Lipidome Evolution in Mammalian Tissues

Ekaterina Khrameeva,[†,1,2] Ilia Kurochkin,[†,1] Katarzyna Bozek,[3] Patrick Giavalisco,[*,4,7] and Philipp Khaitovich[*,1,5,6]

[1]Center for Data-Intensive Biomedicine and Biotechnology, Skolkovo Institute of Science and Technology, Moscow, Russia

[2]A.A.Kharkevich, Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia

[3]Biological Physics Theory Unit, Okinawa Institute of Science and Technology, Graduate University, Onna-Son, Kunigami-Gun, Okinawa, Japan

[4]Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany

[5]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

[6]CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai, China

[7]Current affiliation: Max Planck Institute for Biology of Ageing, Cologne, Germany

[†]These authors contributed equally to this work.

***Corresponding authors:** E-mails: giavalisco@age.mpg.de; khaitovich@eva.mpg.de.

Associate editor: Joel Dudley

Abstract

Lipids are essential structural and functional components of cells. Little is known, however, about the evolution of lipid composition in different tissues. Here, we report a large-scale analysis of the lipidome evolution in six tissues of 32 species representing primates, rodents, and bats. While changes in genes' sequence and expression accumulate proportionally to the phylogenetic distances, <2% of the lipidome evolves this way. Yet, lipids constituting this 2% cluster in specific functions shared among all tissues. Among species, human show the largest amount of species-specific lipidome differences. Many of the uniquely human lipidome features localize in the brain cortex and cluster in specific pathways implicated in cognitive disorders.

Key words: **humans, brain, lipidome, molecular evolution, mass spectrometry.**

Introduction

Genome sequence and phenotype differences among species are relatively well studied. This knowledge forms the foundation for evolutionary models that explain observed genetic and phenotypic divergence and help to identify genetic regions or phenotypic traits influenced by a specific type of evolutionary selection. Under most, though not all, current models, the majority of genetic differences observed among species are considered to be evolutionarily neutral, that is, have no effect on the individuals' survival and reproductive abilities, while a few differences are adaptive, that is, increase the individual's fitness (Otto 2000; Eyre-Walker 2006; Hahn 2008; Orr 2009). Genetic changes with substantial deleterious effects are removed by purifying selection and are not observed in comparisons among species. Accordingly, some of the main features of genetic evolution include: 1) an overwhelming correlation between phylogenetic distances and genetic distances; 2) reduced diversity (variation within species) and divergence (differences between species) of functional sequences compared with nonfunctional ones; and 3) reduced diversity but the increased divergence of sequences containing adaptive variants.

In contrast to genetic differences, most changes that occur at the level of the phenotype affect individuals' survival and reproductive chances and only a few changes are neutral (Nei 2007; Steiner and Tuljapurkar 2012). While the genetic information largely determines the phenotype, several intermediate steps enable the transition of this information to the phenotype level. These steps include changes in epigenetic modification levels of DNA and histone proteins, RNA and protein expression levels, as well as concentrations of small molecules, metabolites, and lipids, that participate in all biological processes. Some of these steps, including RNA expression levels have been examined, resulting in evolutionary models parallel to the one based on genetic data (Wray et al. 2003; Jordan et al. 2005; Stern and Orgogozo 2008). Yet, the driving forces determining evolutionary changes of metabolite and lipid concentrations have not been explored.

In this study, we assess the evolution at another intermediate step connecting the genome with the phenotype—the level of lipid concentration changes. Lipids represent the hydrophobic fraction of small biological molecules with a molecular weight <1500 Da, known as metabolites. Lipids are essential components of cell plasma membranes, energy storage blocks and signaling messengers in different cells and tissues, especially those of the brain (Simons and Toomre 2000). Their critical role in cell functioning is supported by their involvement in metabolic and neurological disorders, such as Alzheimer's or Parkinson's diseases, as well as in diabetes and cancer (Han et al. 2002; Wenk 2005; Adibhatla et al. 2006; Ariga et al. 2008; Colsch et al. 2008; Haughey et al. 2010; Lamari et al. 2013). Preliminary comparisons conducted in three to four

Article

mammalian species demonstrated the rapid evolution of lipid concentration levels, in particular in the human lineage (Bozek et al. 2015; Li et al. 2017). Nevertheless, the evolutionary processes driving lipid concentration divergence among species, including humans, remain unexplored. Here, we addressed this question by analyzing the lipidome compositions of six tissues in 32 mammalian species based on data collected using liquid chromatography coupled with high-precision mass spectrometry.

Results

Lipidome Data Description

We analyzed lipid concentrations in the liver (LV), muscle (ML), kidney (KD), heart (HT), brain cortex (CX), and cerebellum (CB) using 669 samples obtained from 32 mammalian species representing three phylogenetic clades: rodents, primates, and bats (Bozek et al. 2017) (fig. 1*a* and supplementary table 1, Supplementary Material online). Our analysis was based on the quantification of >18,000 (median = 18,708) mass spectrometry features in each tissue corresponding to hydrophobic compounds (lipids) detected in at least 50% of the individuals of each species. Among these lipids, close to 7,000 (median = 6,994) were annotated in each tissue using computational matching to the hydrophobic compound database (Fahy et al. 2009) (fig. 1*b*; supplementary fig. 1 and table 2, Supplementary Material online). A multidimensional scaling analysis revealed that the samples segregated predominantly according to organ origin (fig. 1*c*).

