and *ABCG2-like*) in the 41 birds. However, phylogenetic trees also clustered the ABCG2 genes of 3 birds (*Coturnix japonica*, *Meleagris gallopavo* and *Gallus gallus*) were clustered in the clades of *ABCG2-like* genes (species marked with red circles in Fig. 1), suggesting that these three *ABCG2* genes and *ABCG2-like* genes have high homology. Although the clade nodes between the two subfamilies and between different species did not always exhibit sufficient phylogenetic resolution (the bootstrap support for some nodes was less than 50%), we could still deduce that the *ABCG2* and *ABCG2-like* genes of birds likely originated from a common ancestor. We subsequently consulted the comprehensive bird phylogeny described by Prum et al. [29] and found that the phylogenetic relationships among *ABCG2-like* genes were closer to the comprehensive phylogenetic relationship among birds

Fig. 1 Evolutionary relationships of the *ABCG2* and *ABCG2-like* genes in birds. The phylogenetic tree was constructed with the ML method using MEGA7 software. The numbers on the nodes represent the bootstrap support. The outermost ring contains the Latin name and gene name of the birds and outgroup species, and *ABCG2-like* is abbreviated as *ABCG2-L*. The red circles indicate the birds whose *ABCG2-like* genes were lost

than those among *ABCG2* genes. Therefore, we speculated that the *ABCG2-like* genes were more conserved than the *ABCG2* genes during the evolution of birds.

Selection pressure analysis

Using the CodeML program, positive and purifying selection analyses of the *ABCG2* and *ABCG2-like* gene sequences were performed for the various birds, as shown in Tables 1 and 2. In the M0 (one ratio) model, the ω

values of *ABCG2* and *ABCG2-like* genes were 0.25797 and 0.23234, respectively (Tables 1 and 2), and these values were far less than 1. Therefore, the M0 model provides no direct evidence that the *ABCG2* and *ABCG2-like* genes were affected by positive selection pressure. Positive selection analysis was used to identify positive selection sites (Fig. 2) [30].

A likelihood ratio test (LRT) was performed to compare the M1a (nearly neutral) and M2a (positive

Table 1 Selection pressure analysis of sites in the coding region of *ABCG2* genes

Model code	lnL	Parameters	Number of positive selection sites
M0 (one ratio)	−19,249.376879	$\omega = 0.25797$	None
M1a (nearly neutral)	−18,586.64149	$P0 = 0.70173, P1 = 0.29827$ $\omega0 = 0.07200, \omega1 = 1$	Not allowed
M2a (positive selection)	−18,561.41171	$P0 = 0.69472, P1 = 0.27126, P2 = 0.03402$ $\omega0 = 0.07356, \omega1 = 1, \omega2 = 2.61739$	13
M3 (discrete)	−11,953.59808	$P0 = 0.28795, P1 = 0.18745, P2 = 0.27619$ $\omega0 = 0.03433, \omega1 = 0.03433, \omega2 = 0.03433$	Not allowed
M7 (beta)	−18,532.88959	$p = 0.24865, q = 0.61421 (p1 = 0.1),$ $\omega = 0.00101$	Not allowed
M8 (beta&$\omega > 1$)	−18,489.18864	$p0 = 0.9112692, p = 0.33179, q = 1.27694, (p1 = 0.08874)$ $\omega = 1.77811$	56

Note: lnL refers to the logarithm of the maximum likelihood (ML)

selection) models (Additional file 3, Tables S4 and S5). The statistical values (ΔlnL) of *ABCG2* and *ABCG2-like* genes were 25 ($p < 0.01$) and 13 ($p < 0.01$), respectively. Therefore, the M2a model was superior to the M1a model.

In addition, the LRT comparison of the M7 (beta) and M8 (beta and $\omega > 1$) models revealed a significant difference between the models. The ΔlnL values of the *ABCG2* and *ABCG2-like* genes were 43 ($p < 0.01$) and 22 ($p < 0.01$), respectively (Additional file 3, Tables S4 and S5), and the M8 model was superior to the M7 model.

Based on the M2a model, a total of 13 positive selection sites (in codons) were found for the *ABCG2* genes (Additional file 4, Tables S6 and S7). Among these sites, three were statistically significant ($p < 0.05$), and two sites were extremely significant ($p < 0.01$). Moreover, a total of 16 positive selection sites were found for the *ABCG2-like* genes (Additional file 4, Tables S6 and S7), and only three of these were statistically significant ($p < 0.05$).

However, we found a total of 56 positive selection sites in the coding regions of the *ABCG2* genes (Additional file 4, Tables S6 and S7) with the M8 model. Six positive selection sites were statistically significant ($p < 0.05$), and

eight positive selection sites were extremely significant ($p < 0.01$). Furthermore, a total of 36 positive selection sites were found for the *ABCG2-like* genes (Additional file 4, Tables S6 and Table S7). Among these, two were statistically significant positive selection sites ($p < 0.05$), and two were extremely significant positive selection sites ($p < 0.01$). Moreover, the analysis of the M3 model revealed that approximately 95.7 and 98.5% of the sites in the *ABCG2* and *ABCG2-like* gene sequences, respectively, were affected by negative selection.

These results showed that the *ABCG2* and *ABCG2-like* genes were mainly subject to strong purifying selection, but the *ABCG2* genes were affected by stronger positive selection pressure in birds compared with the *ABCG2-like* genes. Additionally, the *ABCG2-like* gene sequences were more conserved than the *ABCG2* gene sequences in birds.
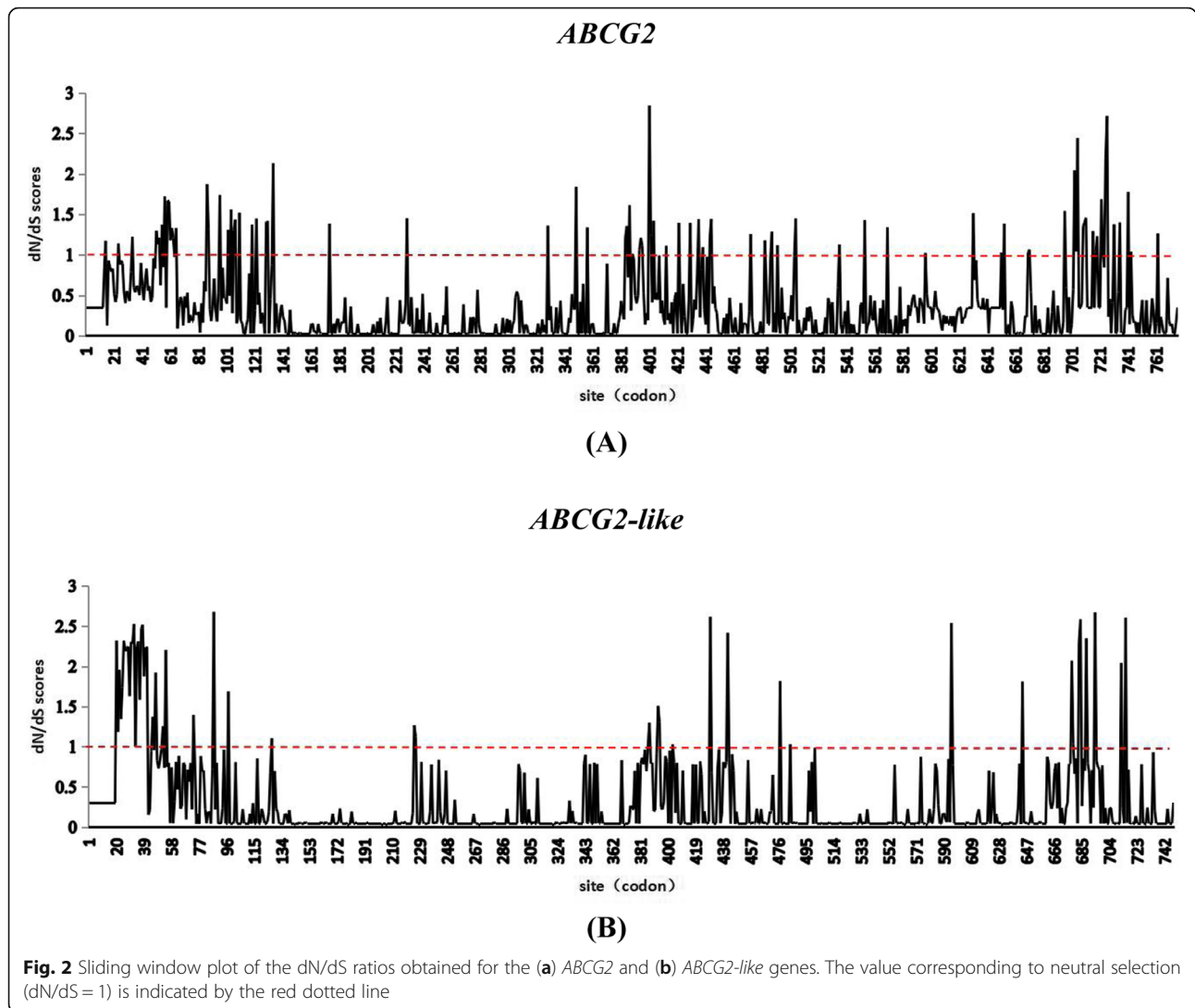
Chromosomal synteny analysis of *ABCG2* and *ABCG2-like* genes

