Evolutionary history of mammalian UDP-glucuronosyltransferase (UGT)1 and UGT2 families: the emergence of UGT2B subfamily in eutherians after the diversification of flowering plants

Yusuke K. Kawai[1], Kasumi Sano[2], Yoshinori Ikenaka[2,3,4], Shouta M.M. Nakayama[2], Mitsuki Kondo[2], Akira Kubota[1], Mayumi Ishizuka[2#]

[1] Department of Veterinary Medicine, Obihiro University of Agriculture and Veterinary Medicine, Nishi-2-11, Inada-cho, Obihiro, Hokkaido, 080-8555, Japan
[2] Laboratory of Toxicology, School of Veterinary Medicine, Hokkaido University, Kita-18 Nishi-9, Kita-ku, Sapporo, Hokkaido, 060-0818, Japan
[3] Water Research Group, Unit for Environmental Science and Management, North-West University, Potchefstroom, North West Province, South Africa
[4] Translational Research Unit, Veterinary Teaching Hospital, Faculty of Veterinary Medicine, Hokkaido University, Kita-18 Nishi-9, Kita-ku, Sapporo, Hokkaido, 060-0818, Japan.
[#] Corresponding author

Keywords: UDP-glucuronosyltransferase; UGT; Molecular evolution; Xenobiotic metabolism; feeding habits

Abstract

The UDP-glucuronosyltransferase (UGT) gene family is responsible for the transfer of glucuronic acid to exogenous and endogenous chemicals. Based on the highly diversified number of genes, the mammalian UGT1A and UGT2B subfamily genes are believed to be involved in the conjugation reactions of xenobiotic metabolism. However, it is speculated that the UGT2 family genes are not involved in the xenobiotic metabolism of avian species due to the less diverse number of genes. In this study, we aimed to investigate the evolutionary history of mammalian UGT1 and UGT2 family genes and determine when the diversification of *UGT2B* genes occurred. We also attempted to identify the main factors responsible for the diversification of *UGT* genes. By examining the genomic information and feeding habits of 67 species representing each mammalian family, we discovered that the *UGT2B* genes emerged in the Eutheria on or after Cretaceous period and that their number were higher in plant-eating mammals (herbivore or omnivore) than in carnivorous mammals. We also found that the *UGT2B* genes in some herbivorous mammals underwent positive selection. In contrast, the diversity of the UGT1 family genes was inherited from the common ancestor of birds and mammals. Thus, our findings suggest that the emergence of angiosperms (flowering plants) and the occurrence of “animal–plant warfare” influenced the evolution of this gene family involved in the xenobiotic metabolism of eutherians. Furthermore, future research investigating the marsupials and birds that do not possess *UGT2B* genes is required to elucidate the mechanisms underlying the metabolism of chemical substances in these species.

Introduction

In vertebrates, the uridine diphosphate (UDP)-glycosyltransferase (UGT) superfamily conjugate sugars from UDP-sugar donors to endogenous and exogenous chemicals. Mammalian species possess the UGT1, UGT2, UGT3, and UGT8 families (1). The UGT1 and UGT2 families, which mainly use UDP-glucuronic acid for conjugation, are called glucuronosyltransferases. The UGT1 family is composed of the UGT1A subfamily, while the UGT2 family is divided into two subfamilies, UGT2A and UGT2B (1). Many UGT1A and UGT2B enzymes play an important role in conjugating glucuronic acids in the liver (2), while UGT2As, except for UGT2A3, are responsible for the conjugation of sugars in the olfactory neuroepithelium (3, 4). The *UGT1A* genes are composed of five exons, including the variable 1st exon and the constant 2nd–5th exons. Similarly, in humans and mice, the *UGT2A1* and *UGT2A2* genes consist of the variable 1st exons and constant 2nd–6th exons. In contrast, the *UGT2A3* and *UGT2B* genes are composed of unigenes (1).

Previous studies reported that mammalian *UGT1A* and *UGT2B* are diversified among species and may have evolved due to the influence of feeding habits. In the order Carnivora, *UGT1A6*, which plays an important role in metabolizing phenolic compounds, became a pseudogene in several pure meat-eating species (5). Furthermore, the number of *UGT1A* genes is generally fewer in pinniped mammals than in omnivorous species (6), while the number of *UGT2B* genes is usually lower in carnivores than in omnivores (7). It was also found that the conjugation activities of liver microsomes in carnivorous species are less than those in omnivorous species (6, 7).

The diversification of *UGT2B* genes is believed to be a mammal-specific event. Similar to mammals, the number of *UGT1E* genes and the conjugation activity are low in carnivorous birds. However, unlike in mammals, the number of UGT2 family genes ($n \leq 3$) is generally consistent in many bird species (8, 9). These studies suggest that the mammalian and avian UGT1 and UGT2 family genes are independently diversified and evolved. In particular, the differences between mammalian UGT2B and avian UGT2 family genes imply that the ancestor *UGT2* genes did not play a role in xenobiotic metabolism and that the mammalian *UGT2B* genes gained this novel function after the divergence of mammals and birds. However, the factors that influenced the diversification of mammalian UGT1 and UGT2 family genes and the time period that this occurred are still unclear.

Previous studies have suggested that the number of *UGT* genes is related to animal feeding habits and that “animal–plant warfare” may have caused the evolution of *UGT* genes (10). Hence, we focused on feeding habits in this study. Furthermore, herbivorous and omnivorous animals are exposed to various plants chemicals, and diets are a major source of xenobiotics. It was predicted that each plant species synthesizes 4.7 (=98,800/20,741) unique metabolites on average (11) and that ~374,000 plant species exist on Earth (12). Angiospermae, the class of flowering plants, is the major plant group on Earth that consists of ~300,000 species (12). Many herbivorous and omnivorous species consume flowering plants, which are believed to have emerged and diverged around the Upper Triassic and Cretaceous periods (13). These suggest that the variety of chemicals to which herbivorous and omnivorous animals were exposed increased after the Cretaceous, as flowering plants diversified.

In the present study, we aimed to elucidate the evolutionary history of mammalian UGT1 and UGT2 family genes and verify the hypothesis that the development of flowering plants in the Cretaceous period caused the emergence and divergence of mammalian *UGT2B* genes. We determined the number of UGT1 and UGT2 family genes in mammals based on a public genome database by tracing their evolutionary history, investigating the relationship between the number of genes and feeding habit, and estimating selection pressures.

2. Materials and Methods

2.1 Homology search and genomic localization of *UGT* sequences

To determine the number of mammalian UGT1 and UGT2 family genes, we retrieved the *UGT1* and *UGT2* sequences from 67 species (Table S1) using homology searches. These species represented the 67 mammalian families with assembly information in the NCBI genome project (as of January 2020) (14) and were selected based on assembly quality, such as scaffold N50. For the homology searches, we used query sequences from the UGT1 and UGT2 family genes of seven vertebrates, namely anole lizard (*Anolis carolinensis*), Tropical clawed frog (*Xenopus tropicalis*), zebrafish (*Danio rerio*), chicken (*Gallus gallus*), platypus (*Ornithorhynchus anatinus*), mouse (*Mus muscles*), and human (*Homo sapiens*), that were annotated in Ensembl (release 98) (15) (Data S1). The TBLASTN searches (16, 17) were performed for each species using the GenBank RefSeq database (18), with e-value < 1e-6 as the identity threshold. Following BLAST search, we excluded the genes with distinctly different annotations and retrieved the locus information of the remaining genes from GenBank using reutils package (19) Data visualization was performed using genoPlotR (20) from R version 4.02 (21).

