RESEARCH ARTICLE

Open Access

Gain and loss events in the evolution of the apolipoprotein family in vertebrata

Jia-Qian Liu[1], Wen-Xing Li[2,3], Jun-Juan Zheng[2,3], Qing-Nan Tian[1*], Jing-Fei Huang[2*] and Shao-Xing Dai[2*]

Abstract

Background: Various apolipoproteins widely distributed among vertebrata play key roles in lipid metabolism and have a direct correlation with human diseases as diagnostic markers. However, the evolutionary progress of apolipoproteins in species remains unclear. Nine human apolipoproteins and well-annotated genome data of 30 species were used to identify 210 apolipoprotein family members distributed among species from fish to humans. Our study focused on the evolution of nine exchangeable apolipoproteins (ApoA-I/II/IV/V, ApoC-I~IV and ApoE) from *Chondrichthyes*, *Holostei*, *Teleostei*, *Amphibia*, *Sauria* (including *Aves*), *Prototheria*, *Marsupialia* and *Eutheria*.

Results: In this study, we reported the overall distribution and the frequent gain and loss evolutionary events of apolipoprotein family members in vertebrata. Phylogenetic trees of orthologous apolipoproteins indicated evident divergence between species evolution and apolipoprotein phylogeny. Successive gain and loss events were found by evaluating the presence and absence of apolipoproteins in the context of species evolution. For example, only ApoA-I and ApoA-IV occurred in cartilaginous fish as ancient apolipoproteins. ApoA-II, ApoE, and ApoC-I/ApoC-II were found in *Holostei, Coelacanthiformes*, and *Teleostei,* respectively, but the latter three apolipoproteins were absent from *Aves*. ApoC-I was also absent from *Cetartiodactyla*. The apolipoprotein ApoC-III emerged in terrestrial animals, and ApoC-IV first arose in *Eutheria*. The results indicate that the order of the emergence of apolipoproteins is most likely ApoA-I/ApoA-IV, ApoE, ApoA-II, ApoC-I/ApoC-II, ApoA-V, ApoC-III, and ApoC-IV.

Conclusions: This study reveals not only the phylogeny of apolipoprotein family members in species from *Chondrichthyes* to *Eutheria* but also the occurrence and origin of new apolipoproteins. The broad perspective of gain and loss events and the evolutionary scenario of apolipoproteins across vertebrata provide a significant reference for the research of apolipoprotein function and related diseases.

Keywords: Phylogenesis, Gain and loss events, Divergence, Apolipoprotein, Vertebrata

Background

Apolipoproteins are components of plasma lipoproteins and are mainly synthesized in the small intestine and liver. The major human apolipoproteins include ApoA, ApoB, ApoC, and ApoE. ApoA-I and ApoA-II are the major proteins of plasma high-density lipoproteins (HDLs); both ApoA-I and ApoA-IV are found in chylomicrons and HDLs, while a large portion of ApoA-IV is lipid free [1–3]. Human ApoA-V is present at a very low concentration in plasma and as a fraction of very low-density lipoproteins (VLDLs), HDLs and chylomicrons [4, 5]. ApoCs are constituents of chylomicrons, VLDLs, and HDLs [6], and ApoE is a component of various lipoprotein particles, including VLDLs and their components, chylomicrons and their components and HDLs [7], in humans.

As a crucial structural component of lipoproteins, apolipoproteins form different kinds of lipoproteins and play significant roles in lipid transport and lipoprotein metabolism [6, 8–12]. Abnormalities in apolipoproteins are a risk factor for a variety of human diseases. For example, a deficiency in ApoC-II and 24 genetic mutations can cause hypertriglyceridemia (HTG) [13]. The level of ApoC-III in human plasma is elevated in hyperlipidemia and diabetes [14, 15]. Low-density lipoprotein (LDL) accumulation in plasma resulting from a deficiency in LDL

* Correspondence: tianqn@zzu.edu.cn; huangjf@mail.kiz.ac.cn; daishx08@gmail.com
[1]School of Life Sciences, Zhengzhou University, Zhengzhou 450001, China
[2]Key Laboratory of Animal Models and Human Disease Mechanisms, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China
Full list of author information is available at the end of the article

Liu *et al. BMC Evolutionary Biology*   (2019) 19:209

Page 2 of 10

receptors (ApoB and ApoE) is related to the onset of accelerated coronary artery disease [16]. In addition, ApoE gene allelic variants [17, 18] and single nucleotide polymorphisms (SNPs) in different regions [19], together with protein level, are associated with the risk of Alzheimer's disease [20–24].

The human exchangeable apolipoproteins APOA/C/E (ApoA, ApoC, and ApoE) have the same genomic structure and are members of a multigene family. In humans, apolipoprotein genes have similar structures, including 3′- and 5′-untranslated regions, 3–4 introns, and 2–3 exons. The exons encode signal peptides, propeptides and mature peptides. Apolipoproteins have a common 33-codon block and variable repeats in the mature peptide region. Thus, their protein domains consist of one 33-amino acid repeat and various 22- and 11-amino acid repeats. Genes for human apolipoproteins belonging to the apolipoprotein A1/A4/E family are located on three different chromosomes. In *Homo sapiens*, the *APOE/C1/C2/C4* gene cluster is located on chromosome 19q13.32. The gene *APOC1* (6.6 kb), together with its downstream pseudogene, the *APOC1'* (6.0 kb) gene, is located in the downstream gene *APOE* (4.7 kb) [25]. The gene *APOC2* (4.7 kb) is located between the gene *APOC1* and gene *APOC4* (4.2 kb) [26–28]. Another gene cluster of human apolipoproteins is the *APOA1/APOC3/APOA4* cluster on chromosome 11q23.3, which spans a range of approximately 43.8 kb [29, 30]. The gene *APOC3* (4.1 kb) is located upstream of *APOA4* and 4735 bp downstream of the *APOA*1 (2.9 kb) gene. The *APOA5* gene (4.0 kb) is located 31.4 kb downstream of the *APOA4* gene (3.4 kb). Only the *APOA2* (1.7 kb) gene is located on chromosome 1q23.3.

A previous study showed that the apolipoprotein LAL1 (lamprey apolipoprotein 1) in *Petromyzon marinus* (lamprey) is similar to human ApoA-I/II/IV, ApoE and ApoC-III [31]. This suggests that the APOA/C/E family and LAL1 likely share common ancestors. Members of the mouse apolipoprotein family are highly conserved with members of the human apolipoprotein family [32, 33]. The apolipoprotein ApoA genes of primates and hedgehogs independently underwent convergent evolution through different duplication and modification events [34, 35]. Another study showed that the evolution of ApoE is highly influenced by feeding habits, and ApoE of frogs as an ancestor is less related to ApoE of other species [36].