A chromosomal synteny analysis was performed with several representative birds using the Genome Data Viewer (GDV) from the NCBI and Ensembl 94. Conserved synteny dot plots showed that the neighborhood

Table 2 Selection pressure analysis of sites in the coding region of *ABCG2-like* genes

Model code	lnL	Parameters	Number of positive selection sites
M0 (one ratio)	−12,339.59447	$\omega = 0.23234$	None
M1a (nearly neutral)	−11,969.60128	$P0 = 0.77409, P1 = 0.22591$ $\omega0 = 0.03929, \omega1 = 1$	Not allowed
M2a (positive selection)	−11,956.39673	$P0 = 0.77460, P1 = 0.19600, P2 = 0.02940$ $\omega0 = 0.04171, \omega1 = 1, \omega2 = 3.43990$	16
M3 (discrete)	−18,486.07613	$P0 = 0.30973, P1 = 0.31583, P2 = 0.24289, \omega0 = 0, \omega1 = 0.11046, \omega2 = 0.44219$	Not allowed
M7 (beta)	−11,978.85492	$p = 0.10690, q = 0.34331 (p1 = 0.10000), \omega = 0$	Not allowed
M8 (beta&$\omega > 1$)	−11,956.1596	$p0 = 0.95146, p = 0.13558, q = 0.57516, (p1 = 0.04854) \omega = 2.55211$	36

Note: lnL refers to the logarithm of the maximum likelihood (ML)

Ma *et al. BMC Evolutionary Biology*   (2020) 20:85

Page 5 of 14

Fig. 2 Sliding window plot of the dN/dS ratios obtained for the (**a**) *ABCG2* and (**b**) *ABCG2-like* genes. The value corresponding to neutral selection (dN/dS = 1) is indicated by the red dotted line
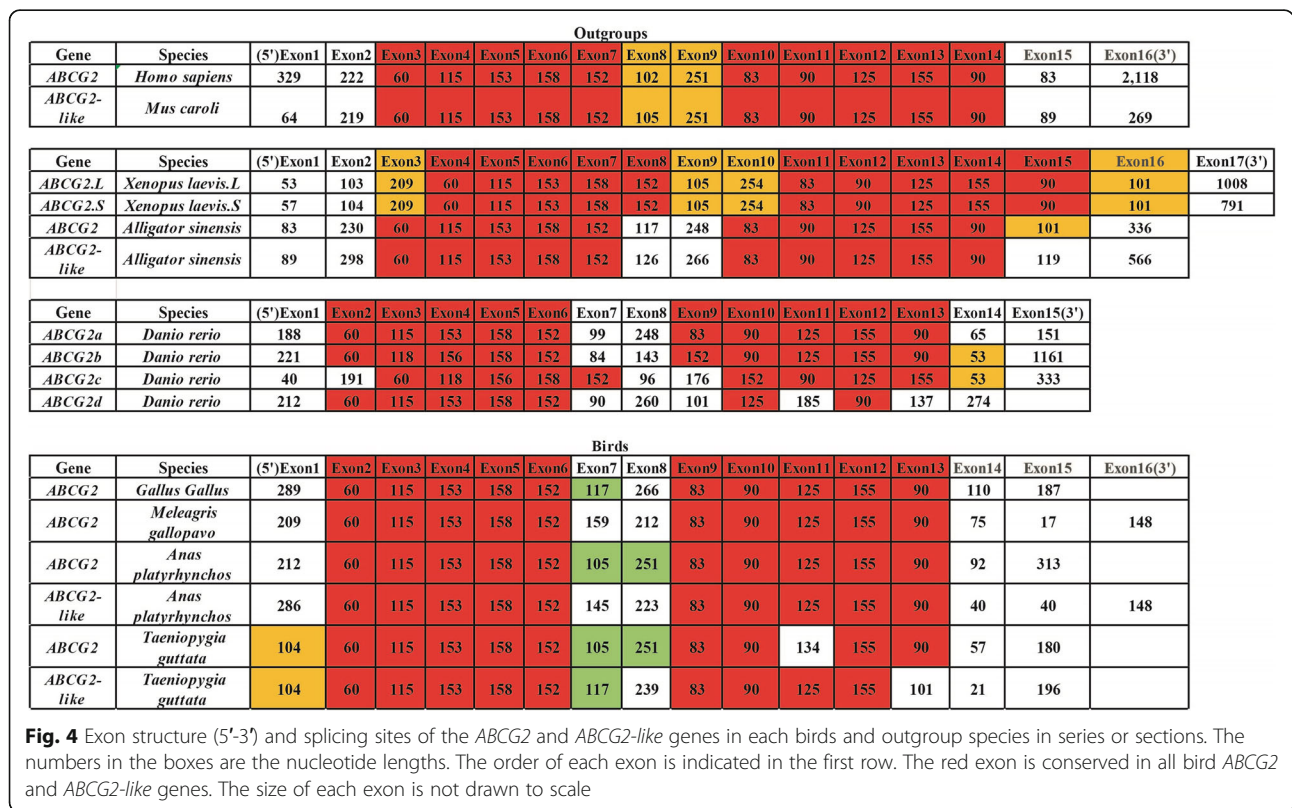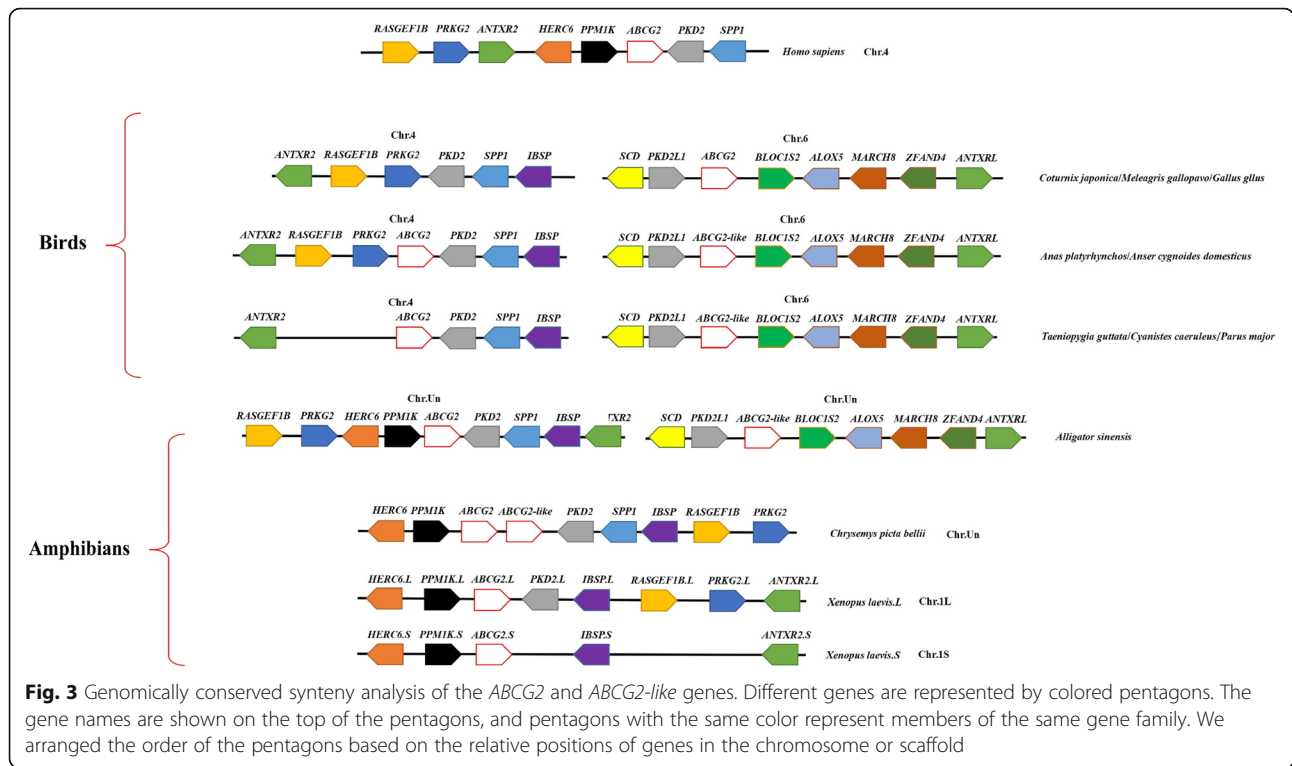
regions of the *ABCG2* and *ABCG2-like* genes in seven birds were similar to those in *Homo sapiens*, *Alligator sinensis*, *Chrysemys picta bellii* and *Xenopus laevis* (Fig. 3). Conserved chromosome segments (Fig. 3) were found in *Anas platyrhynchos* and *Taeniopygia guttata*, and the conserved chromosome segments are *ABCG2-PKD2-SPP1-IBSP* (chromosome 4) and *BLOC1S2-ABCG2-like-PKD2L1-SCD* (chromosome 6). Conserved chromosome segments (Fig. 3) were also found in *Gallus gallus*, *Meleagris gallopavo* and *Coturnix japonica* (deletion of the *ABCG2-like* gene occurred in the genomes of *Gallus gallus*, *Meleagris gallopavo* and *Coturnix japonica*), and these conserved chromosome segments are *PKD2-SPP1-IBSP* (chromosome 4) and *BLOC1S2-ABCG2-PKD2L1-SCD* (chromosome 6). These results reflected the overall evolutionary conservation of *ABCG2* and *ABCG2-like* genes in birds. However, deletion of the *ABCG2* gene occurred in some birds, which was contrary to the characteristics of genome conservation, and