2.2 Phylogenetic analysis of mammalian *UGT* genes

The gene location and BLASTN searches were used to classify *UGT* genes into the UGT1 and UGT2 families. Maximum likelihood-based phylogenetic analysis was performed for the classified *UGT1* and *UGT2* genes, which were divided into the variable 1st exons and other constant exons and separately analyzed for phylogeny construction. Briefly, the amino acid sequences were aligned with MAFFT version 7.2 (22) using the auto option and then trimmed with trimAl using the automated1 option (23). For the model selection and phylogenetic analysis of *UGT1A*, the 1st and 2nd–5th exons with >200 amino acid sequences were used. For *UGT2*, the 1st and 2nd–6th exons with >200 and >250 amino acid sequences, respectively, were analyzed (Data S2). The best-fit model was selected using the Bayes information criterion (BIC) calculated by CodeML on Aminosan (24, 25). Based on the model, the phylogenies of UGT1 and UGT2 families were inferred by maximum likelihood method, with 100 bootstrap replicates, using RAxML version 8.2.10 (26). The models used in this analysis were shown in Table S2.

2.3 Number of *UGT* genes

The number of UGT genes were based on the phylogenetic analysis and gene localization on the chromosome. In this study, the genes were categorized into three types: pseudogenes, partial genes, and functional genes. First, we identified the pseudogenes according to the GenBank annotation. Second, the short sequences that were excluded from the phylogenetic analysis were considered as partial genes. Third, the remaining genes were predicted to be functional genes. Fourth, sequences containing multiple identities in one annotation were considered as multiple genes (Data S3). The mammalian phylogenetic tree was obtained from TimeTree database (27–29). Data visualization incorporating the mammalian phylogenetic tree was performed using ggtree (30–32) from R version 4.02 (21).

2.4 Relationship between the number of genes and feeding habit

The dietary information was obtained from PHYLACINE 1.2.1 (33). The feeding habits were divided into the carnivore group that does not eat plants (0%) and the herbivore or omnivore group that eat plants (10~100%) (Table S1). For phylogenetic correction, mammalian phylogenetic tree in TimeTree (Hedges, Dudley, and Kumar 2006; Hedges et al. 2015; Kumar et al. 2017) was used and Phylogenetic ANOVA was performed to compare the number of genes between the two groups using phythools (34) from R version 4.02 (21).

2.5 Estimating selection pressures

The RELAX program (35) in HyPhy (36) was used to detect relaxed selection on the 1st exons of *UGT1A6* genes in carnivorous species. The adaptive branch-site random effects likelihood (aBSREL) approach (37) in HyPhy was used to detect episodic positive selection on all branches of the 1st exon regions of eutherian or placental mammal (Infraclass Eutheria) *UGT2B* genes or marsupial (Infraclass Marsupialia/Metatheria) *UGT2* genes in herbivorous or omnivorous mammals. The false discovery rate (FDR) method (38) was employed for the correction of the *p*-values from all tests of branches after multiple testing. Sequences and phylogeny used in the analysis were shown in Data S4–S7. The RELAX and aBSREL analysis were performed in the Datamonkey server (39).

Results

3.1 Phylogeny and gene locus of UGT1 and UGT2 family genes

The maximum likelihood-based phylogenetic analysis of the variable 1st exons revealed that the *UGT1A* genes of eutherians and marsupials grouped into 4 major clades, specifically "*UGT1A1*," "*UGT1A2–UGT1A5*," "*UGT1A6*," and "*UGT1A7–UGT1A10*." Four platypus *UGT1A* genes composed one clade with the mammalian "*UGT1A1*" and "*UGT1A2–UGT1A5*" genes, while the remaining platypus *UGT1A* gene clustered with the avian *UGT1E* group V gene (Fig. 1, Fig. S1). Phylogenetic analysis of the constant 2nd–5th exons showed that the eutherian *UGT1A* genes formed one clade, and marsupial and platypus *UGT1A* genes composed one clade with avian *UGT1E* genes (Fig. S2). The locus information of the *UGT1A* genes showed conserved synteny among mammalian species, of which most were located between ubiquitin-specific protease 40 (*USP40*) and Transient receptor potential melastatin 8 (*TRPM8*) (Fig. 2, Fig. S3). In most species, *UGT1A* genes are usually composed of one set of 2nd–5th exons and variable 1st exons, except in the Ord's kangaroo rat (*Dipodomys ordii*), naked mole-rat (Heterocephalus glaber), and Chinese tree shrew (*Tupaia chinensis*) that may have multiple sets of 2nd–5th exons (Fig. S2, Fig. S3). Furthermore, the 2nd–5th *UGT1A* exons do not exist in the big brown bat (*Eptesicus fuscus*), Sunda pangolin (*Manis javanica*), northern greater galago (*Otolemur garnettii*), and black flying fox (*Pteropus alecto*).

Phylogenetic analysis of the constant 2nd–6th exons revealed that the mammalian *UGT2* genes clustered into four major clades, platypus *UGT2*, marsupial *UGT2*, eutherian *UGT2A*, and eutherian *UGT2B* (Fig. 3, Fig. S4). Moreover, the eutherian *UGT2B* clade included two subclades; *UGT2B-I* also known as *UGT2B* subfamily and *UGT2B-II* also known as *UGT2C* or *UGT2E* subfamily. The phylogenetic tree showed that the UGT2B subfamily is a unique eutherian gene group and that the 1st exons of marsupial *UGT2* genes grouped with the eutherian clades for *UGT2A* and *UGT2B-I* genes (Fig. S5). The locus information indicated the conserved synteny of *UGT2* genes among mammalian species, of which most were located between sulfotransferase family 1B member 1 (*SULT1B1*) and YTH domain-containing 1 (*YTHDC1*) (Fig. 2, Fig. S6). Similar to the *UGT1A* in all species, the marsupial *UGT2* and eutherian *UGT2A* genes were composed of variable 1st exons and constant 2nd–6th exons, while eutherian *UGT2B* genes were found to be variable unigenes.