As early as 1988, Wen-Hsiung Li et al. summarized the knowledge of the biosynthesis, structure, and structure-function relationships of the APOA/C/E family and proposed a hypothetical scheme for the evolution of this family [27]. Since then, numerous studies have explored the evolution of the APOA/C/E family in different aspects and specific lineages. However, systematic

comparisons and analyses of all members of the APOA/C/E family in various species throughout vertebrata are still lacking. The determination of the evolution of each apolipoprotein in vertebrata is important for understanding the implication of the function of the APOA/C/E family and the adaptation of specific species. In recent years, a large amount of genomic data has become available and provides a favorable opportunity to study apolipoproteins in a broad perspective. In this study, we systematically analyzed all members of the APOA/C/E family across vertebrata throughout *Chondrichthyes, Holostei, Teleostei, Amphibia, Sauria (including Aves)* and *Mammalia* via a comparative genomic approach. The analysis revealed the evolutionary relationships and the gain and loss events for apolipoprotein family members ApoA (I, II, IV, V), ApoC (I, II, III, IV), and ApoE and uncovered the connection between the evolution of apolipoprotein family members and their biological function in the process of species formation.

Results

Overall distribution of the apolipoprotein family in vertebrata

After databases were searched and data were filtered, we obtained all genome data and protein sequences of apolipoproteins for 30 species across vertebrata (Fig. 1). The following analysis is based on the data of these 30 species. After redundant and partial sequences were eliminated, we identified 210 protein sequences belonging to ApoA (I, II, IV, V), ApoC (I~IV), and ApoE in these species by BLASTP as described in the Methods section. The overall distribution of the nine apolipoproteins in vertebrata was greatly varied. The apolipoproteins ApoA-I and ApoA-IV are present in all 30 species. The apolipoprotein ApoC-III exists extensively in *Amniota*. However, ApoC-I, ApoC-II and ApoE are absent in *Aves*. The apolipoproteins ApoC-I and ApoC-IA are not widely distributed and are present only in specific clades. For example, ApoC-IA is identified only in nonhuman *Primates*.

Divergence and convergence between species evolution and apolipoprotein phylogeny

During the process of species evolution, the gene duplication of apolipoproteins resulted in many orthologous and paralogous genes. For each apolipoprotein family member, we gathered the sequences of its orthologous genes to construct phylogenic trees. According to the trees (Fig. 2), each apolipoprotein generally clustered into a different clade (*Chondrichthyes, Holostei, Teleostei, Amphibia, Sauria, Prototheria (Ornithorhynchus), Metatheria (Monodelphis), Laurasiatheria (Perissodactyla, Cetartiodactyla,* and *Carnivora),* and *Euarchontoglires (Primates* and *Glires)),* which is consistent with the pattern of species

Species Name	Genome Assembly & Annotation	ApoA-I	ApoA-II	ApoA-IV	ApoA-V	ApoC-I	ApoC-IA	ApoC-II	ApoC-III	ApoC-IV	ApoE
Homo sapiens	Chromosome	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓
Pan troglodytes	Chromosome	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gorilla gorilla	Chromosome	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pongo abelii	Chromosome	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Macaca mulatta	Chromosome	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓
Papio anubis	Chromosome	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓
Mus musculus	Chromosome	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓
Rattus norvegicus	Chromosome	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓
Oryctolagus cuniculus	Chromosome	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓
Bos taurus	Chromosome	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓
Ovis aries	Chromosome	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓
Capra hircus	Chromosome	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓
Sus scrofa	Chromosome	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓
Equus caballus	Chromosome	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓
Felis catus	Chromosome	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓
Canis lupus familiaris	Chromosome	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓
Monodelphis domestica	Chromosome	✓	✓	✓	✓	✓	✗	✓	✓	✗	✓
Ornithorhynchus anatinus	Chromosome	✓	✓	✓	✗	✗	✗	✓	✓	✗	✓
Gallus gallus	Chromosome	✓	✓	✓	✓	✗	✗	✗	✓	✗	✗
Meleagris gallopavo	Chromosome	✓	✗	✓	✓	✗	✗	✗	✓	✗	✗
Anolis carolinensis	Chromosome	✗	✗	✓	✗	✓	✗	✓	✓	✗	✓
Chrysemys picta	Chromosome	✓	✓	✓	✓	✓	✗	✗	✓	✗	✓
Xenopus Silurana tropicalis	Chromosome	✓	✓	✓	✓	✓	✗	✓	✗	✗	✓
Takifugu rubripes	Chromosome	✓	✗	✓	✗	✗	✗	✓	✗	✗	✓
Oryzias latipes	Chromosome	✓	✗	✓	✗	✓	✗	✗	✗	✗	✓
Danio rerio	Chromosome	✓	✓	✓	✗	✓	✗	✓	✗	✗	✓
Lepisosteus oculatus	Chromosome	✓	✓	✓	✗	✗	✗	✗	✗	✗	✓
Latimeria chalumnae	Scaffold	✓	✗	✓	✗	✗	✗	✗	✗	✗	✓
Rhincodon typus	Scaffold	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗
Callorhinchus milii	Scaffold	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗

Fig. 1 The overall distribution of the apolipoprotein family in vertebrata. A total of 10 members of the apolipoprotein family were analyzed in 30 species. The genome assembly and annotation of the top 26 species are at the chromosome level, and the other 4 species are at the scaffold level. The red check and the black 'x' indicate the presence and absence of an apolipoprotein in a specific species, respectively

evolution. However, some apolipoproteins had phylogenic divergences with the evolutionary progress of species.

In the maximum-likelihood (ML) tree of the ApoA-II protein sequences, the *Glires* clade clustered with *Eutheria* (even with a very low bootstrap value) (Fig. 2b) rather than clustering with *Primates* in the *Euarchontoglires* clade. Except for ApoA-I (Fig. 2a), *Glires* consistently clustered as a peripheral clade instead of clustering with the *Primates* clade. Moreover, in ApoE ML trees (Fig. 2i), *Oryctolagus cuniculus* (rabbit) had low homology with *Mus* and *Rattus*, although rabbit and mouse both belong to *Glires*. The special evolutionary status of the *Glires* apolipoprotein may reflect its functional divergence from its orthologous proteins.