the complex reasons for this result would be worth further study. Moreover, the *ABCG2* and *ABCG2-like* genes were found adjacent to each other on the same chromosome in *Chrysemys picta bellii*. In addition, only a single copy of the *ABCG2* gene was found in the *Xenopus laevis* genome.

Exon/intron structure and splicing site analysis

In the current study, the exon/intron structures of each *ABCG2* or *ABCG2-like* gene sequence were obtained from the 41 birds and outgroup species (Figs. 4b and 5b). An exon/intron structure analysis showed that the CDSs were interrupted by several introns. Compared with *Homo sapiens* and *Mus caroli*, birds harbored significantly shorter *ABCG2* and *ABCG2-like* genes.

According to the structure analysis, 22 *ABCG2* gene sequences and 24 *ABCG2-like* gene sequences were composed of 15 exons, whereas 13 *ABCG2* gene

Fig. 3 Genomically conserved synteny analysis of the *ABCG2* and *ABCG2-like* genes. Different genes are represented by colored pentagons. The gene names are shown on the top of the pentagons, and pentagons with the same color represent members of the same gene family. We arranged the order of the pentagons based on the relative positions of genes in the chromosome or scaffold

Outgroups

Gene	Species	(5')Exon1	Exon2	Exon3	Exon4	Exon5	Exon6	Exon7	Exon8	Exon9	Exon10	Exon11	Exon12	Exon13	Exon14	Exon15	Exon16(3')
ABCG2	*Homo sapiens*	329	222	60	115	153	158	152	102	251	83	90	125	155	90	83	2,118
ABCG2-like	*Mus caroli*	64	219	60	115	153	158	152	105	251	83	90	125	155	90	89	269

Gene	Species	(5')Exon1	Exon2	Exon3	Exon4	Exon5	Exon6	Exon7	Exon8	Exon9	Exon10	Exon11	Exon12	Exon13	Exon14	Exon15	Exon16	Exon17(3')
ABCG2.L	*Xenopus laevis.L*	53	103	209	60	115	153	158	152	105	254	83	90	125	155	90	101	1008
ABCG2.S	*Xenopus laevis.S*	57	104	209	60	115	153	158	152	105	254	83	90	125	155	90	101	791
ABCG2	*Alligator sinensis*	83	230	60	115	153	158	152	117	248	83	90	125	155	90	101	336	
ABCG2-like	*Alligator sinensis*	89	298	60	115	153	158	152	126	266	83	90	125	155	90	119	566	

Gene	Species	(5')Exon1	Exon2	Exon3	Exon4	Exon5	Exon6	Exon7	Exon8	Exon9	Exon10	Exon11	Exon12	Exon13	Exon14	Exon15(3')
ABCG2a	*Danio rerio*	188	60	115	153	158	152	99	248	83	90	125	155	90	65	151
ABCG2b	*Danio rerio*	221	60	118	156	158	152	84	143	152	90	125	155	90	53	1161
ABCG2c	*Danio rerio*	40	191	60	118	156	158	152	96	176	152	90	125	155	53	333
ABCG2d	*Danio rerio*	212	60	115	153	158	152	90	260	101	125	185	90	137	274	

Birds

Gene	Species	(5')Exon1	Exon2	Exon3	Exon4	Exon5	Exon6	Exon7	Exon8	Exon9	Exon10	Exon11	Exon12	Exon13	Exon14	Exon15	Exon16(3')
ABCG2	*Gallus Gallus*	289	60	115	153	158	152	117	266	83	90	125	155	90	110	187	
ABCG2	*Meleagris gallopavo*	209	60	115	153	158	152	159	212	83	90	125	155	90	75	17	148
ABCG2	*Anas platyrhynchos*	212	60	115	153	158	152	105	251	83	90	125	155	90	92	313	
ABCG2-like	*Anas platyrhynchos*	286	60	115	153	158	152	145	223	83	90	125	155	90	40	40	148
ABCG2	*Taeniopygia guttata*	104	60	115	153	158	152	105	251	83	90	134	155	90	57	180	
ABCG2-like	*Taeniopygia guttata*	104	60	115	153	158	152	117	239	83	90	125	155	101	21	196	

Fig. 4 Exon structure (5'-3') and splicing sites of the *ABCG2* and *ABCG2-like* genes in each birds and outgroup species in series or sections. The numbers in the boxes are the nucleotide lengths. The order of each exon is indicated in the first row. The red exon is conserved in all bird *ABCG2* and *ABCG2-like* genes. The size of each exon is not drawn to scale