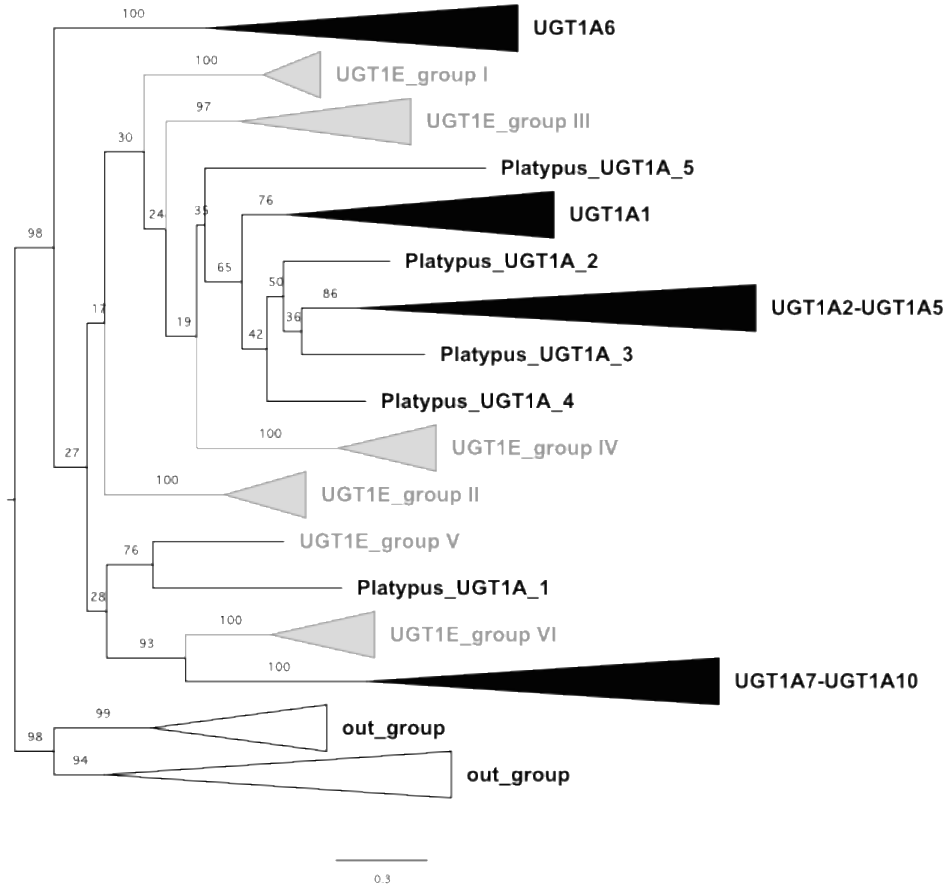

Figure 1. Phylogenetic tree of UDP-glucuronosyltransferase 1 (UGT1) family genes constructed from the amino acid sequence alignments of the 1st exons in mammalian and avian *UGT1* genes. The black and gray triangles indicate the mammalian *UGT1A* and avian *UGT1E* clades, respectively. The white triangles represent the outgroups. The *UGT1E* clade names were based on Kawai et al. (2018). Bootstrap values with 100 replicates are shown next to the branches as percentage. The tree is drawn to scale with branch length indicating the expected number of substitutions per site.

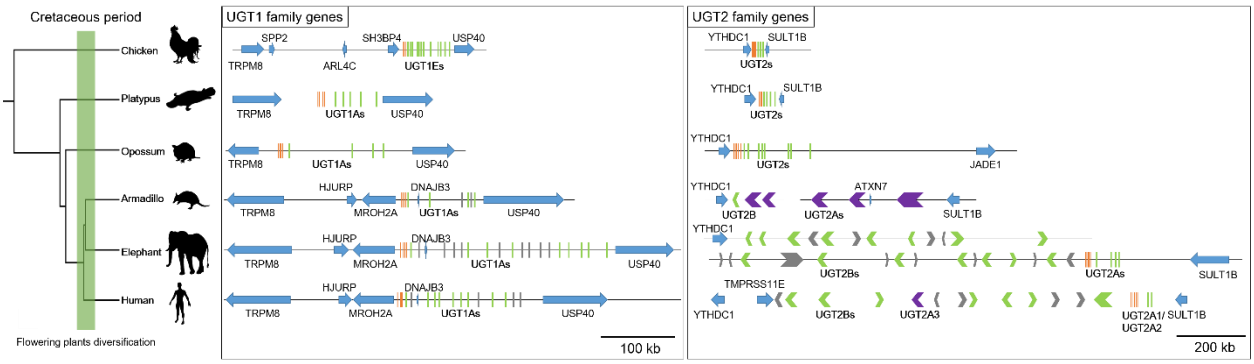

Figure 2. Phylogenetic tree demonstrating the synteny of UGT1 and UGT2 family genes in chicken (*Gallus gallus*), platypus (*Ornithorhynchus anatinus*), opossum (*Monodelphis domestica*), armadillo (*Dasypus novemcinctus*), elephant (*Loxodonta africana*), and human (*Homo sapiens*). *UGT* genes with multiple variable 1st exons and constant 2nd–6th exons are represented by green and orange vertical lines, respectively. The *UGT2B* and *UGT2A* unigenes are represented by green and purple arrowheads, while the other genes located in the same chromosome are represented by blue arrows. The scale bar represents 200 kilobase pairs of nucleotides.

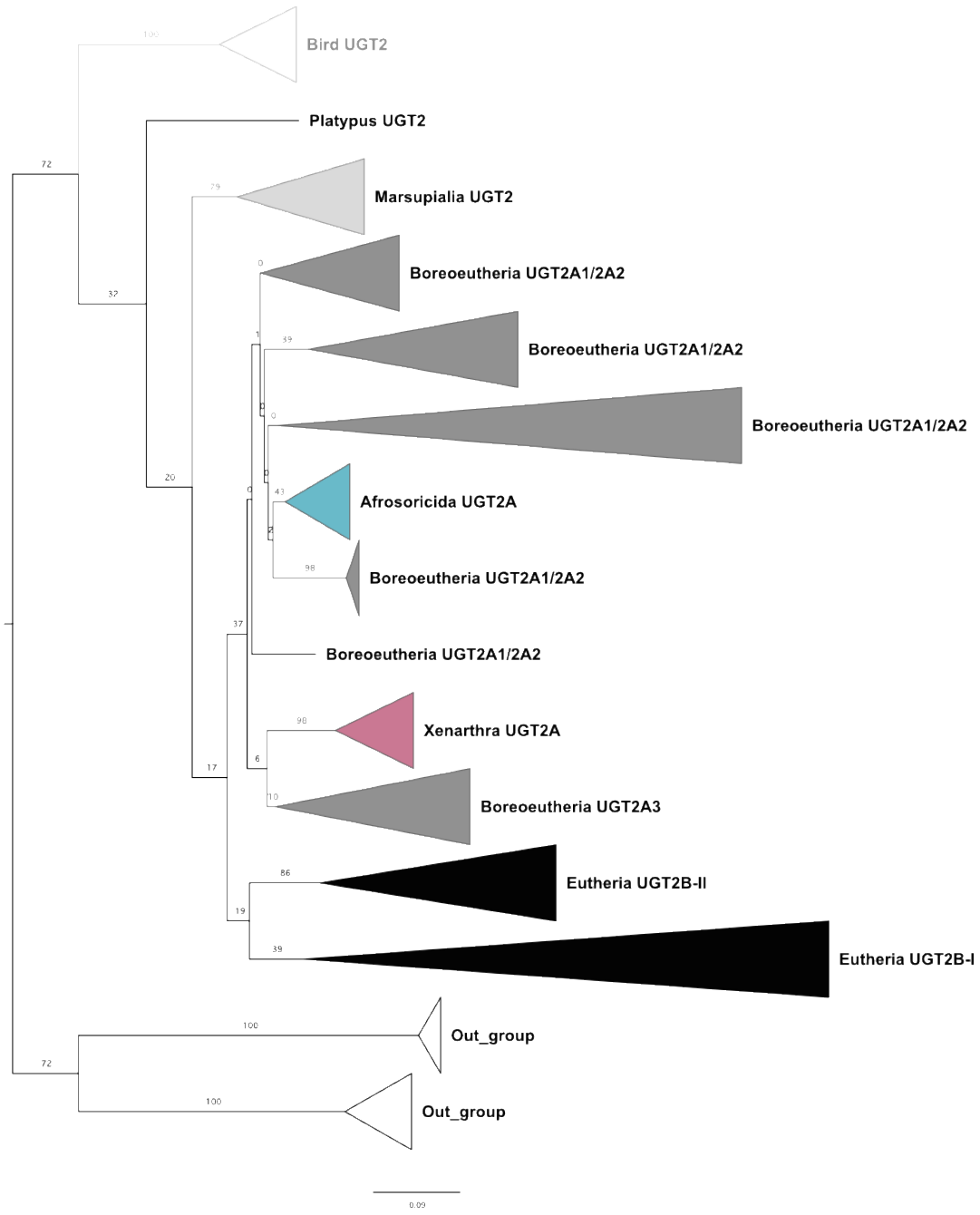

Figure 3. Phylogenetic tree of UDP-glucuronosyltransferase 2 (UGT2) family genes constructed from the amino acid sequence alignments of the 2nd–6th exons in mammalian and avian *UGT2* genes. The white, light gray, dark gray, blue, red, and black triangles indicate the avian (Class Aves) *UGT2*, marsupial (Infraclass Marsupialia) *UGT2*, boreoeutherian (Magnorder Boreoeutheria) *UGT2A*, afrosoricidan (Order Afrosoricida) *UGT2A*, xenarthran (Magnorder Xenarthra) *UGT2A*, and eutherian (Infraclass Eutheria) *UGT2B* clades, respectively. Bootstrap values with 100 replicates are shown next to the branches as percentage. The tree is drawn to scale with branch length indicating the expected number of substitutions per site.