In addition to the unique *Glires* clade, apolipoprotein in platypus (*Ornithorhynchus*) and opossum (*Monodelphis*), which evolved as early mammals, also had phylogenic divergences compared with species evolution. For example, in the ApoA-I ML tree (Fig. 2a), opossum clustered at the periphery of *Sauria*.

In addition, it is noteworthy that the branch length of different clades in the trees of different apolipoproteins varied greatly. This suggests that the evolutionary rate of this family varied depending on species specificity and living environment. For example, the branch length of the *Eutheria* clade is shorter than that of the other branches in the ApoA-I and ApoE trees.

Frequent loss of members in the evolution of the apolipoprotein family in vertebrata

To study the gain and loss events in the evolution of apolipoproteins, we evaluated the presence and absence of these proteins for each species (Fig. 1) in the context of species evolution. Although the functions of these proteins are conserved and, importantly, were revealed in a previous study, we found multiple gain and loss events in the evolution of apolipoproteins (Fig. 3). As indicated by the marks on the tree, there are eight gain events of apolipoproteins. The earliest members that appeared in *Chondrichthyes* (shark) were ApoA-I and
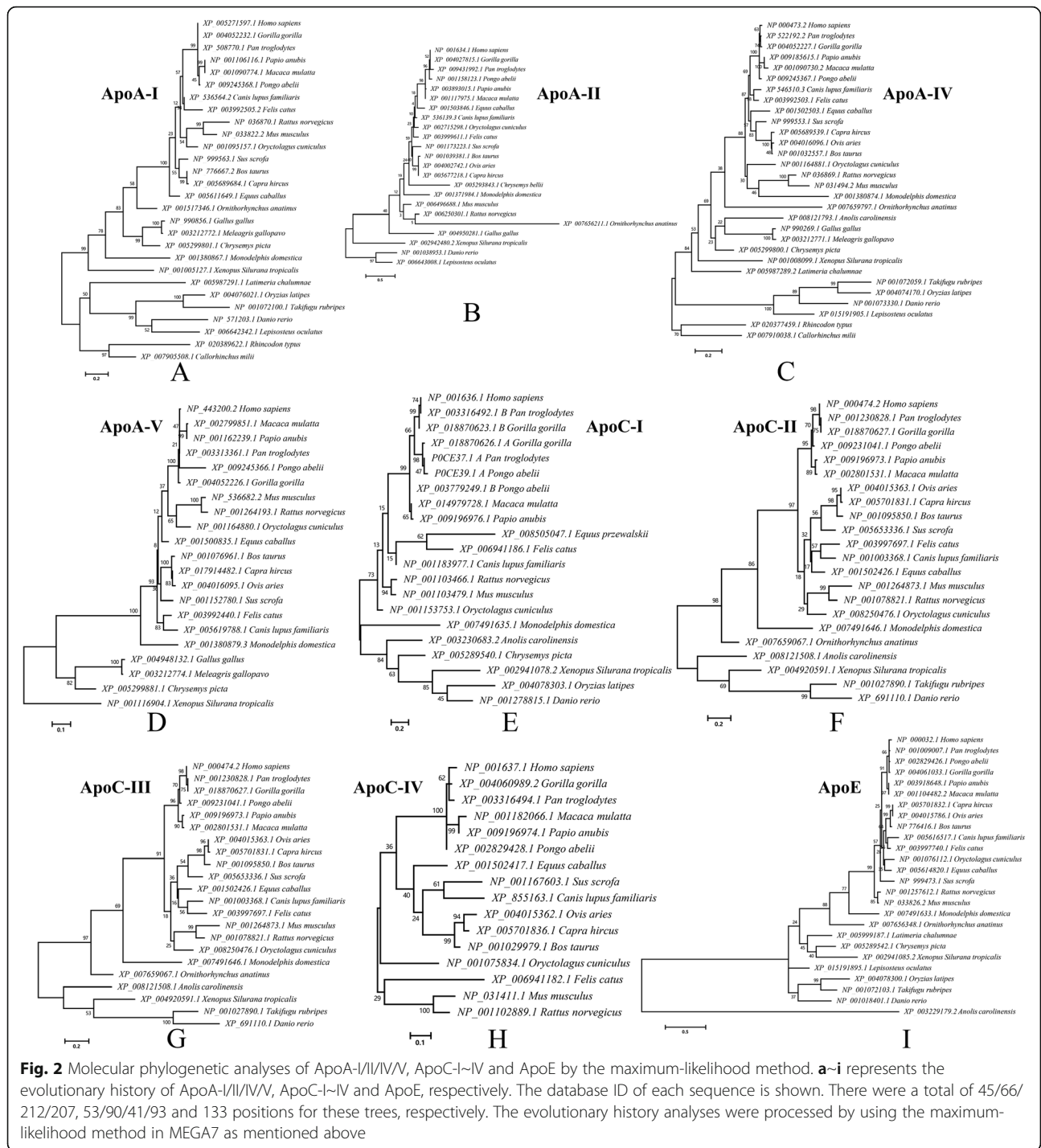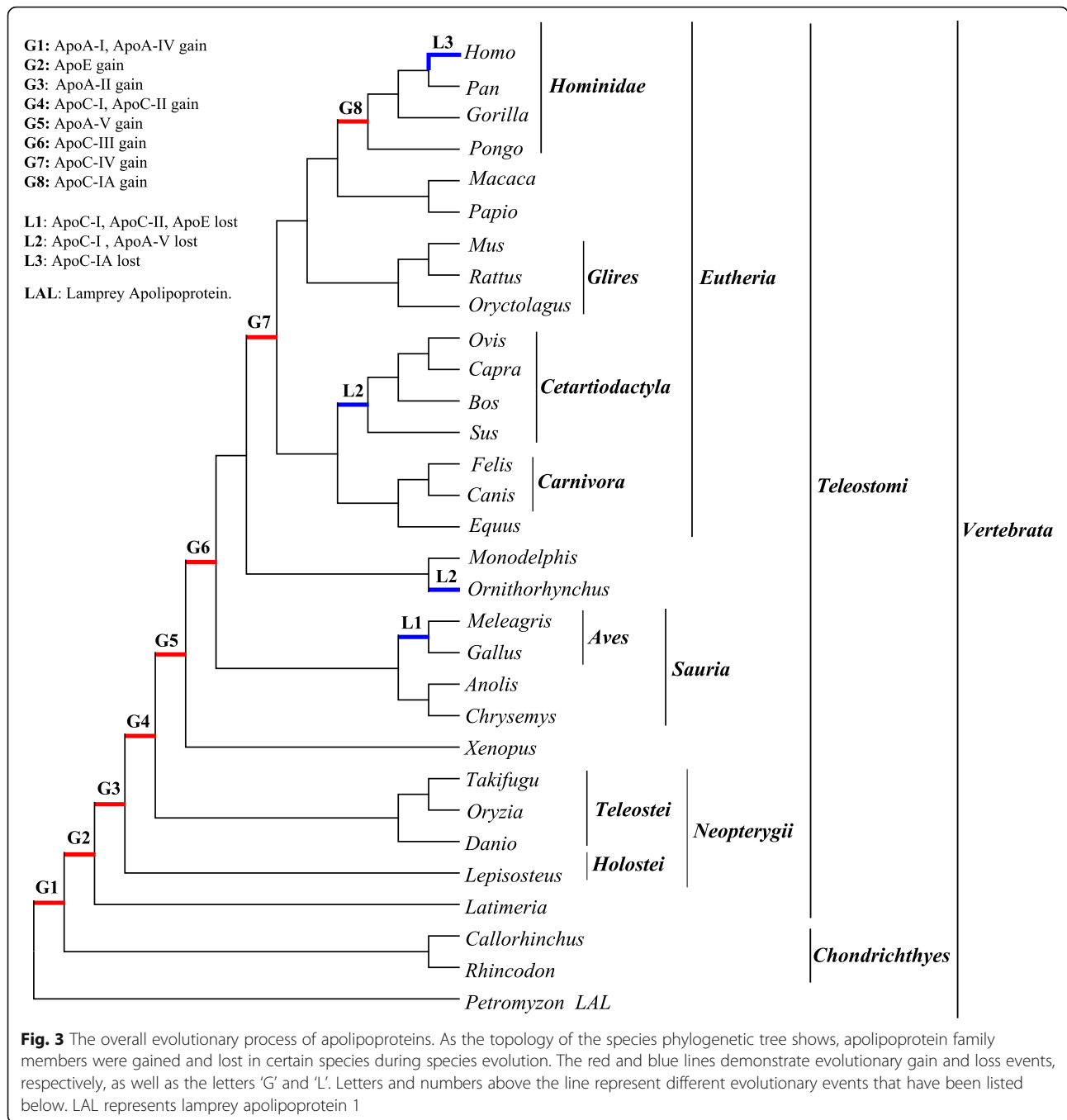
Fig. 2 Molecular phylogenetic analyses of ApoA-I/II/IV/V, ApoC-I~IV and ApoE by the maximum-likelihood method. **a~i** represents the evolutionary history of ApoA-I/II/IV/V, ApoC-I~IV and ApoE, respectively. The database ID of each sequence is shown. There were a total of 45/66/212/207, 53/90/41/93 and 133 positions for these trees, respectively. The evolutionary history analyses were processed by using the maximum-likelihood method in MEGA7 as mentioned above