Ma *et al. BMC Evolutionary Biology* (2020) 20:85

Page 7 of 14

Fig. 5 Phylogenetic analysis (**a**), gene exons/intron structure (**b**) and protein domain structures (**c**) of the *ABCG2* genes in birds. The phylogenetic tree of the amino acid sequences was constructed with the NJ method using MEGA7 software. The numbers on the nodes represent the bootstrap support. The red boxes and black lines (drawn to scale, as shown in the figure below) represent the exons and introns, respectively. The protein motifs are shown as colored boxes based on the MEME identification

sequences and 15 *ABCG2-like* gene sequences were composed of 16 exons. Only two *ABCG2-like* genes were composed of 17 exons, and one *ABCG2* gene was composed of 14 exons (Figs. 5b and 6b). Most members of each individual subfamily contained more than seven similar and conserved exons. The exons marked in red, as shown in Fig. 4, were conserved in length and consisted of 60, 115, 153, 158, 152, 83, 90, 125, 155, 156 and 90 nucleotides. However, the lengths of the first and last exons of the CDSs of the *ABCG2* or *ABCG2-like* genes varied significantly among birds ($p < 0.05$), and these exons were identified in the *ABCG2* or *ABCG2-like* gene sequences of all the studied birds, *Homo sapiens*, *Mus caroli* and *Danio rerio*. At least five conserved exons were copied in series or piecewise in the *ABCG2* and *ABCG2-like* genes, supporting the hypothesis of a common ancestral relationship among *ABCG2* and *ABCG-like* gene sequences (Fig. 4).

Amino acid sequence domain analysis of ABCG2 and ABCG2-like proteins

The protein domains of the ABCG2 and ABCG2-like protein sequences from the 41 birds and outgroup species were predicted in this study (Figs. 5c and 6c). A total of 40 motifs, named 1–40, were identified in the full-length amino acid sequences of the ABCG2 and ABCG2-like proteins (Additional file 5).

Most amino acid sequences of the *ABCG2* gene subfamily shared more than 12 common motifs, which indicated that both ABCG2 and ABCG2-like proteins were highly conserved (Figs. 5c and 6c). However, some small differences were found in the N-terminal region. Most ABCG2-like amino acid sequences of birds harbored a specific motif (named 14 in Figs. 5c and 6c), whereas the ABCG2 amino acid sequences of birds did not have this motif in the N-terminal region.

In addition, the ABCG2 protein sequences of *Anas platyrhynchos* and *Anser cygnoides domesticus* contain seven and eight transmembrane helical structures, respectively, and the ABCG2 protein sequences of *Parus major* contain four transmembrane helical structures. The amino acid sequences of the *ABCG2* genes in other birds have five transmembrane helices, and those of *ABCG2-like* genes in all birds also have five transmembrane helices (Additional file 6). In summary, with the exception of the ABCG2 amino acid sequences of *Anas platyrhynchos*, *Anser cygnoides domesticus* and *Parus major*, the ABCG2 and ABCG2-like amino acid

Fig. 6 Phylogenetic analysis (**a**), gene exons/intron structure (**b**) structures and protein domain structures (**c**) of the *ABCG2-like* genes in birds. The phylogenetic tree of the amino acid sequences was constructed with the NJ method using MEGA7 software. The numbers on the nodes represent the bootstrap support. The red boxes and black lines (drawn to scale) represent the exons and introns, respectively. The protein motifs are shown as colored boxes based on the MEME identification

sequences of most birds have similar transmembrane structures. Moreover, these sequences have transmembrane structures similar to those of the ABCG2 and ABCG2-like proteins of some amphibians.

These results suggested that ABCG2 proteins are homologous to ABCG2-like proteins and that those in the same subgroup might have similar functions. However, functional differences between the two proteins cannot be excluded. Although the function of these conserved motifs has not been elucidated, some of the motifs might determine differences in the transport functions of the two proteins.

Phosphorylation site analysis of *ABCG2* and ABCG2-*like* genes

The potential serine (S), threonine (T), and tyrosine (Y) phosphorylation sites (Additional file 7, Table S8) in the ABCG2 and ABCG2-like protein sequences of the 41 birds were predicted. Approximately 11, 6 and 17% of the S, T and Y residues, respectively, in the ABCG2 amino acid sequences were predicted as phosphorylated sites, and approximately 11, 11 and 13% of the S, T and Y residues, respectively, in the ABCG2-like protein sequences were predicted as phosphorylated sites.

The number of S phosphorylation sites in the ABCG2 protein varied greatly among the different studied birds. Similar results were also obtained for the ABCG2-like protein; however, compared with the number of S sites in ABCG2 proteins, that in the ABCG2-like proteins was

less variable across the birds. According to the results (Additional file 8), all S residues were concentrated in the NBD region of the amino acid sequence, and a small number of phosphorylation sites were located in the TMD regions. Compared with the ABCG2 protein, the ABCG2-like protein contained more phosphorylation sites in the NBD region (Additional file 8). In terms of the total number of phosphorylation sites, we found that the TMD regions of the ABCG2 and ABCG2-like proteins were highly conserved. The differences in the number and distribution of phosphorylation sites between ABCG2 and ABCG2-like amino acid sequences might be related to their functions.

Gene expression patterns in *Anas platyrhynchos* and conversion analysis

To establish the occurrence of gene conversion between paralogs, the nucleotide sequences of the *ABCG2* gene subfamily in some birds and outgroup species were analyzed (Additional file 1, Table S8) using GENECONV and SIMPLOT. The results from GENECONV showed no clear evidence of gene conversion events during the evolutionary process of birds, and no significant evidence demonstrates genetic conversion between the *ABCG2* gene of *Gallus gallus* and the *ABCG2-like* gene of *Anas platyrhynchos*. SIMPLOT also showed high shared sequence identity between the paralogs in the birds and outgroup species, and the results provided no clear evidence of gene conversion (Additional file 9).

Data on the expression of *ABCG2* and *ABCG2-like* genes in mallards were obtained from duckbase (http://duckbase.org/rnaseqExpression; Additional file 10). The expression levels of the *ABCG2* and *ABCG2-like* genes were highest in the spleen and liver, respectively, whereas the *ABCG2* gene was barely expressed in some tissues in mallards (Fig. 7). The expression patterns of these two genes were significantly different in mallard.

Discussion

As an important transmembrane transporter protein, Zhou et al. [31] found that increasing the expression of the ABCG2 transport protein during the process of red blood cell maturation can reduce the level of intracellular protoporphyrin IX. Although the *ABCG2* gene has been investigated in various studies, the evolutionary process of the *ABCG2* gene subfamily members and the functional differences in this gene among birds have never been studied. In the present study, phylogenetic methods and comparative genomics were used to investigate the molecular evolution characteristics of *ABCG2* and *ABCG2-like* genes in birds. After an extensive database statistical analysis, *ABCG2* subfamily genes were found to be widely present in chordates and vertebrates. Overall, *ABCG2* or *ABCG2-like* genes were found in 41 birds, and *ABCG2-like* genes were lost in only five birds.

Origin, duplication events and conversion of the *ABCG2* and *ABCG2-like* genes in birds

The genomic structure of birds is relatively evolutionarily stable in evolution [32], and chromosomes 1–10 and Z are the ancestors of almost all chromosomes in birds [33]. Moreover, two rounds (2R) of genome duplication occurred during the early diversification of chordates and vertebrates [34, 35], which provides a theoretical basis for studying the evolution of the *ABCG2* and *ABCG2-like* genes in birds.