3.2 Number of *UGT* genes

The different numbers of *UGT1A* genes in the studied mammalian species suggest that the number of functional genes varies among species, ranging from one in the minke whale (*Balaenoptera acutorostrata*) to 14 in the Common degu (*Octodon degus*) (Fig. 4). In the eutherian and marsupial *UGT1A* subgroups, the number of functional genes in "*UGT1A1*" and "*UGT1A6*" are relatively consistent, while those in "*UGT1A2–UGT1A5*" and "*UGT1A7–UGT1A10*" are variable (Fig. 5). Furthermore, the number of UGT2 family genes in mammalian species indicates the high variability of functional *UGT2* genes among mammals, which ranged from zero in the Pacific white-sided dolphin (*Lagenorhynchus obliquidens*) to 24 in the European rabbit (*Oryctolagus cuniculus*) (Fig. 4). In Eutheria, the number of *UGT2A* genes was relatively consistent compared to that of *UGT2B* genes (Fig. 6).

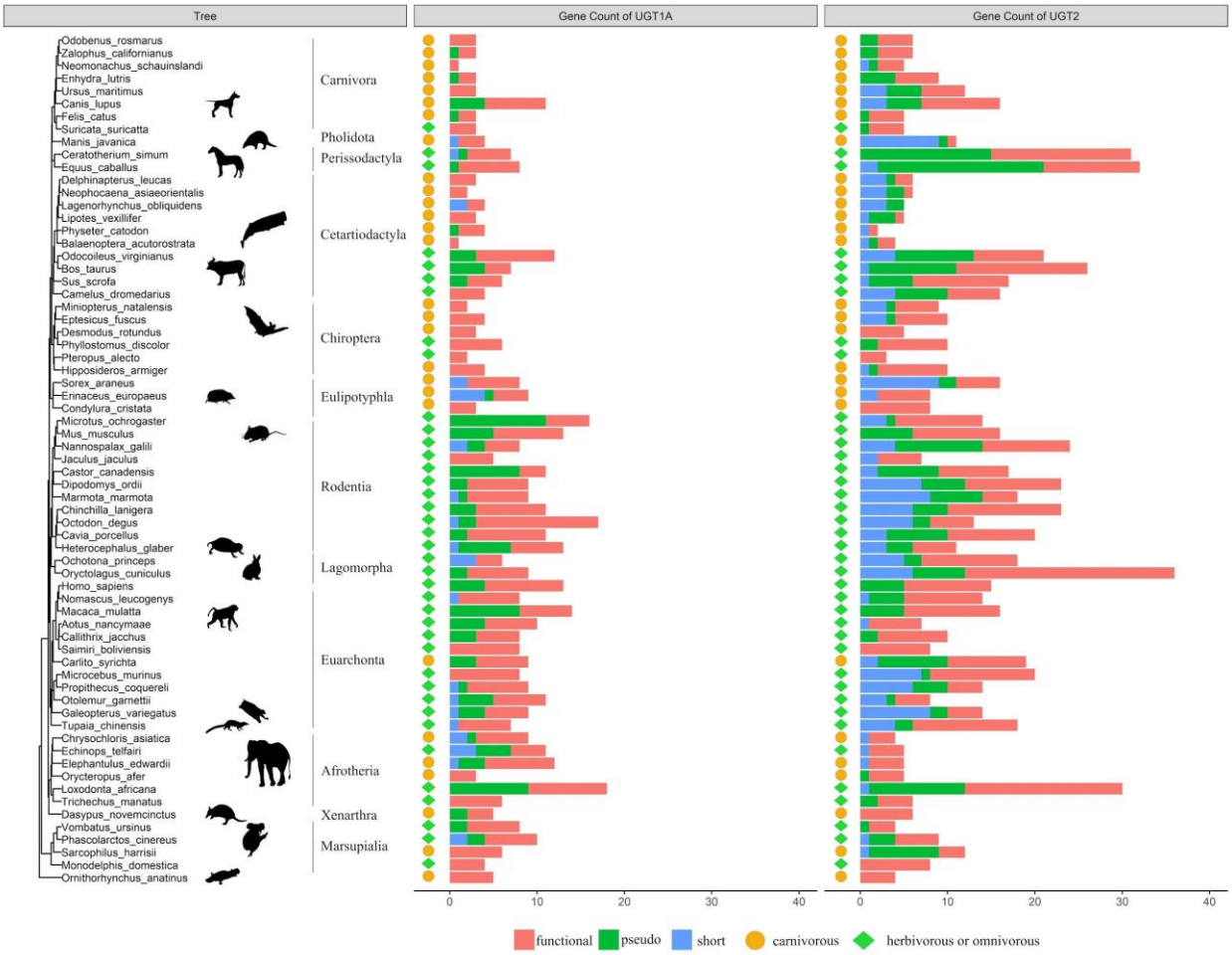

Figure 4. The numbers of UGT1 and UGT2 family genes among mammalian species. The data were based on GenBank annotations. The functional, pseudo-, and short genes (excluded from phylogenetic analysis) are represented by red, green, and blue bars, respectively. The feeding habit is marked on the left of the bar graphs, with orange circles and light green diamonds representing the carnivorous and herbivorous or omnivorous animals, respectively.