ApoA-IV, and ApoE and ApoA-II were originally present in the ancient fish lineages spot gar and coelacanth, respectively. Then, ApoC-I and ApoC-II emerged in *Teleostei*. Subsequently, ApoA-V, ApoC-III and ApoC-IV emerged in *Amphibia*, *Sauria* and *Eutheria*, respectively. In addition, ApoC-I derived its acidic form, ApoC-IA, distinctly from the original basic form in *Hominidae*.

However, at the branch of *Homo*, ApoC-IA was lost, and the coding gene of ApoC-IA became a pseudogene.
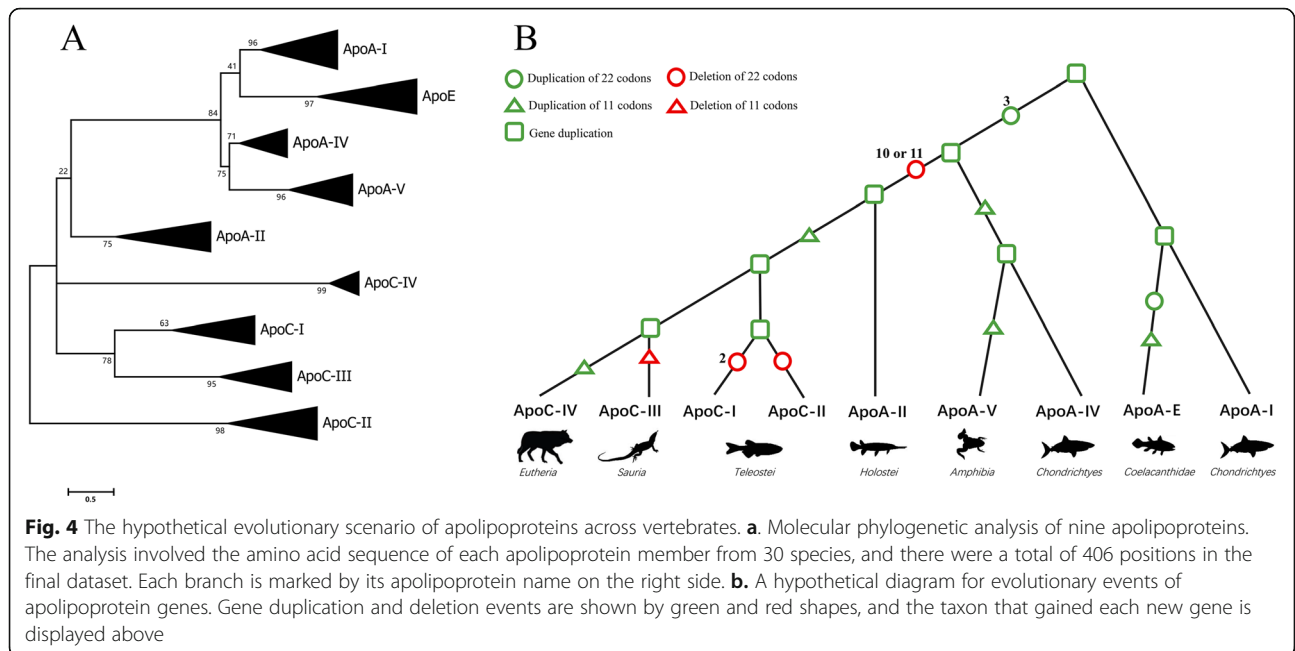
During species evolution, ApoC-I, ApoC-II, and ApoE were lost in *Aves* (L1 event); subsequently, ApoC-I was lost in *Cetartiodactyla* (L2 event). ApoA-V appears in *Amphibia* but disappears together with ApoC-I in platypus (*Ornithorhynchus*). Although ApoC-III and ApoC-

Fig. 3 The overall evolutionary process of apolipoproteins. As the topology of the species phylogenetic tree shows, apolipoprotein family members were gained and lost in certain species during species evolution. The red and blue lines demonstrate evolutionary gain and loss events, respectively, as well as the letters 'G' and 'L'. Letters and numbers above the line represent different evolutionary events that have been listed below. LAL represents lamprey apolipoprotein 1

IV appeared in *Sauria* and *Eutheria,* respectively (G3 and G4 event), and later than other apolipoproteins, they remain stable in the process of species evolution. These apolipoproteins emerged in diverse speciation events and were lost at different times. This result means that apolipoprotein family members did not all simultaneously occur at the beginning of the species and were not always preserved during evolution.