Based on a chromosomal synteny analysis (Fig. 3), *ABCG2* or *ABCG2-like* genes share similar gene neighborhoods in birds (including early birds), *Alligator sinensis* and *Xenopus laevis*. Furthermore, two members of the *ABCG2* gene subfamily were adjacent to each other in *Chrysemys picta bellii* and shared similar chromosomal neighborhoods with the *ABCG2* gene of birds. First, these results showed highly conserved chromosomal synteny in the neighborhood regions of *ABCG2* or *ABCG2-like* genes in birds. Moreover, these results confirmed that the characteristics of ancestral chromosomes were highly conserved and exhibit low segment recombination rates in birds [33, 36]. Although the NCBI database did not specify which chromosome in *Chrysemys picta bellii* harbored the *ABCG2* and *ABCG2-like* genes, Matusda Y et al. [37] found that chromosome 4 of birds is highly homologous with
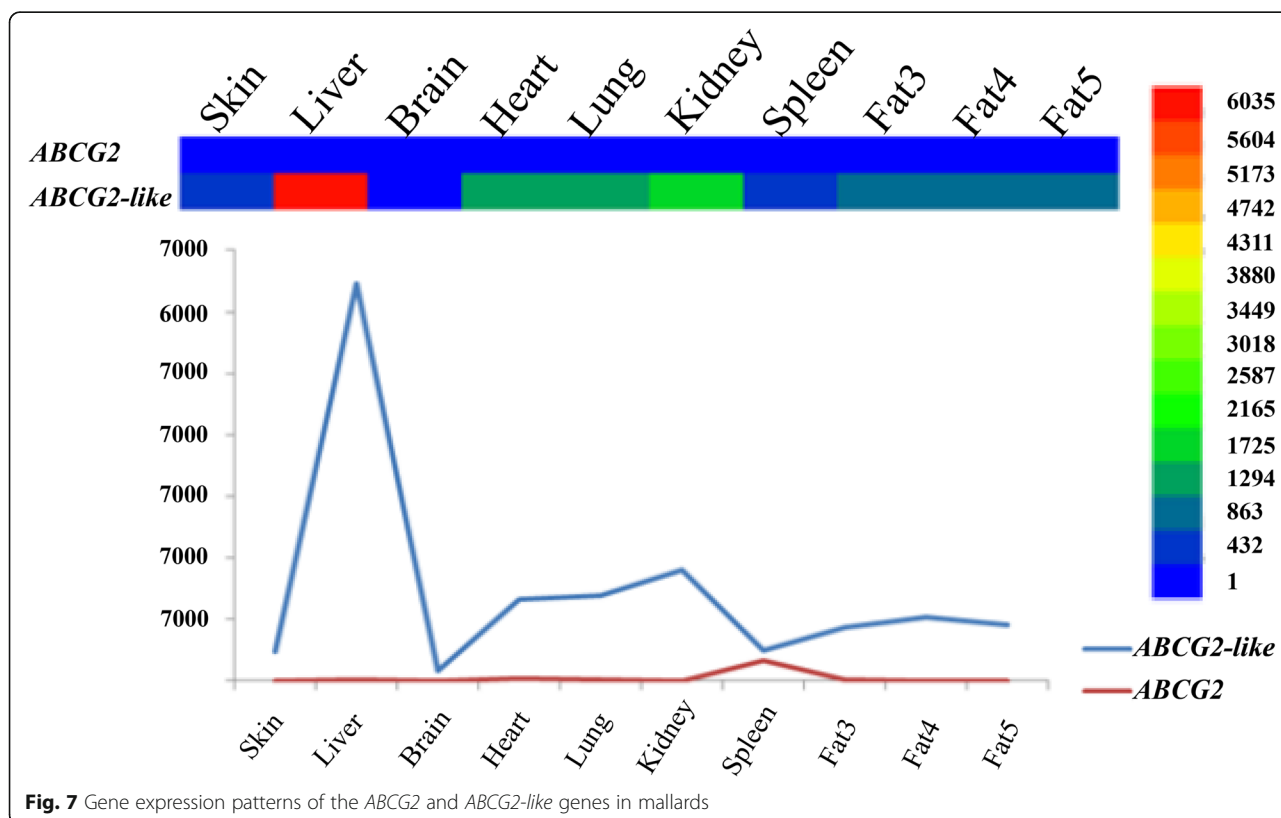
Fig. 7 Gene expression patterns of the *ABCG2* and *ABCG2-like* genes in mallards

chromosome 4 of turtles (particularly between chickens and soft-shelled turtles). Chromosome 4 is very old in bird genomes and can be characterized by an early origin and strong evolutionary conservatism [38]. Based on the above-described results, we hypothesized that the *ABCG2* and *ABCG2-like* genes in *Chrysemys picta bellii* are also located on chromosome 4. We also assumed that the ancestors of the *ABCG2* gene subfamily members existed on chromosome 4 or the ancestor of chromosome 4 for a long time during the evolutionary process (Fig. 3), and this assumption is supported by the results from a phylogenetic analysis.

We also found that even more *ABCG2* gene subfamily members were located on the same chromosome in some fishes (Additional file 11). For example, the *ABCG2* gene and multiple *ABCG2-like* genes were located on chromosome 25 in *Astyanax mexicanus*, which suggested that multiple members of the *ABCG2* gene subfamily already existed in the same chromosome in early chordates. However, the evolutionary connection between the chromosome in fishes and that in amphibians remains unclear, and we can trace the origin of *ABCG2* and *ABCG2-like* genes back only to amphibians. We hypothesize that during the evolution of birds from fishes, most members (or multiple copies of a single member) of the *ABCG2* gene subfamily were lost, and ultimately, only two *ABCG2* gene subfamily members were retained in birds. Moreover, only one copy of the *ABCG2* gene in frog genomes is located on chromosome 1. These findings produce an unusual situation. The reasons for this phenomenon are very complicated, and no clear conclusion has been reached. Some studies have suggested that chromosome fusion occurred during the process of speciation in frogs [39], which might have caused the deletion of *ABCG2* gene subfamily members in frogs. These findings provides a reference for explaining the large number of deletions of *ABCG2* gene subfamily members during the evolution of birds from fishes.

Some of the *ABCG2-like* genes in birds are located on chromosome 6. *ABCG2-like* neighborhoods similar to those in birds have been found in some alligator genomes (Fig. 3). Based on the results from the phylogenetic and chromosomal synteny analyses, we hypothesized that the presence of *ABCG2* and *ABCG2-like* genes on different chromosomes could be traced back to either after the differentiation of turtle and birds or after the 2R WGD event.

We also found that one or more members of the *ABCG2* gene subfamily were also located on microchromosomes in some fishes, e.g., two *ABCG2-like* genes are located on the LG5 chromosome (Additional file 11). Some of these members have similar neighborhood segments among bird lineages and amphibians

(Additional file 11). Therefore, we speculated that multiple microchromosomes containing *ABCG2-like* genes were further fused into a complete chromosome and that multiple *ABCG2-like* genes were lost during this process. Ultimately, only one *ABCG2-like* gene was retained in birds, and this gene did not originate from the WGD event. The timing of the appearance of chromosomes 6 to 9 in birds has not yet been determined [33]. Studying the origin of *ABCG2* and *ABCG2-like* genes can provide a reference for studying the origin of chromosome 6 in birds.

The current study provides no clear evidence of genetic conversion between *ABCG2* and *ABCG2-like* genes (Additional file 9), which supports the hypothesis of an independent origin of *ABCG2* and *ABCG2-like* genes. Olsen et al. [22] found that the *ABCG2-PKD2-SPP1* segment was located in a QTL. In the present study, the above segment was also found in some birds. Therefore, it can be speculated that the *ABCG2* gene controls similar quantitative traits in birds.