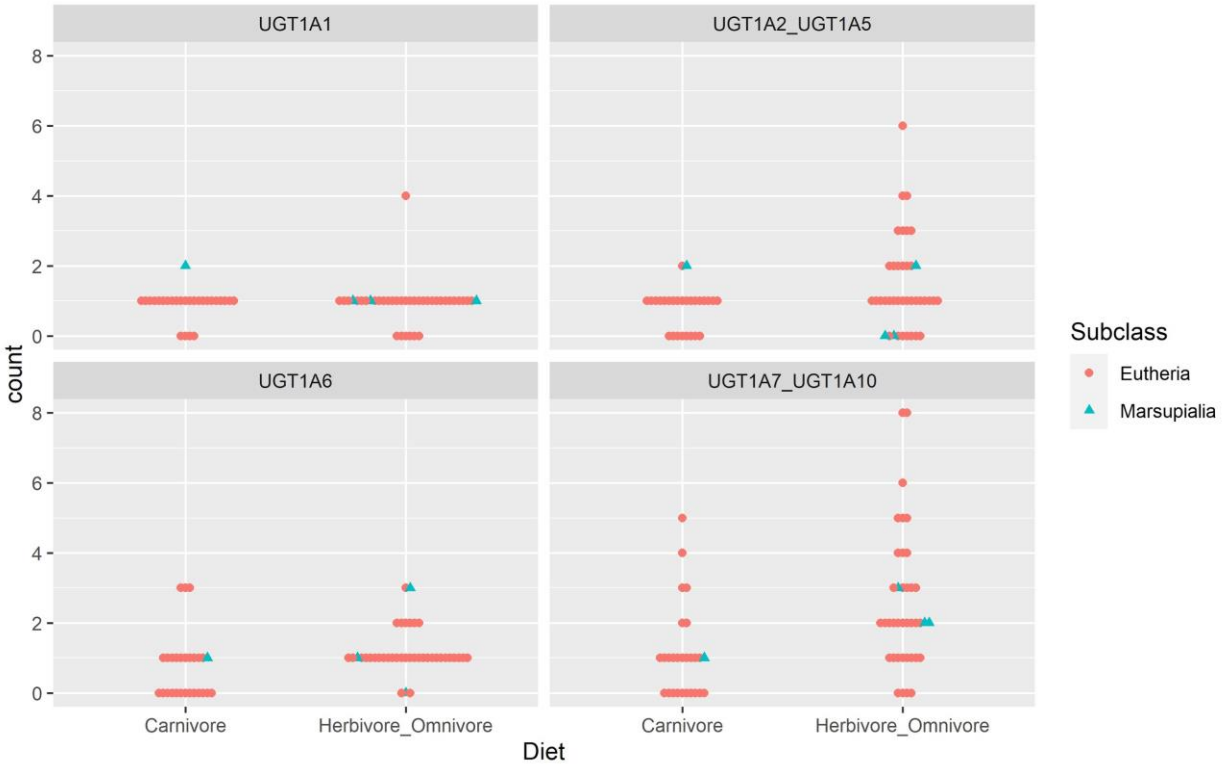

Figure 5. The relationship between the number of *UGT1A* genes and feeding habit. The numbers of "*UGT1A1*" and "*UGT1A6*" subgroup genes are relatively consistent, while those of "*UGT1A2–UGT1A5*" and "*UGT1A7–UGT1A10*" subgroup genes are relatively variable. There is no significant difference between the numbers of "*UGT1A*" subgroup genes in carnivorous and herbivorous or omnivorous mammals; all *P*-values are more than 0.05.

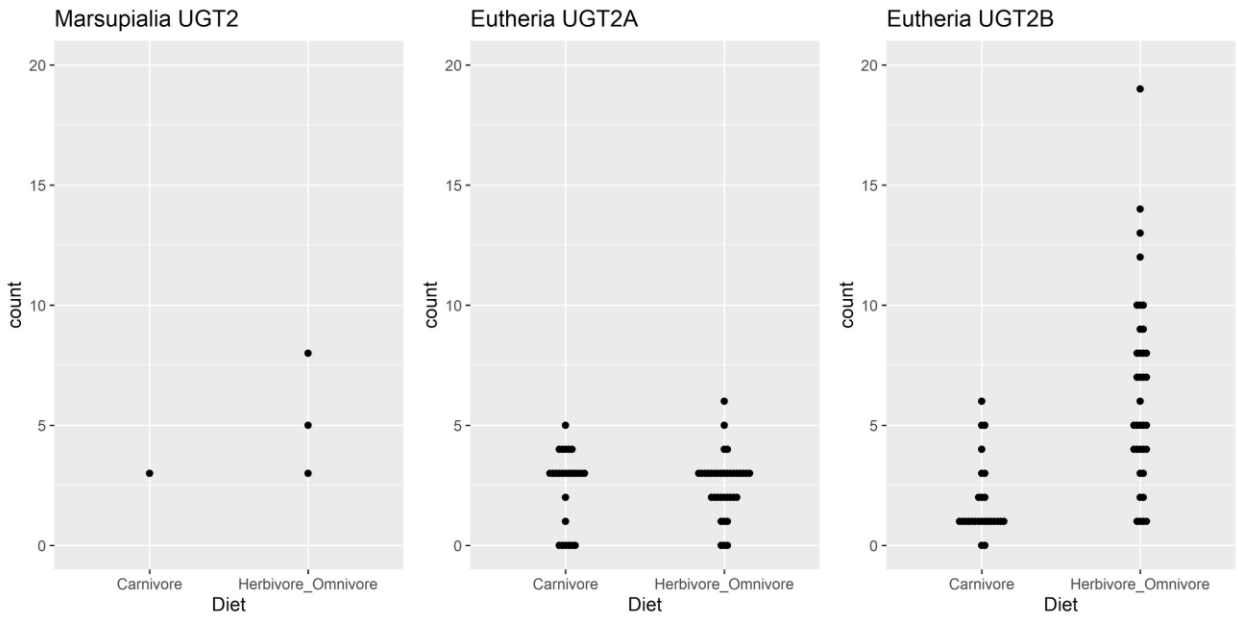

Figure 6. The relationship between the number of UGT2 family genes and feeding habit. In Marsupialia, the number of *UGT2* genes was higher in the herbivore or omnivore group than in the carnivore group, although the sample size was small. In Eutheria, the number of *UGT2A* genes showed no significant difference between the two groups (P = 0.922), while the number of *UGT2B* genes was significantly higher in herbivores or omnivores than in carnivores (P = 0.001).

3.3 Relationship between the number of *UGT* genes and feeding habit

Results of the phylogenetically corrected ANOVA revealed that the number of genes in each *UGT1A* subgroup was not significantly different between carnivores and herbivores or omnivores, indicating that the number of eutherian *UGT2A* genes were not significantly different between feeding habits. In contrast, the number of eutherian *UGT2B* genes was significantly different between feeding habits. We could not analyze the number of *UGT2* genes in Marsupialia because of the limited genomic and dietary information on marsupials. However, we observed that the number of *UGT2* genes in pure carnivores is generally lower than that of herbivores or omnivores (Fig. 6).

3.4 Relationship between selection pressures on *UGT* genes and feeding habit

The selection rate analysis revealed that there was no significant relaxed selection on *UGT1A6* in carnivorous mammals. In eutherian *UGT2B* genes, we found evidence of episodic diversifying selection on five out of 368 branches that were selected from the herbivorous or omnivorous mammals. The five branches consisted of the 1st exons in the *UGT2B* genes of four species, namely the pale spear-nosed bat (*Phyllostomus discolor*), European rabbit, guinea pig (*Cavia porcellus*), and Sunda flying lemur (*Galeopterus variegatus*). In marsupial *UGT2* genes, we also found evidence of episodic diversifying selection on four out of 35 branches that were selected from the herbivorous or omnivorous mammals. All four branches included the *UGT2* genes of koala (*Phascolarctos cinereus*).

Discussion

The locus information indicated that almost all *UGT1A* genes were composed of variable 1st exons and constant 2nd–5th exons. Furthermore, the evolutionary history of *UGT1A* genes was deduced from the phylogenetic analysis of the 1st exons, revealing that the common ancestor of mammals and birds possessed at least seven UGT1 family genes. This is in agreement with a previous report on the common ancestor of vertebrates possessing multiple UGT1 family genes (40). The phylogenetic analysis also indicated that the UGT1A subfamily genes would show the birth–death evolution after the divergence of mammals and birds (41).

The *UGT1A* genes of marsupials and eutherians can be divided into four subgroups, namely "*UGT1A1*," "*UGT1A2–UGT1A5*," "*UGT1A6*," and "*UGT1A7–UGT1A10*", corresponding to the human *UGT1A* (1). The number of *UGT1A* genes were diversified among species. However, the diversification patterns differed among the *UGT1A* subgroups. For example, the "*UGT1A1*" and "*UGT1A6*" subgroups were consistent among mammalian species. In particular, *UGT1A1* was found to be conserved in almost all mammals. In contrast, the "*UGT1A2–UGT1A5*" and "*UGT1A7–UGT1A10*" subgroups showed diversity among mammalian species. Therefore, the common ancestor of eutherians and marsupials might have possessed four UGT1A subgroups, with their functions conserved throughout mammalian evolutionary history.