Evolutionary relationships among the apolipoproteins ApoA, ApoC and ApoE

A comprehensive ApoC/C/E unrooted phylogenetic tree including 210 sequences from 30 species was constructed and analyzed (Fig. 4a). As shown in the tree, each branch consists of sequences from all species in which the apolipoprotein exists and clusters well together. The tree also indicates that ApoA-I, ApoE, ApoA-IV and ApoA-V

Fig. 4 The hypothetical evolutionary scenario of apolipoproteins across vertebrates. **a**. Molecular phylogenetic analysis of nine apolipoproteins. The analysis involved the amino acid sequence of each apolipoprotein member from 30 species, and there were a total of 406 positions in the final dataset. Each branch is marked by its apolipoprotein name on the right side. **b**. A hypothetical diagram for evolutionary events of apolipoprotein genes. Gene duplication and deletion events are shown by green and red shapes, and the taxon that gained each new gene is displayed above

cluster into one clade with high bootstrap values. Moreover, the former two and the latter two branches cluster with each other. The other members, ApoC-I~IV and ApoA-II, cluster as periphery clades, and ApoC-I is clustered closer to ApoC-III than other apolipoproteins. The phylogenetic tree reflects the relationship of apolipoproteins at the protein level.

The ApoA-I, A-II, A-V, ApoC-I~III and ApoE genes have three introns separating four exons, but the ApoA-IV and ApoC-IV genes have two introns and three exons. These genes also partly share similar repeat patterns and have a common block of 33 codons in the third exon. According to these structural characteristics, an evolutionary schematic diagram of apolipoproteins at the gene level has been proposed (Fig. 4b). The structure and length of the ancestral apolipoprotein gene are similar to ApoA-I or ApoA-IV; this primordial gene duplicated in *Chondrichthyes*, with one lineage becoming ApoA-I and one lineage becoming ApoA-IV after losing the first intron, gaining three duplications of 22 codons and one duplication of 11 codons in the fourth exon. Furthermore, in the ApoA-I lineage, the fourth exon obtained one 22-codon duplication and one 11-codon duplication, and then the duplicated gene became the ApoE gene in *Coelacanthidae*. In the ApoA-IV lineage, the gene duplicated into ApoA-V with one 11-codon duplication when *Amphibia* arose. In the latter lineage, substantial changes took place in gene length and gene duplication, and ten or eleven deletions of 11 codons occurred in the fourth exon.

Following duplication, one lineage became ApoA-II in *Holostei*, and the other became the common ancestor of ApoC. In the latter lineage, the duplication of 11 codons in the fourth exon occurred in *Teleostei*, and then the gene was further duplicated into two genes. These two resultant genes deleted one and two 22-codon repeats, leading to ApoC-II and ApoC-I, respectively. In the other lineage, the gene duplicated, with the deletion of 11 codons in the gene in *Sauria* and with the duplication of 11 codons in the gene in *Eutheria,* and became the present ApoC-III and ApoC-IV genes, respectively.

Discussion

Apolipoprotein family members are constantly amplified and changed with the evolution of species and perform important physiological functions in these species. Although Wen-Hsiung Li et al. summarized the structural features and evolution of apolipoprotein in the 1980s, their study was limited to three species (humans, dogs and mice) and few apolipoproteins [29, 31–33]. Because of the lack of genome data for many species at that time, the conclusions of those studies are limited and cannot be applied to all vertebrates. In this research, we presented new insights that are very different from those of previous studies by the integrated analysis of a large amount of genome data of all vertebrates. We collected all ApoA, ApoC, and ApoE family members (ApoA-I/II/ IV/V, ApoC-I~IV and ApoE) from *Chondrichthyes, Holostei, Teleostei, Amphibia, Sauria* (including *Aves*), *Prototheria* (*Ornithorhynchus*), *Metatheria* (*Monodelphis*), *Laurasiatheria* (*Perissodactyla, Cetartiodactyla, and Carnivora), and Euarchontoglires* (*Glires* and *Primates*). Representative species can indicate the dynamic changes in apolipoprotein family members at various

stages and nodes of a vertebrate phylogeny more comprehensively.

These results suggest that the ancestral members of the apolipoprotein are most likely ApoA-I and/or ApoA-IV, and other members emerged subsequently after gene duplication. The order of emergence is roughly ApoA-I/ApoA-IV—ApoE—ApoA-II—ApoC-I/ApoC-II—ApoA-V—ApoC-III—ApoC-IV. This new discovery is quite different from the previous hypotheses of Wen-Hsiung Li et al., who suggested that the evolutionary change in apolipoprotein structure and length was from less structured to more structured and from short to long, respectively, by analyzing only the structural characteristics of human apolipoproteins. However, according to our study, the oldest members of the apolipoprotein family are ApoA-I and ApoA-IV, which appeared in ancient cartilaginous fish and have long amino acid sequences and complex structural compositions.

In the context of the divergence and consistency between species evolution and apolipoprotein phylogeny, almost nine apolipoproteins exhibit the phenomenon in which the clusters of protein sequences from different species are inconsistent with the evolutionary progress of these species. The *Glires* clade of apolipoproteins is usually separate from *Euarchontoglires* or *Primates* and does not cluster well in *Eutheria*. Indeed, through a literature search and analysis, we found that the divergent species in the genera *Mus* and *Rattus* are quite different from other species of *Eutheria* in their life history [37, 38]. In terms of body size, reproduction capacity and longevity, *Mus* and *Rattus* species differ significantly from other species. Compared with other species of *Eutheria*, species of *Mus* and *Rattus* are one-fiftieth in size, exhibit 2 to 3 times the frequency of reproduction, and have one-tenth of the average life-span; their peculiar life history inevitably has resulted in their unique mechanisms of growth and metabolism. Thus, we speculate that this diversity may lead to a marked distinction in the metabolism and transport function of lipids between *Glires* and *Primates*. This finding may explain the main reason for the effectiveness of drugs related to lipid metabolism in mouse or rat models but the ineffectiveness of such drugs in humans.