Influence of positive and purifying selection on the *ABCG2* and *ABCG2-like* genes in birds

A selection pressure analysis of the *ABCG2* and *ABCG2-like* genes showed that both genes were more affected by purifying selection than by positive selection pressure. These results further supported the hypothesis that both *ABCG2* and *ABCG2-like* genes are highly conserved in birds. We also inferred that both *ABCG2* and *ABCG2-like* genes have some similar functions but exhibit some differences, e.g., regulation of some quantitative traits. The selection pressure sites were mostly concentrated at the N terminus (NBD region), which also indicated a difference in the N terminus between the two genes that might affect their function. Compared with the *ABCG2-like* genes, the *ABCG2* genes were under stronger positive selection pressure. Therefore, the *ABCG2* gene was more evolutionarily active. This finding motivates studies of *ABCG2* and *ABCG2-like* gene deletions in some birds.

Gene and protein structures of the *ABCG2* and *ABCG2-like* genes in birds and functional differences between these genes during evolution

An exon/intron structure analysis can provide valuable information on duplication events within gene families that occurred during eukaryotic evolution, and the gain and loss of introns reflect positive or negative correlations with the CDS evolutionary rate [40]. In addition, an exon/intron structure analysis provides a theoretical reference for exploring the functional differences in gene families. Based on the gene structure analysis, all *ABCG2* or *ABCG2-like* genes, with the exception of the *Balearica regulorum gibbericeps ABCG2* gene, contained more

Ma *et al. BMC Evolutionary Biology*   (2020) 20:85

Page 11 of 14

than 15 exons in their CDSs (Figs. 5 and 6). A sequence length of more than seven exons was conserved in birds (Fig. 7). However, birds contained much shorter introns than other outgroup species (Figs. 5 and 6), which suggests that the *ABCG2* and *ABCG2-like* genes have similar functions. We thus conclude that the *ABCG2* and *ABCG2-like* genes perform unique functions in birds.

Through an analysis of the ABCG2 and ABCG2-like protein domains, one additional motif was found in the N terminus of *ABCG2-like* (Figs. 5 and 6; Additional file 5), and this motif could be responsible for the functional differences between the two genes. However, more than 12 similar motifs were found to exist in the whole proteins of various birds (Figs. 5 and 6; Additional file 5), supporting the notion of strong conservation and similarity among ABCG2 and ABCG2-like proteins, and the analysis also revealed some inaccurate gene annotations. A detailed analysis clarified some of the inaccurate annotations, such as *ABCG2* of *Gallus gallus*, *Meleagris gallopavo*, and *Coturnix japonica*, which should be *ABCG2-like* genes based on a phylogenetic analysis of the three species in the same clade. The above gene and protein structural analyses provide evidence for the origin of *ABCG2* and *ABCG2-like* genes from 2R WGD.

The posttranslational modification of proteins can affect the biochemical properties of proteins and plays an important role in maintaining biological processes in cells. Protein phosphorylation can change the structure and functions of a protein [41]. A protein can have one or more phosphorylation sites, which make the protein structurally diverse. A phosphorylation site analysis of the amino acid sequences based on the *ABCG2* and *ABCG2-like* genes (Additional files 7 and 8) suggested that the phosphorylation sites were also concentrated in the ATP-binding domain (N terminus) of ABCG2 and ABCG2-like proteins, but there were also certain differences. The number of S and T sites in the *ABCG2* gene was significantly larger than that in the *ABCG2-like* gene, which indicated that the structural differences between *ABCG2* and *ABCG2-like* genes at the N terminus might affect the actual functions of these genes in birds.

Expression and function of *ABCG2* and *ABCG2-like* genes

Studying the expression of gene families in animal tissues can provide a reference for exploring the functional differences between gene families. The expression level of the *ABCG2-like* gene was significantly higher than that of the *ABCG2* gene in most bird tissues (e.g., liver, heart, lung, and kidney; Fig. 7). The *ABCG2* gene was almost not expressed in some tissues. We speculated that the *ABCG2-like* gene might play a more important role than the *ABCG2* gene or functionally replace this gene in some tissues of mallards. Alternatively, the *ABCG2-like* gene replaced the function of the *ABCG2* gene in

some tissues of mallard. Even though some genomes of birds have lost the *ABCG2* and *ABCG2-like* genes, we speculated that ABCG2 might be functionally redundant in some birds and that its deletion can thus be tolerated. The expression patterns of *ABCG2* and *ABCG2-like* genes in birds require further research. Overall, the results of the present study provide new ideas for future studies on the deletion of *ABCG2* gene subfamily members throughout the evolutionary process of vertebrates and chordates and on the deletion of *ABCG2* or *ABCG2-like* genes in some birds.

Conclusion

The diversity of the *ABCG2* gene subfamily members has declined in birds, and most birds have only retained *ABCG2* and *ABCG2-like* genes. Here, we speculated that the *ABCG2* and *ABCG2-like* genes might have originated from the same ancestral chromosome and that these two genes might have been produced via genome duplication events during the evolution of amphibians to birds. The protein sequences of the *ABCG2* and *ABCG2-like* genes were structurally conserved and homologous, but these sequences were less conserved at the N terminus. These results indicate that the functions of *ABCG2* and *ABCG2-like* genes are generally similar; however, differences in the N-terminal structure might have led to the functional differences between the two genes in some birds.

Methods

Acquisition and identification of the *ABCG2* and *ABCG2-like* genes in birds

The nucleic acid coding sequences and amino acid sequences used in this study were obtained from NCBI (https://www.ncbi.nlm.nih.gov/). We used the ABCG2 amino acid sequence of *Anas platyrhynchos* in a BLASTP search and obtained 41 ABCG2 amino acid sequences and 36 ABCG2-like amino acid sequences from 41 representative birds. All obtained sequences had E-scores less than 0.01. The corresponding nucleic acid sequences were obtained using tblastn [42]. The amino acid sequences identified by BLAST were used in a BLASTX search to ensure that the nucleic acid and amino acid sequences of each *ABCG2* and *ABCG2-like* gene and protein matched each other [43]. We completed all of the sequence searches in August 2018.

Phylogenetic analysis

Nucleic acid sequences (CDSs) were used for the phylogenetic analysis of *ABCG2* and *ABCG2-like* genes in birds. A nucleic acid sequence alignment was performed using Clustal Omega included in MEGA7 (Additional file 12). Revised sequence alignments were then submitted to MEGA7 to select the appropriate DNA

Ma *et al. BMC Evolutionary Biology*   (2020) 20:85

Page 12 of 14

evolution model according to our dataset [44]. Here, we found that the nucleic acid sequence group of the *ABCG2* gene subfamily followed a K2 + G + I model.

MEGA7 software was used to construct a bootstrap (1000 replicate) tree [45] of the nucleic acid sequences, and the ML method was used in the phylogenetic analysis. The ML search was started with the tree generated using BIONJ [46], and the optimal tree was determined through a heuristic search using the nearest-neighbor interchange (NNI) algorithm [43].

Selective pressure analysis

In the selective pressure analysis, we obtained values of omega (nonsynonymous/synonymous replacement rate ratio, dN/dS) to analyze the evolutionary selection pressure at the molecular level. The variable omega (ω) intuitively reflects the evolutionary trend of organisms at the codon level, and omega> 1, omega = 1 and omega< 1 represent genes subjected to positive selection, neutral selection and negative selection (purification selection) during evolution, respectively [47].

The ω values were calculated using the CodeML program in the PAML 4.9 package [48]. We selected the M0, M1a, M2a, M3, M7 and M8 models (site models) [49, 50] for the following reasons: (1) the M0 model allows uniform selection pressure between different sites in a sequence, and the M1a, M7, and M8 models do not allow sites with $\omega > 1$; (2) the M3 model assumes variable selection pressure between sites; and (3) the M2a and M8 models allow sites with $\omega > 1$ [51]. With LRTs, we compared three groups (M1a versus M2, M0 versus M3 [52–55], and M7 versus M8 [51]) to infer the most suitable model.