Our results suggest that UGT1A1, which is known to conjugate bilirubin (42), was important for mammalian species; hence, the number of UGT1A1 was conserved during evolution. However, platypus does not possess the *UGT1A1* ortholog, suggesting that the transformation of biliverdin to bilirubin is only necessary in viviparous mammals for the antioxidant effects that protect neonates from birth-induced oxidative stress (43). It is also possible that another *UGT1A* ortholog may be involved in the bilirubin conjugation in platypus.

Traditionally, it was thought that human UGT1A2–UGT1A5 enzymes conjugate endogenous substrates, such as bile acids, while UGT1A7–UGT1A10 conjugate exogenous substances, such as phenol (44, 45). However, UGT1A3, which was found in the "*UGT1A2–UGT1A5*" subgroup, reportedly conjugates other exogenous chemicals, such as amines, flavonoids, and hydroxylated benzo[a]pyrene (42). Furthermore, our results suggest that the enzymes in the "*UGT1A2–UGT1A5*" subgroup are required for the metabolism of exogenous chemicals. However, the number of genes in the "*UGT1A2–UGT1A5*" subgroup was not significantly different between carnivorous and herbivorous or omnivorous mammals.

UGT1A6 is a major drug-metabolizing enzyme (46), with its gene known as a pseudogene in numerous carnivorous species (5, 6). However, UGT1A6 was not duplicated in many herbivorous or omnivorous species. In the present study, we did not observe significant differences between the numbers of genes of the *UGT1A* subgroups. There was also no relaxation of selection pressure on *UGT1A6* in carnivores. These results corroborate the previous report that failed to find a significant relationship between gene loss of *UGT1A6* and feeding habits (47), suggesting that *UGT1A6* is partially involved in the metabolism of plant chemicals but mainly involved in the conjugation of exogenous chemicals that existed before the emergence of flowering plants, such as benzo[a]pyrene metabolites (46).

The "*UGT1A7–UGT1A10*" subgroup genes were discovered to be more abundant in herbivores or omnivores than in carnivores, although there was no significant difference between the two. In humans, UGT1A7–UGT1A10 conjugate not only carcinogenic compounds

(48, 49) but also dietary compounds, such as flavonols and isoflavones (50, 51). Therefore, the production of new plant chemicals may have caused an increase in the number of "*UGT1A7–UGT1A10*" subgroup genes in herbivorous or omnivorous mammals.

The phylogeny of the constant 2nd–5th exons in *UGT1A* genes, of which birds, platypus, and marsupials formed one clade, did not match the phylogeny of mammalian species. The 2nd–5th exons are believed to correspond to the cofactor (UDP-glucuronic acid or UDP-glucose) binding region in UGT enzymes (10, 52). Hence, it is possible that the preference for UDP-glucuronic acid has changed in the eutherian lineage. Several species possess multiple sets of 2nd–5th *UGT1A* exons, while some species do not have these exons, which is probably due to the limited assembly information available in the database. Therefore, future studies are needed to verify whether these *UGT1A* genes are functional in all mammals.

The locus information revealed that some *UGT2A* genes were composed of variable 1st exons and constant 2nd–6th exons, while other *UGT2A* and *UGT2B* genes were unigenes composed of 1st–6th exons. Therefore, we deduced the evolutionary history of UGT2 family genes from the phylogenetic analysis of the constant 2nd–6th exons and discovered that mammalian *UGT2* genes grouped into four major clades, namely the platypus *UGT2*, marsupial *UGT2*, eutherian *UGT2A*, and eutherian *UGT2B*. This result suggests that after the divergence of Marsupialia and Eutheria, the UGT2 family was divided into two eutherian subgroups, UGT2A and UGT2B. Moreover, the appearance of *UGT2* genes in platypus and marsupials, which were composed of variable 1st exons and constant 2nd–6th exons, suggested that *UGT2* unigene duplication occurred during eutherian evolution.

The differences in the number of genes per clade demonstrate that the diversification of eutherian *UGT2B* genes was influenced by feeding habits, in which the number of *UGT2B* genes was higher in herbivores or omnivores than in carnivores. In Eutheria, *UGT2B* genes function in the metabolism of steroid hormones in the liver (53, 54) and some are known to metabolize flavonoids (55) that are known to be steroid-like chemicals (56).Therefore, this difference suggests that the increase in the number of *UGT2B* genes may have resulted from the need to metabolize exogenous chemicals in food, including flavonoids. Furthermore, marsupial *UGT2* genes may have also diversified due to changes in feeding habits. However, because of the limited number of species and incomplete genome assembly, we could not confirm its correlation to feeding habits.

The number of *UGT2A* genes was consistent among mammalian species ($n \leq 4$), except in armadillo, European rabbit, and Chinese tree shrew. In humans, the *UGT2A1* and *UGT2A2* genes function during the local detoxification of polycyclic aromatic hydrocarbons (PAHs) in the olfactory neuroepithelium (3, 57). Therefore, the consistent number of eutherian *UGT2A* genes indicate that these are not related to the conjugation of food chemicals. In contrast, the human *UGT2A3*, a unigene composed of 1st–6th exons, is primarily expressed in the intestine, liver, and kidney (4, 58). Therefore, the *UGT2A* genes in armadillo, European rabbit, and Chinese tree shrew, which were duplicated similar to *UGT2B* genes, may be specifically associated with xenobiotic metabolism in the liver.

Plants synthesize many secondary metabolites for protection against herbivorous animals, including vertebrates and insects (59). Flowering plants (Class Angiospermae) are a major group that diversified during the Cretaceous period and became a primary food source on Earth (13). Before the Cretaceous period, Eutheria and Marsupialia diverged, and both

mammalian groups were independently required to ingest and metabolize the secondary metabolites from flowering plants for continued survival.

In this study, we hypothesized that the *UGT1A* genes may have played a role in the xenobiotic metabolism in the common ancestor of mammals before the emergence of flowering plants. Although some UGT1A isozymes are known to conjugate the secondary metabolites of flowering plants, such as cannabinol and nicotine (60, 61), we did not observe a relationship between the number of UGT1A subgroup genes and feeding habit. This weak relationship suggests that ancestral *UGT1A* genes might have conjugated certain chemicals, such as polycyclic aromatic hydrocarbons (PAHs), and toxic metabolites from the ferns (Class Gymnospermae) that existed before Cretaceous period and there are historical constraints in the *UGT1A* genes.