The functions of apolipoproteins not only include transferring lipids, regulating lipoprotein metabolism, and acting as receptor ligands and enzyme cofactors but also involves immune responses related to the pathogenicity factor lipopolysaccharide (LPS) [39, 40]. ApoA-I and ApoA-II, which are present in *Teleostei*, also have a crucial function in antibiosis [41, 42]. ApoC-I and ApoE are differentially expressed after bacterial infection in fish [43], and ApoA-IV is associated with food intake in zebrafish [44]. In addition, the apolipoproteins Apo-II and Apo-IV have been found to be involved in estrogen regulation and egg production [45, 46], which do not exist in *Mammalia*.

These apolipoproteins, which are primitively present in *Chondrichthyes*, have important physiological roles, and their functions have also been preserved in some species during evolution. However, our results also show that ApoC-I, ApoC-II and ApoE are absent or were lost from *Aves*. Thus, the reserved exchangeable apolipoproteins may be responsible for the lipid metabolism and fat storage needed for migration. For example, ApoA-I can bind cholesterol, high-density lipoprotein (HDL) particle receptors and phospholipids and participate in the reverse transport of cholesterol from tissues to the liver by promoting cholesterol efflux from tissues and by acting as a cofactor for the lecithin cholesterol acyltransferase (LCAT). ApoA-IV can also bind copper ions and has protein-homodimerization activity. ApoA-V can bind heparin, lipase, low-density lipoprotein (LDL) particle receptor and phosphatidylcholine.

Furthermore, some functions of apolipoprotein are unique, and once they are lost, corresponding phenotypes will appear. For example, ApoE is related to the formation of bones and appeared in bony fish, not in cartilaginous fish [47, 48]. ApoE may provide a basis for bone structure changes in the evolution of cartilaginous fish to bony fish. However, ApoE is absent in birds and may be associated with specific phenotypes, such as hollow bones without bone marrow and slender ductile bone walls. ApoC-III is the essential component of triglyceride-rich very low-density lipoproteins (VLDLs) and HDLs in plasma. ApoC-III plays a multifaceted role in triglyceride homeostasis, promotes hepatic very low-density lipoprotein 1 (VLDL1) assembly and secretion, and attenuates the hydrolysis and clearance of triglyceride-rich lipoproteins (TRLs) [49]. Animals, from aquatic to terrestrial, consume more energy to overcome changing temperatures and living conditions. ApoC-III may promote the accumulation of triglycerides and store more energy to fight hunger, but at the same time, it is also prone to inducing obesity and cardiovascular disease when food is adequate.

It is believed that relevant evidence for the absence or presence of other members can also be found, which is also the focus and challenge of future studies.

We noticed that the sequence of apolipoprotein family members evolved from long (ApoA and ApoE) to short (ApoA-II, ApoC-I, ApoC-II, ApoC-III). Long apolipoproteins mainly play a role of lipids binding and transport. Short apolipoproteins have multiple regulatory functions such as lipase inhibitor activity (ApoA-II, ApoC-I, ApoC-III), lipase activator activity (ApoC-II), signaling receptor binding (ApoA-II), phospholipase activator activity(ApoC-II), phosphatidylcholine-sterol O-acyltransferase activator activity (ApoC-I), heat shock protein binding (ApoA-II).

Liu *et al. BMC Evolutionary Biology*   (2019) 19:209

Page 8 of 10

Therefore, based the function annotation and the sequence shortening events, we propose a hypothesis that besides the functions of lipids binding and transport, short apolipoproteins may evolve new regulatory functions and contribute to more complex and flexible lipids metabolism in vertebrate species.

Conclusions

To summarize, we reported a lot of gain and loss events of the exchangeable apolipoproteins from *Chondrichthyes* to *Eutheria*, which hasn't previously been reported. New members occurred at the nodes of speciation. Following ApoA-I and ApoA-IV, ApoE arose in Coelacanthiformes; ApoA-II arose in Holostei; ApoC-I arose in Teleostei with ApoC-II; ApoC-III and ApoC-IV emerged in Sauria and Eutheria, respectively; And ApoC-IA emerged in Hominidae. Furthermore, members also lost in the specific nodes of a vertebrate phylogeny. ApoC-I, ApoC-II and ApoE were absent from Aves; ApoC-I/ApoA-V was absent from Ornithorhynchus and Cetartiodactyla; And ApoC-IA lost in human. This is the first clarification of these gain and loss events, which will benefit the research in related fields. Additionally, we also noticed that the sequence of apolipoprotein family members evolved from long (ApoA and ApoE) to short (ApoA-II and ApoC), according to the species in which they primordially existed. Short apolipoproteins ApoA-II and ApoC may evolve new regulatory functions besides the functions of lipids binding and transport. Further data mining and functional research will focus on the function and evolution of major apolipoproteins related to human disease.

Methods

Collection of genome data and sequence retrieval

To study the evolution of the apolipoprotein family across vertebrata, we obtained genome information from the Ensembl database (http://www.ensembl.org/index.html). After the level of genome assembly and annotation were evaluated, only the species with the level 'Chromosome' were used for further analysis (Fig. 1). We downloaded the genome and protein sequences of these species. All members of the apolipoprotein family in *Homo sapiens* were retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/). These human apolipoprotein sequences (ApoA-I:NP_000030.1, ApoA-II:NP_001634.1, ApoA-IV:NP_000473.2, ApoA-V:NP_443200.2, ApoC-I: NP_001636.1, ApoC-II:NP_000474.2, ApoC-III: NP_000031.1, ApoC-IV:NP_001637.1, and ApoE:NP_000032.1) were used as query sequences against all protein sequence data using both local and online BLASTP searches under the default parameters. The details for sequence retrieval were described in our previous study [50]. To confirm the loss of genes in specific species, we first used these protein sequences to search for genome sequences by local tblastn. Then, we used these apolipoproteins to search for all available nucleotide databases in NCBI by online tblastn.