We then used M2a and M8 to identify the positive selection sites. The Bayes empirical Bayes (BEB) calculation method was adopted to identify the positive selection sites, and the posterior probability (PP) of these sites was analyzed [56]. We considered only positive sites with a PP > 95%.

Sequence analysis

First, we used GSDS online analysis software (http://gsds.cbi.pku.edu.cn/) [57] to analyze the exon/intron structure and exon distribution patterns of *ABCG2* and *ABCG2-like* genes. To study the conserved motifs of ABCG2 and ABCG2-like proteins, MEME online analysis software (http://meme.nbcr.net/meme/intro.html) was used [58] to predict protein structural domains. The optimized parameters of MEME were as follows: maximum number of motifs, 40 [59]; optimal motif width, 10–100 residues [59], and optimal width of each motif, 10–100 residues [59]. TMHMM software (http://www.cbs.dtu.dk/services/TMHMM/) was used to predict the transmembrane structure domain of the ABCG2 and

ABCG2-like amino acid sequences [60]. The ABCG2 and ABCG2-like protein sequence phosphorylation sites were predicted using the online program KinasePhos 2.0 (http://kinasephos2.mbc.nctu.edu.tw/) [61] with the default parameters. The aligned sequences were then examined for possible gene conversion events by constructing a sliding window genetic diversity plot (SIMPLOT 3.5.1) [62]. GENECONV 1.8 [63] software was used for conversion analysis using the default parameters and a global segment *p* value ($p < 0.05$) corrected with 10,000 pseudoreplicates.

Abbreviations

ABC: ATP-binding cassette; ABCG2: ATP-binding cassette subfamily G member 2; ABCG2-L: ATP-binding cassette subfamily G member 2-like; BCRP: breast cancer resistance protein; BEB: Bayes empirical Bayes; CDS: coding sequence; dN: nonsynonymous; dS: synonymous; GDV: Genome Data Viewer; LECA: last eukaryotic common ancestor; LRT: likelihood ratio test; ML: maximum likelihood; NBD: nucleotide-binding domain; NCBI: National Center for Biotechnology Information; NNI: nearest-neighbor interchange; QTL: quantitative trait locus; PP: posterior probability; S: serine; T: threonine; TMD: transmembrane domain; WGD: whole-genome duplication; Y: tyrosine; 2R: two rounds; ω: omega

Ma *et al. BMC Evolutionary Biology*   (2020) 20:85

Page 13 of 14

Availability of data and materials

All data and materials are shown within the manuscript or additional files. These data and materials are fully available without restriction.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Farm Animal Genetic Resources Exploration and Innovation Key Laboratory of Sichuan Province, Sichuan Agricultural University, Chengdu, Sichuan 611130, P.R. China. [2]Institute of Animal Nutrition, Key Laboratory for Animal Disease-Resistance Nutrition of China, Ministry of Education, Sichuan Agricultural University, Chengdu, P.R. China.

References

1. Doyle LA, Yang W, Abruzzo LV, Krogmann T, Gao Y, Rishi AK, Ross DD. A multidrug resistance transporter from human MCF-7 breast cancer cells. Proc Natl Acad Sci U S A. 1998;95(26):15665–70.
2. Ni Z, Bikadi Z, Rosenberg MF, Mao Q. Structure and function of the human breast cancer resistance protein (BCRP/ABCG2). Curr Drug Metab. 2010;11(7):603–17.
3. Sarkadi B, Homolya L, Szakács G, Váradi A. Human multidrug resistance ABCB and ABCG transporters: participation in a chemoimmunity defense system. Physiol Rev. 2006;86(4):1179–236.
4. Theodoulou FL, Kerr ID. ABC transporter research: going strong 40 years on. Biochem Soc Trans. 2015;43(5):1033–40.
5. Bhatia A, Schäfer H-J, Hrycyna CA. Oligomerization of the human ABC transporter ABCG2: evaluation of the native protein and chimeric dimers. Biochemistry. 2005;44(32):10893–904.
6. Rocchi E, Khodjakov A, Volk EL, Yang CH, Litman T, Bates SE, Schneider E. The product of the ABC half-transporter gene ABCG2 (BCRP/MXR/ABCP) is expressed in the plasma membrane. Biochem Biophys Res Commun. 2000;271(1):42–6.
7. Doyle LA, Ross DD. Multidrug resistance mediated by the breast cancer resistance protein BCRP (ABCG2). Oncogene. 2003;22(47):7340–58.
8. Zhou S, Schuetz JD, Bunting KD, Colapietro A-M, Sampath J, Morris JJ, Lagutina I, Grosveld GC, Osawa M, Nakauchi H, Sorrentino BP. The ABC transporter Bcrp1/ABCG2 is expressed in a wide variety of stem cells and is a molecular determinant of the side-population phenotype. Nat Med. 2001;7(9):1028–34.
9. Ifergan I, Jansen G, Assaraf YG. Cytoplasmic confinement of breast cancer resistance protein (BCRP/ABCG2) as a novel mechanism of adaptation to short-term folate deprivation. Mol Pharmacol. 2005;67(4):1349–59.
10. Ifergan I, Shafran A, Jansen G, Hooijberg JH, Scheffer GL, Assaraf YG. Folate deprivation results in the loss of breast cancer resistance protein (BCRP/ABCG2) expression. A role for BCRP in cellular folate homeostasis. J Biol Chem. 2004;279(24):25527–34.
11. Scharenberg C, Mannowetz N, Robey RW, Brendel C, Repges P, Sahrhage T, Jähn T, Wennemuth G. ABCG2 is expressed in late spermatogenesis and is associated with the acrosome. Biochemi Biophys Res Commun. 2009;378(2):302–7.
12. Shen B, Dong P, Li D, Gao S. Expression and function of ABCG2 in head and neck squamous cell carcinoma and cell lines. Experimental Therapeutic Med. 2011;2(6):1151–7.
13. Ebert B, Seidel A, Lampen A. Identification of BCRP as transporter of benzo[a]pyrene conjugates metabolically formed in Caco-2 cells and its induction by ah-receptor agonists. Carcinogenesis. 2005;26(10):1754–63.
14. Dean M. Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates. Annu Rev Genomics Hum Genet. 2005;6(1):123–42.
15. Xiong J, Feng J, Yuan D, Zhou J, Miao W. Tracing the structural evolution of eukaryotic ATP binding cassette transporter superfamily. Sci Rep. 2015;5:16724.
16. Mable BK, Alexandrou MA, Taylor MI. Genome duplication in amphibians and fish: an extended synthesis. J Zool. 2011;284(3):151–82.
17. Long M, Thornton K. Gene duplication and evolution. Science. 2002;297(5583):945–7.
18. Chen ZJ, Wang J, Tian L, Lee H-S, Wang JJ, Chen M, Lee JJ, Josefsson C, Madlung A, Watson B. The development of an Arabidopsis model system for genome-wide analysis of polyploidy effects. Biol J Linn Soc. 2004;82(4):689–700.
19. Seret ML, Diffels JF, Goffeau A, Baret PV. Combined phylogeny and neighborhood analysis of the evolution of the ABC transporters conferring multiple drug resistance in hemiascomycete yeasts. BMC Genomics. 2009;10(1):459.
20. Annilo T, Chen ZQ, Shulenin S, Costantino J, Thomas L, Lou H, Stefanov S, Dean M. Evolution of the vertebrate ABC gene family: analysis of gene birth and death ☆. Genomics. 2006;88(1):1–11.
21. Tantia MS, Vijh RK, Mishra BP, Mishra B, Kumar SB, Sodhi M. DGAT1 and ABCG2 polymorphism in Indian cattle (Bos indicus) and buffalo (Bubalus bubalis) breeds. BMC Vet Res. 2006;2(1):32.
22. Alim MA, Xie Y, Fan Y, Wu X, Sun D, Zhang Y, Zhang S, Zhang Y, Zhang Q, Liu L. Genetic effects of ABCG2 polymorphism on milk production traits in the Chinese Holstein cattle. J Appl Anim Res. 2013;41(3):333–8.
23. Cohen-Zinder M, Seroussi E, Larkin DM, Loor JJ, Everts-van der Wind A, Lee J-H, Drackley JK, Band MR, Hernandez A, Shani M, Lewin HA, Weller JI, Ron M. Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome Res. 2005;15(7):936–44.
24. Woodward OM, Köttgen A, Coresh J, Boerwinkle E, Guggino WB, Köttgen M, Burg MB. Identification of a Urate transporter, ABCG2, with a common functional polymorphism causing gout. Proc Natl Acad Sci U S A. 2009;106(25):10338–42.
25. Candeil L, Gourdier I, Peyron D, Vezzio N, Copois V, Bibeau F, Orsetti B, Scheffer GL, Ychou M, Khan QA. ABCG2 overexpression in colon cancer cells resistant to SN38 and in irinotecan-treated metastases. Int J Cancer. 2004;109(6):848–54.
26. Blazquez AG, Oscar B, Romero MR, Ruben R, Monte MJ, Javier V, Macias RIR, Doris C, Marin JJ. Characterization of the role of ABCG2 as a bile acid transporter in liver and placenta. Mol Pharmacol. 2012;81(2):273–83.
27. Maliepaard M, Scheffer GL, Faneyte IF, Gastelen MA, Van PAC, Schinkel AH, Vijver MJ, Van De SRJ, Schellens JH. Subcellular localization and distribution of the breast cancer resistance protein transporter in normal human tissues. Cancer Res. 2001;61(8):3458–64.
28. Qiang GH, Yu DC, Ding XW. Expression of ABCG2 in human liver Cancer cell lines and its related functions. Chin J Bases Clin Gen Surg. 2012;19(02):146–50.
29. Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP, Lemmon EM, Lemmon AR. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature. 2015;526(7574):569–73.
30. Gao F, Zou W, Xie L, Zhan J. Adaptive evolution and demographic history contribute to the divergent population genetic structure ofPotato virus Ybetween China and Japan. Evol Appl. 2017;10(4):379–90.
31. Zhou S, Zong YP, Nair G, Stewart CF, Sorrentino BP. Increased expression of the Abcg2 transporter during erythroid maturation plays a role in decreasing cellular protoporphyrin IX levels. Blood. 2005;105(6):2571–6.
32. Shibusawa M, Nishibori M, Nishida-Umehara C, Tsudzuki M, Masabanda J, Griffin DK, Matsuda Y. Karyotypic evolution in the Galliformes: an examination of the process of karyotypic evolution by comparison of the