However, a previous study suggested that the *UGT2* genes in the common ancestor of mammals do not have diversity and do not play a role in hepatic xenobiotic metabolism (8). In this study, phylogenetic analysis of the 1st exons in *UGT2* genes that is related to aglycone recognition (64), reveals that marsupial *UGT2* genes, which are composed of variable 1st exons and constant exons similar to eutherian *UGT2A1* and *UGT2A2*, formed two different clades with eutherian *UGT2A* and *UGT2B* genes. This result suggests that in the common ancestor of Eutheria and Marsupialia possessed both *UGT2A*-like and *UGT2B*-like 1st exons. Moreover, the phylogenetic tree of the 2nd–6th exons showed *UGT2A* and *UGT2B* genes were separated in the eutherian lineage. These phylogenetic inferences suggest that in the eutherian lineage 2nd–6th exons were duplicated and separated from original *UGT2* genes with *UGT2B*-like 1st exon as unigenes. Unlike splicing variant genes such as *UGT1* family genes, unigene of *UGT2B*, which means that the distance between the 1st exon and the 2nd exon does not separate even when the number of genes increases, may have created room for more genes to be generated by duplication. The phylogenetic tree of the 2nd–6th exons also demonstrate that several duplications of *UGT2B* genes separately occurred in each eutherian family. This suggests that *UGT2B* genes have independently evolved in each lineage to facilitate the adaptation to species-specific environments and that the common ancestor of eutherian mammals “invented” the *UGT2B* genes as unigenes for metabolizing variable flowering plant chemicals. Furthermore, in each lineage, the function of the UGT2B subgroup expanded for conjugating species-specific xenobiotics. In Eutheria, there are some exceptions to this hypothesis, such as the armadillos that possess diversified *UGT2A* genes, not *UGT2B*. Armadillos are carnivorous mammals that mainly eats small animals, such as insects (62). Therefore, the high number of *UGT2A* genes in armadillo may have been caused by specific endogenous chemicals. Furthermore, since armadillos are xenarthran mammals that diverged from other eutherians in the Cretaceous period (63), this species might reflect the ambiguity of the functions between ancestral *UGT2B* and *UGT2A* in xenobiotic metabolism.

We tested the hypothesis that exposure to flowering plant chemicals was the major selection pressure on the 1st exons of *UGT2* genes in Eutheria and Marsupialia using aBSREL analysis. Our data suggest that there was no positive selection on the *UGT2B* genes in several herbivorous eutherian species. This may have been caused by the huge number of *UGT2B* genes analyzed, and the corrected multiple comparison tests reduced the power of detection. However, we discovered that there was episodic diversifying selection in four herbivorous or

omnivorous species, which supports our hypothesis that the plant chemicals in food may have influenced the divergence of *UGT2B* genes. Moreover, we found evidence of episodic diversifying selection on the branches of *UGT2* genes in koala. Koalas are mammals that eat the leaves of eucalyptus trees that contain toxic chemicals, such as phenolic compounds and terpenes (65, 66). Although the genome assembly is incomplete, the selection evidence suggests that the *UGT2* genes in koala may have also evolved to adapt to these plant chemicals. Thus, our results suggest that convergent/parallel evolution occurred between the *UGT2* genes of marsupials and eutherians after adaptation to conjugating plant chemicals.

In summary, we traced the evolutionary history of mammalian *UGT1A* and *UGT2* genes by analyzing their phylogenies, determining the relationship between number of genes and feeding habit, and estimating selection pressures. The results revealed that *UGT2B* genes emerged in Eutheria after the divergence from Marsupialia. Our findings also suggest that parallel evolution may have occurred in the *UGT2* genes of eutherian and marsupial lineages after the diversification of flowering plants in the Cretaceous period.

Acknowledgments

This work was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (MEXT) awarded to M. Ishizuka (No. 18K19847 and 21H04919), and Hokkaido University SOUSEI Support Program. This research was also supported by Program for supporting introduction of the new sharing system (JPMXS0420100619). This research was also supported by MEXT awarded to Y. Kawai (No. 20K15848). The English in this manuscript was proofread by Editage (https://www.editage.jp/).

References

1. P. I. Mackenzie, *et al.*, Nomenclature update for the mammalian UDP glycosyltransferase (UGT) gene superfamily. *Pharmacogenet. Genomics* **15**, 677–685 (2005).

2. A. Rowland, J. O. Miners, P. I. Mackenzie, The UDP-glucuronosyltransferases: Their role in drug metabolism and detoxification. *Int. J. Biochem. Cell Biol.* **45**, 1121–1132 (2013).

3. G. Jedlitschky, A. J. Cassidy, M. Sales, N. Pratt, B. Burchell, Cloning and characterization of a novel human olfactory UDP-glucuronosyltransferase. *Biochem. J* **340 (Pt 3)**, 837–843 (1999).

4. M. H. Court, S. Hazarika, S. Krishnaswamy, M. Finel, J. A. Williams, Novel Polymorphic Human UDP-glucuronosyltransferase 2A3: Cloning, Functional Characterization of Enzyme Variants, Comparative Tissue Expression, and Gene Induction. *Mol. Pharmacol.* **74**, 744–754 (2008).

5. B. Shrestha, *et al.*, Evolution of a major drug metabolizing enzyme defect in the domestic cat and other felidae: phylogenetic timing and the role of hypercarnivory. *PLoS One* **6**, e18046 (2011).

6. M. Kakehi, *et al.*, Uridine Diphosphate-Glucuronosyltransferase (UGT) Xenobiotic Metabolizing Activity and Genetic Evolution in Pinniped Species. *Toxicol. Sci.* **147**, 360–369 (2015).

7. T. Kondo, *et al.*, Uridine Diphosphate-Glucuronosyltransferase (UGT) 2B Subfamily Interspecies Differences in Carnivores. *Toxicol. Sci.* **158**, 90–100 (2017).

8. Y. K. Kawai, Y. Ikenaka, M. Ishizuka, A. Kubota, The evolution of UDP-glycosyl/ glucuronosyltransferase 1E (UGT1E) genes in bird lineages is linked to feeding habits but UGT2 genes is not. *PLoS One* **13**, 1–13 (2018).

9. Y. K. Kawai, *et al.*, Characterization of function and genetic feature of UDP-glucuronosyltransferase in avian species. *Comparative Biochemistry and Physiology Part - C: Toxicology and Pharmacology* **217**, 5–14 (2019).

10. K. W. Bock, Vertebrate UDP-glucuronosyltransferases: functional and evolutionary aspects. *Biochem. Pharmacol.* **66**, 691–696 (2003).

11. F. M. Afendi, *et al.*, KNApSAcK family databases: integrated metabolite-plant species databases for multifaceted plant research. *Plant Cell Physiol.* **53**, e1 (2012).

12. M. J. M. Christenhusz, J. W. Byng, The number of known plants species in the world and its annual increase. *Phytotaxa* **261**, 201–217 (2016).

13. H. T. Li, *et al.*, Origin of angiosperms and the puzzle of the Jurassic gap. *Nature Plants* **5**, 461–470 (2019).

14. P. A. Kitts, *et al.*, Assembly: a resource for assembled genomes at NCBI. *Nucleic Acids Res.* **44**, D73-80 (2016).

15. F. Cunningham, *et al.*, Ensembl 2019. *Nucleic Acids Res.* **47**, D745–D751 (2018).

16. C. Camacho, *et al.*, BLAST plus: architecture and applications. *BMC Bioinformatics* **10**, 1 (2009).

17. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

18. N. A. O'Leary, *et al.*, Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733-45 (2016).

19. G. Schofl, reutils: Talk to the NCBI EUtils (2016).

20. L. Guy, J. R. Kultima, S. G. E. Andersson, J. Quackenbush, GenoPlotR: comparative gene and genome visualization in R. *Bioinformatics* **27**, 2334–2335 (2011).

21. R Core Team, *R: A Language and Environment for Statistical Computing* (2020).

22. K. Katoh, D. M. Standley, MAFFT Multiple Sequence Alignment Software Version 7 : Improvements in Performance and Usability Article Fast Track. **30**, 772–780 (2013).

23. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).

24. A. S. Tanabe, Kakusan4 and Aminosan: Two programs for comparing nonpartitioned, proportional and separate models for combined molecular phylogenetic analyses of multilocus sequence data. *Mol. Ecol. Resour.* **11**, 914–921 (2011).

25. Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

26. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

27. S. B. Hedges, J. Dudley, S. Kumar, TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22**, 2971–2972 (2006).

28. S. B. Hedges, J. Marin, M. Suleski, M. Paymer, S. Kumar, Tree of life reveals clock-like speciation and diversification. *Mol. Biol. Evol.* **32**, 835–845 (2015).