Phylogenetic analysis of the apolipoprotein family

Phylogenetic analysis of the apolipoprotein family was performed using MEGA7 [51]. The amino acid or nucleotide sequences were multiply aligned using MUSCLE [52] with the default parameters. Alignment gaps and unmatched regions were eliminated manually. Phylogenetic trees were constructed by using the maximum-likelihood (ML) method. The ML method is based on the Jones-Taylor-Thornton (JTT) matrix-based mode or Poisson model [53]. The phylogenetic test was performed using the bootstrap method with 1000 bootstrap replications. The ML heuristic method was performed with the nearest-neighbor-interchange (NNI). Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model and selecting topology with superior log likelihood value. The tree was drawn to scale with branch lengths measured in the number of substitutions per site. Positions including gaps and missing data were completely deleted.

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

Author details
[1]School of Life Sciences, Zhengzhou University, Zhengzhou 450001, China. [2]Key Laboratory of Animal Models and Human Disease Mechanisms, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China. [3]Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming 650204, China.

References
1. Phillips MC. Thematic review series: high density lipoprotein structure, function, and metabolism new insights into the determination of HDL structure by apolipoproteins. J Lipid Res. 2013;54(8):2034–48.
2. Utermann G, Beisiegel U. Apolipoprotein A-IV: a protein occurring in human mesenteric lymph chylomicrons and free in plasma. Isolation and quantification. Eur J Biochem. 1979;99(2):333–43.
3. Tabet F, Rye KA. High-density lipoproteins, inflammation and oxidative stress. Clin Sci (Lond). 2009;116(2):87–98.
4. O'Brien PJ, Alborn WE, Sloan JH, Ulmer M, Boodhoo A, Knierman MD, Schultze AE, Konrad RJ. The novel apolipoprotein A5 is present in human serum, is associated with VLDL, HDL, and chylomicrons, and circulates at very low concentrations compared with other apolipoproteins. Clin Chem. 2005;51(2):351–9.
5. Ishihara M, Kujiraoka T, Iwasaki T, Nagano M, Takano M, Ishii J, Tsuji M, Ide H, Miller IP, Miller NE, et al. A sandwich enzyme-linked immunosorbent assay for human plasma apolipoprotein A-V concentration. J Lipid Res. 2005;46(9):2015–22.
6. Jong MC, Hofker MH, Havekes LM. Role of ApoCs in lipoprotein metabolism - functional differences between ApoC1, ApoC2, and ApoC3. Arterioscler Thromb Vasc Biol. 1999;19(3):472–84.
7. Strittmatter WJ, Hill CB. Molecular biology of apolipoprotein E. Curr Opin Lipidol. 2002;13(2):119–23.
8. Mahley RW, Innerarity TL, Rall SC Jr, Weisgraber KH. Plasma lipoproteins: apolipoprotein structure and function. J Lipid Res. 1984;25(12):1277–94.
9. Liu M, Subbaiah PV. Activatioprostate cancer using magnetic resonance imagingn of plasma lysolecithin acyltransferase reaction by apolipoproteins A-I, C-I and E. Biochim Biophys Acta. 1993;1168(2):144–52.
10. Swaney JB, Weisgraber KH. Effect of apolipoprotein C-I peptides on the apolipoprotein E content and receptor-binding properties of beta-migrating very low density lipoproteins. J Lipid Res. 1994;35(1):134–42.
11. Windler E, Havel RJ. Inhibitory effects of C apolipoproteins from rats and humans on the uptake of triglyceride-rich lipoproteins and their remnants by the perfused rat liver. J Lipid Res. 1985;26(5):556–65.
12. Ooi EM, Chan DC, Hodson L, Adiels M, Boren J, Karpe F, Fielding BA, Watts GF, Barrett PH. Triglyceride-rich lipoprotein metabolism in women: roles of apoC-II and apoC-III. Eur J Clin Investig. 2016;46(8):730–6.
13. Wolska A, Dunbar RL, Freeman LA, Ueda M, Amar MJ, Sviridov DO, Remaley AT. Apolipoprotein C-II: new findings related to genetics, biochemistry, and role in triglyceride metabolism. Atherosclerosis. 2017;267:49–60.
14. Onat A, Hergenc G, Ayhan E, Ugur M, Kaya H, Tuncer M, Can G. Serum apolipoprotein C-III in high-density lipoprotein: a key diabetogenic risk factor in Turks. Diabet Med. 2009;26(10):981–8.
15. Cohn JS, Tremblay M, Batal R, Jacques H, Rodriguez C, Steiner G, Mamer O, Davignon J. Increased apoC-III production is a characteristic feature of patients with hypertriglyceridemia. Atherosclerosis. 2004;177(1):137–45.
16. Goldstein JL, Brown MS. The LDL receptor defect in familial hypercholesterolemia. Implications for pathogenesis and therapy. Med Clin North Am. 1982;66(2):335–62.
17. Munoz SS, Garner B, Ooi L. Understanding the role of ApoE fragments in Alzheimer's disease. Neurochem Res. 2018.
18. Huebbe P, Rimbach G. Evolution of human apolipoprotein E (APOE) isoforms: gene structure, protein function and interaction with dietary factors. Ageing Res Rev. 2017;37:146–61.
19. Kulminski AM, Huang J, Wang J, He L, Loika Y, Culminskaya I. Apolipoprotein E region molecular signatures of Alzheimer's disease. Aging Cell. 2018;17: e12779.
20. Laws SM, Hone E, Gandy S, Martins RN. Expanding the association between the APOE gene and the risk of Alzheimer's disease: possible roles for APOE promoter polymorphisms and alterations in APOE transcription. J Neurochem. 2003;84(6):1215–36.
21. Raber J, Wong D, Yu GQ, Buttini M, Mahley RW, Pitas RE, Mucke L. Apolipoprotein E and cognitive performance. Nature. 2000;404(6776):352–4.
22. Holtzman DM, Herz J, Bu GJ. Apolipoprotein E and Apolipoprotein E Receptors: Normal Biology and Roles in Alzheimer Disease. Cold Spring Harbor Perspectives in Medicine. 2012:2(3).
23. Strittmatter WJ, Saunders AM, Schmechel D, Pericak-Vance M, Enghild J, Salvesen GS, Roses AD. Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proc Natl Acad Sci U S A. 1993;90(5):1977–81.
24. Rhinn H, Fujita R, Qiang L, Cheng R, Lee JH, Abeliovich A. Integrative genomics identifies APOE epsilon 4 effectors in Alzheimer's disease (retracted article. See vol. 523, 2015). Nature. 2013;500(7460):45–U62.
25. Lauer SJ, Walker D, Elshourbagy NA, Reardon CA, Levy-Wilson B, Taylor JM. Two copies of the human apolipoprotein C-I gene are linked closely to the apolipoprotein E gene. J Biol Chem. 1988;263(15):7277–86.
26. Allan CM, Walker D, Segrest JP, Taylor JM. Identification and characterization of a new human gene (APOC4) in the apolipoprotein E, C-I, and C-II gene locus. Genomics. 1995;28(2):291–300.
27. Wei CF, Tsao YK, Robberson DL, Gotto AM Jr, Brown K, Chan L. The structure of the human apolipoprotein C-II gene. Electron microscopic analysis of RNA:DNA hybrids, complete nucleotide sequence, and identification of 5′ homologous sequences among apolipoprotein genes. J Biol Chem. 1985;260(28):15211–21.
28. Myklebost O, Williamson B, Markham AF, Myklebost SR, Rogers J, Woods DE, Humphries SE. The isolation and characterization of cDNA clones for human apolipoprotein CII. J Biol Chem. 1984;259(7):4401–4.
29. Karathanasis SK. Apolipoprotein multigene family: tandem organization of human apolipoprotein AI, CIII, and AIV genes. Proc Natl Acad Sci U S A. 1985;82(19):6374–8.
30. Cheung P, Kao FT, Law ML, Jones C, Puck TT, Chan L. Localization of the structural gene for human apolipoprotein A-I on the long arm of human chromosome 11. Proc Natl Acad Sci U S A. 1984;81(2):508–11.
31. Li WH, Tanimura M, Luo CC, Datta S, Chan L. The Apolipoprotein multigene family - biosynthesis, structure, structure-function relationships, and evolution. J Lipid Res. 1988;29(3):245–71.
32. Luo CC, Li WH, Moore MN, Chan L. Structure and evolution of the apolipoprotein multigene family. J Mol Biol. 1986;187(3):325–40.
33. Boguski MS, Birkenmeier EH, Elshourbagy NA, Taylor JM, Gordon JI. Evolution of the apolipoproteins. Structure of the rat apo-A-IV gene and its relationship to the human genes for apo-A-I, C-III, and E. J Biol Chem. 1986; 261(14):6398–407.
34. Lawn RM, Boonmark NW, Schwartz K, Lindahl GE, Wade DP, Byrne CD, Fong KJ, Meer K, Patthy L. The recurring evolution of lipoprotein(a). Insights from cloning of hedgehog apolipoprotein(a). J Biol Chem. 1995;270(41):24004–9.
35. Lawn RM, Schwartz K, Patthy L. Convergent evolution of apolipoprotein(a) in primates and hedgehog. Proc Natl Acad Sci U S A. 1997;94(22):11992–7.
36. Kasap M, Sazci A, Akpinar G, Ergul E. Apolipoprotein E phylogeny and evolution. Cell Biochem Funct. 2008;26(1):43–50.
37. Calder WA 3rd. Body size, mortality, and longevity. J Theor Biol. 1983;102(1):135–44.
38. Ramsey JJ, Tran D, Giorgio M, Griffey SM, Koehne A, Laing ST, Taylor SL, Kim K, Cortopassi GA, Lloyd KC, et al. The influence of Shc proteins on life span in mice. J Gerontol A Biol Sci Med Sci. 2014;69(10):1177–85.
39. Henning MF, Herlax V, Bakas L. Contribution of the C-terminal end of apolipoprotein AI to neutralization of lipopolysaccharide endotoxic effect. Innate Immunity. 2011;17(3):327–37.
40. Berbee JFP, Coomans CP, Westerterp M, Romijn JA, Havekes LM, Rensen PCN. Apolipoprotein CI enhances the biological response to LPS via the CD14/TLR4 pathway by LPS-binding elements in both its N- and C-terminal helix. J Lipid Res. 2010;51(7):1943–52.
41. Dietrich MA, Adamek M, Bilinska B, Hejmej A, Steinhagen D, Ciereszko A. Characterization, expression and antibacterial properties of apolipoproteins a from carp (Cyprinus carpio L.) seminal plasma. Fish & Shellfish Immunology. 2014;41(2):389–401.