molecular cytogenetic findings with the molecular phylogeny. Cytogenetic Genome Res. 2004;106(1):111–9.

33. Griffin DK, Robertson L, Tempest HG, Skinner BM. The evolution of the avian genome as revealed by comparative molecular cytogenetics. Cytogenetic Genome Res. 2007;117(1–4):64–77.

34. Sidow A. Gen(om)e duplications in the evolution of early vertebrates. Curr Opin Genet Dev. 1996;6(6):715–22.

35. Meyer A, Van dP. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). Bioessays. 2010;27(9):937–45.

36. Shetty S, Griffin DK, Graves M. Comparative painting reveals strong chromosome homology over 80 million years of bird evolution. Chromosom Res. 1999;7(4):289–95.

37. Matsuda Y, Nishida-Umehara C, Tarui H, Kuroiwa A, Yamada K, Isobe T, Ando J, Fujiwara A, Hirao Y, Nishimura O. Highly conserved linkage homology between birds and turtles: bird and turtle chromosomes are precise counterparts of each other. Chromosom Res. 2005;13(6):601–15.

38. Chowdhary BP, Raudsepp T. HSA4 and GGA4: remarkable conservation despite 300-Myr divergence. Genomics. 2000;64(1):102–5.

39. Voss SR, Kump DK, Putta S, Pauly N, Reynolds A. Origin of amphibian and avian chromosomes by fission, fusion, and retention of ancestral chromosomes. Genome Res. 2011;21(8):1306–12.

40. Du XX, Liu YZ, Liu JX, Zhang QQ, Wang XB. Evolution history of duplicated smad3 genes in teleost: insights from japanese flounder, paralichthys olivaceus. PeerJ. 2016;4:e2500.

41. Wu X, Tian L, Li J, Zhang Y, Han V, Li Y, Xu X, Li H, Chen X, Chen J, Jin WH, Xie YM, Han JH, Zhong CQ. Investigation of receptor interacting protein (RIP3)-dependent protein phosphorylation by quantitative phosphoproteomics. Mol Cell Proteomics. 2012;11(12):1640–51.

42. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25(17):3389–402.

43. Moore GW, Goodman M, Barnabas J. An iterative approach from the standpoint of the additive hypothesis to the dendrogram problem posed by molecular data sets. J Theor Biol. 1973;38(3):423–57.

44. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870–4.

45. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 1985;39(4):783–91.

46. Gascuel O. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol Biol Evolution. 1997;14(7):685–95.

47. Almeida D, Maldonado E, Khan I, Silva L, Gilbert MTP, Zhang G, Jarvis ED, O'Brien SJ, Johnson WE, Antunes A. Whole-genome identification, phylogeny, and evolution of the cytochrome P450 family 2 (CYP2) subfamilies in birds. Genome Biol Evol. 2016;8(4):1115–31.

48. Yang ZH. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91.

49. Nielsen R, Yang ZH. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics. 2009;148(3):929–36.

50. Yang ZH. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus a. J Mol Evol. 2000;51(5):423–32.

51. Wong WS, Yang ZH, Goldman N, Nielsen R. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics. 2004; 168(2):1041–51.

52. Anisimova M, Bielawski JP, Yang ZH. Accuracy and power of the likelihood ratio test in detecting adaptivemolecular evolution. Mol Biol Evol. 2001; 18(8):1585–92.

53. Anisimova M, Bielawski JP, Yang ZH. Accuracy and power of bayes prediction of amino acid sites under positive selection. Mol Biol Evol. 2002; 19(6):950–8.

54. Suzuki Y, Nei M. Reliabilities of parsimony-based and likelihood-based methods for detecting positive selection at single amino acid sites. Mol Biol Evol. 2001;18(12):2179–85.

55. Suzuki Y, Nei M. Simulation study of the reliability and robustness of the statistical methods for detecting positive selection at single amino acid sites. Mol Biol Evol. 2002;19(11):1865–9.

56. Wong WS, Yang ZH, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol. 2005;22(4):1107–18.

57. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2014;31(8):1296–7.

58. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren JY, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37(suppl_2):W202–8.

59. Liu CY, Xie T, Chen CJ, Luan AP, Long JM, Li CH, Ding YQ, He YH. Genome-wide organization and expression profiling of the R2R3-MYB transcription factor family in pineapple (Ananas comosus). BMC Genomics. 2017;18(1):503.

60. Krogh A, Larsson B, Heijne GV, Ell S. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305(3):567–80.

61. Huang HD, Lee TY, Tzeng SW, Horng JT. KinasePhos: a web tool for identifying protein kinase-specific phosphorylation sites. Nucleic Acids Res. 2005;33(suppl_2):W226–9.

62. Sawyer S. Statistical tests for detecting gene conversion. Mol Biol Evolution. 1989;6(5):526–38.

63. Lole K, Bollinger R, RS GD, Kulkarni S, Novak N, Ingersoll R, Sheppard H, Ray SC. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol. 1999;73(1):152–60.

Publisher's Note