29. S. Kumar, G. Stecher, M. Suleski, S. B. Hedges, TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).

30. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T. Lam, ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).

31. G. Yu, T. T.-Y. Lam, H. Zhu, Y. Guan, Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree. *Mol. Biol. Evol.* **35**, 3041–3043 (2018).

32. G. Yu, Using ggtree to Visualize Data on Tree-Like Structures. *Curr. Protoc. Bioinformatics* **69**, e96 (2020).

33. S. Faurby, *et al.*, PHYLACINE 1.2: The Phylogenetic Atlas of Mammal Macroecology. *Ecology* **99**, 2626 (2018).

34. L. J. Revell, phytools: an R package for phylogenetic comparative biology (and other things): phytools: R package. *Methods Ecol. Evol.* **3**, 217–223 (2012).

35. J. O. Wertheim, B. Murrell, M. D. Smith, S. L. Kosakovsky Pond, K. Scheffler, RELAX: detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* **32**, 820–832 (2015).

36. S. L. K. Pond, S. D. W. Frost, S. V. Muse, HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679 (2004).

37. M. D. Smith, *et al.*, Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* **32**, 1342–1353 (2015).

38. Y. Benjamini, Y. Hochberg, Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing Author (s): Yoav Benjamini and Yosef Hochberg Source : Journal of the Royal Statistical Society . Series B (Methodological), Vol . 57 , No . 1 Published by : *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).

39. S. Weaver, *et al.*, Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Mol. Biol. Evol.* **35**, 773–777 (2018).

40. H. Huang, Q. Wu, Cloning and comparative analyses of the zebrafish Ugt repertoire reveal its evolutionary diversity. *PLoS One* **5**, e9144 (2010).

41. M. Nei, A. P. Rooney, Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* **39**, 121–152 (2005).

42. R. Meech, *et al.*, The UDP-Glycosyltransferase (UGT) Superfamily: New Members, New Functions, and Novel Paradigms. *Physiol. Rev.* **99**, 1153–1222 (2019).

43. T. W. Sedlak, S. H. Snyder, Bilirubin benefits: cellular protection by a biliverdin reductase antioxidant cycle. *Pediatrics* **113**, 1776–1782 (2004).

44. Y. Emi, S. Ikushiro, T. Iyanagi, Drug-responsive and tissue-specific alternative expression of multiple first exons in rat UDP-glucuronosyltransferase family 1 (UGT1) gene complex. *J. Biochem.* **117**, 392–399 (1995).

45. J. Trottier, *et al.*, Human UDP-glucuronosyltransferase (UGT)1A3 enzyme conjugates chenodeoxycholic acid in the liver. *Hepatology* **44**, 1158–1170 (2006).

46. K. W. Bock, C. Köhle, UDP-glucuronosyltransferase 1A6: structural, functional, and regulatory aspects. *Methods Enzymol.* **400**, 57–75 (2005).

47. N. Hecker, V. Sharma, M. Hiller, Convergent gene losses illuminate metabolic and physiological changes in herbivores and carnivores. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 3036‑3041 (2019).

48. M. Lacko, *et al.*, Genetic polymorphisms in the tobacco smoke carcinogens detoxifying enzyme UGT1A7 and the risk of head and neck cancer. *Head Neck* **31**, 1274–1281 (2009).

49. R. W. Dellinger, J.-L. Fang, G. Chen, R. Weinberg, P. Lazarus, Importance of UDP-glucuronosyltransferase 1A10 (UGT1A10) in the detoxification of polycyclic aromatic hydrocarbons: decreased glucuronidative activity of the UGT1A10139Lys isoform. *Drug Metab. Dispos.* **34**, 943–949 (2006).

50. E. D. Niemeyer, J. S. Brodbelt, Regiospecificity of human UDP-glucuronosyltransferase isoforms in chalcone and flavanone glucuronidation determined by metal complexation and tandem mass spectrometry. *J. Nat. Prod.* **76**, 1121–1132 (2013).

51. J.-Q. Ruan, R. Yan, Regioselective glucuronidation of the isoflavone calycosin by human liver microsomes and recombinant human UDP-glucuronosyltransferases. *Chem. Biol. Interact.* **220**, 231–240 (2014).

52. C. W. Locuson, T. S. Tracy, Comparative modelling of the human UDP-glucuronosyltransferases: insights into structure and mechanism. *Xenobiotica* **37**, 155–168 (2007).

53. A. Bélanger, G. Pelletier, F. Labrie, O. Barbier, S. Chouinard, Inactivation of androgens by UDP-glucuronosyltransferase enzymes in humans. *Trends Endocrinol. Metab.* **14**, 473–479 (2003).

54. D. Turgeon, J. S. Carrier, E. Lévesque, D. W. Hum, A. Bélanger, Relative enzymatic activity, protein stability, and tissue distribution of human steroid-metabolizing UGT2B subfamily members. *Endocrinology* **142**, 778–787 (2001).

55. D. Turgeon, J.-S. Carrier, S. Chouinard, A. Bélanger, Glucuronidation activity of the UGT2B17 enzyme toward xenobiotics. *Drug Metab. Dispos.* **31**, 670–676 (2003).

56. R. S. Zand, D. J. Jenkins, E. P. Diamandis, Steroid hormone activity of flavonoids and related compounds. *Breast Cancer Res. Treat.* **62**, 35–49 (2000).

57. R. T. Bushey, D. F. Dluzen, P. Lazarus, Importance of UDP-glucuronosyltransferases 2A2 and 2A3 in tobacco carcinogen metabolism. *Drug Metab. Dispos.* **41**, 170–179 (2013).

58. S. Gotoh-Saito, *et al.*, Characterization of human UGT2A3 expression using a prepared specific antibody against UGT2A3. *Drug Metab. Pharmacokinet.* **34**, 280–286 (2019).

59. A. Mithöfer, W. Boland, Plant defense against herbivores: chemical aspects. *Annu. Rev. Plant Biol.* **63**, 431–450 (2012).

60. G. E. Kuehl, S. E. Murphy, N-glucuronidation of nicotine and cotinine by human liver microsomes and heterologously expressed UDP-glucuronosyltransferases. *Drug Metab. Dispos.* **31**, 1361–1368 (2003).

61. A. Mazur, *et al.*, Characterization of human hepatic and extrahepatic UDP-glucuronosyltransferase enzymes involved in the metabolism of classic cannabinoids. *Drug Metab. Dispos.* **37**, 1496–1504 (2009).

62. T. C. da Silveira Anacleto, Food habits of four armadillo species in the Cerrado area, Mato Grosso, Brazil. *ZOOLOGICAL STUDIES-TAIPEI-* **46**, 529 (2007).

63. B. M. Hallström, M. Kullberg, M. A. Nilsson, A. Janke, Phylogenomic data analyses provide evidence that Xenarthra and Afrotheria are sister groups. *Mol. Biol. Evol.* **24**, 2059–2068

(2007).

64. C. Li, Q. Wu, Adaptive evolution of multiple-variable exons and structural diversity of drug-metabolizing enzymes. *BMC Evol. Biol.* **7**, 69 (2007).

65. B. M. Eschler, D. M. Pass, R. Willis, W. J. Foley, Distribution of foliar formylated phloroglucinol derivatives amongst Eucalyptus species. *Biochem. Syst. Ecol.* **28**, 813–824 (2000).

66. C. Külheim, *et al.*, The Eucalyptus terpene synthase gene family. *BMC Genomics* **16**, 450 (2015).