42. Dietrich MA, Nynca J, Adamek M, Steinhagen D, Karol H, Ciereszko A. Expression of apolipoprotein A-I and A-II in rainbow trout reproductive tract and their possible role in antibacterial defence. Fish & Shellfish Immunology. 2015;45(2):750–6.
43. Yang YJ, Fu Q, Zhou T, Li Y, Liu SK, Zeng QF, Wang XZ, Jin YL, Tian CX, Qin ZK, et al. Analysis of apolipoprotein genes and their involvement in disease response of channel catfish after bacterial infection. Dev Comp Immunol. 2017;67:464–70.
44. Otis JP, Zeituni EM, Thierer JH, Anderson JL, Brown AC, Boehm ED, Cerchione DM, Ceasrine AM, Avraham-Davidi I, Tempelhof H, et al. Zebrafish as a model for apolipoprotein biology: comprehensive expression analysis and a role for ApoA-IV in regulating food intake. Dis Model Mech. 2015;8(3): 295–309.
45. Nikolay B, Plieschnig JA, Subik D, Schneider JD, Schneider WJ, Hermann M. A novel estrogen-regulated avian apolipoprotein. Biochimie. 2013;95(12): 2445–53.
46. Ratna WN, Bhatt VD, Chaudhary K, Bin Ariff A, Bavadekar SA, Ratna HN. Estrogen-responsive genes encoding egg yolk proteins vitellogenin and apolipoprotein II in chicken are differentially regulated by selective estrogen receptor modulators. Theriogenology. 2016;85(3):376–83.
47. Niemeier A, Schinke T, Heeren J, Amling M. The role of apolipoprotein E in bone metabolism. Bone. 2012;50(2):518–24.
48. Papachristou NI, Blair HC, Kypreos KE, Papachristou DJ. High-density lipoprotein (HDL) metabolism and bone mass. J Endocrinol. 2017;233(2): R95–R107.
49. Luo M, Peng D. The emerging role of apolipoprotein C-III: beyond effects on triglyceride metabolism. Lipids Health Dis. 2016;15(1):184.
50. Dai S-X, Zhang A-D, Huang J-F. Evolution, expansion and expression of the Kunitz/BPTI gene family associated with long-term blood feeding in Ixodes Scapularis. BMC Evol Biol. 2012;12:4.
51. 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.
52. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.
53. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 1992;8(3):275–82.

Publisher's Note