Lipidome Differences and Phylogenetic Distances among Species

Most of the genome sequence differences and a substantial proportion of gene expression differences among species tend to scale linearly with phylogenetic distances. This indicates that the majority of these changes might have no effect on survival and reproductive abilities (Lanfear et al. 2010; Kuraku et al. 2016). We used linear regression and Blomberg's K (Ma et al. 2015) approaches, analogous to the ones applied in the genome sequence and gene expression analyses, to identify lipids that change linearly with phylogenetic distances between species (phylogeny-dependent lipids). Surprisingly, on an average, only 2% of detected lipids showed this type of concentration differences (R^2: median = 1.9%, maximum = 9.2% in CB; Blomberg's K: median = 2.1%, maximum = 9.1% in CB; fig. 2*a* and supplementary figs. 2–9, Supplementary Material online). Notably, the observed proportions of phylogeny-dependent lipids were still significantly greater than expected by chance in all six tissues (permutations, $P < 0.001$). The result was not caused by the selection of species used in the study: a parallel analysis of the transcriptome and lipidome divergence in a subset of three tissues of eight species demonstrated that 30–37% of detected protein-coding transcripts showed phylogeny-dependent expression. For the lipids, however, the median proportion of phylogeny-dependent differences remained low (3.3%) (fig. 2*b* and supplementary fig. 10, Supplementary Material online).

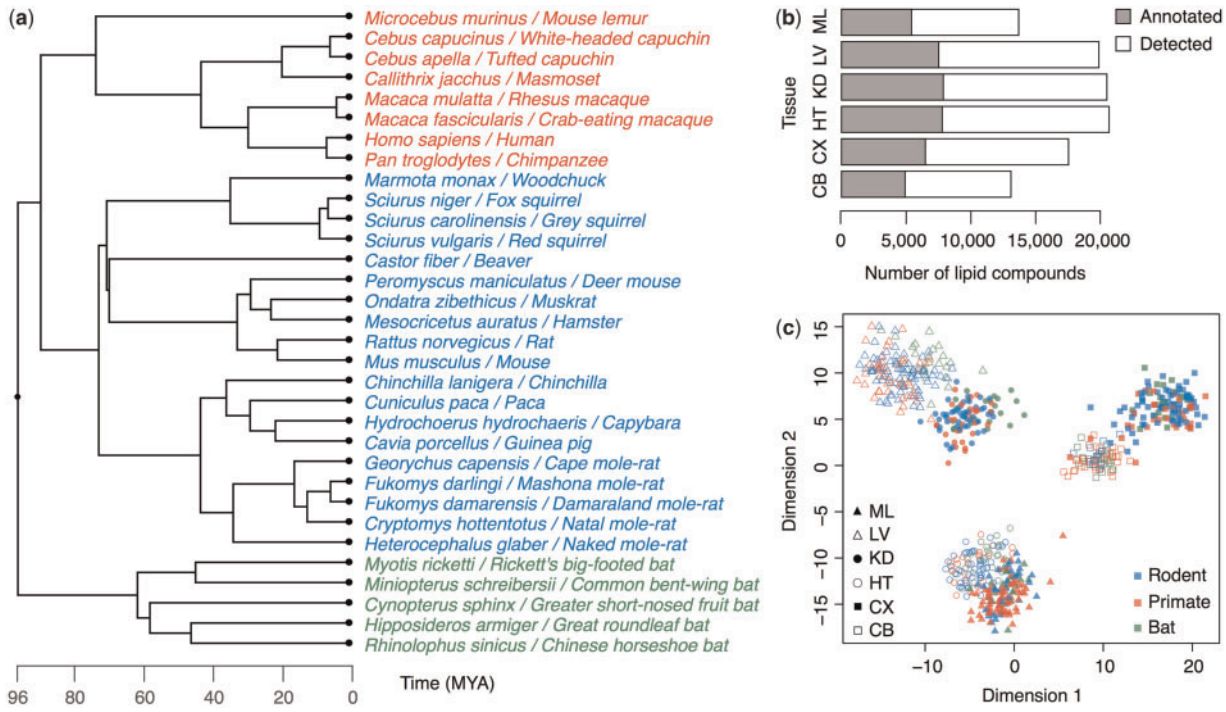

FIG. 1. Data overview. (*a*) Phylogenetic tree of 32 mammalian species used in this study. The colors indicate three main represented clades: rodents (blue), primates (red), and bats (green). (*b*) Numbers of lipids detected in each tissue. Gray bars represent annotated lipids. Here and later the tissues are labeled by a two-letter code: CB, cerebellum; CX, cortex; HT, heart; KD, kidney; LV, liver; ML, muscle. (*c*) The relationship among samples are based on concentrations of 1,231 annotated lipids plotted in two dimensions using a multidimensional scaling algorithm. Symbols represent tissues, colors represent clades, and points represent individual samples.

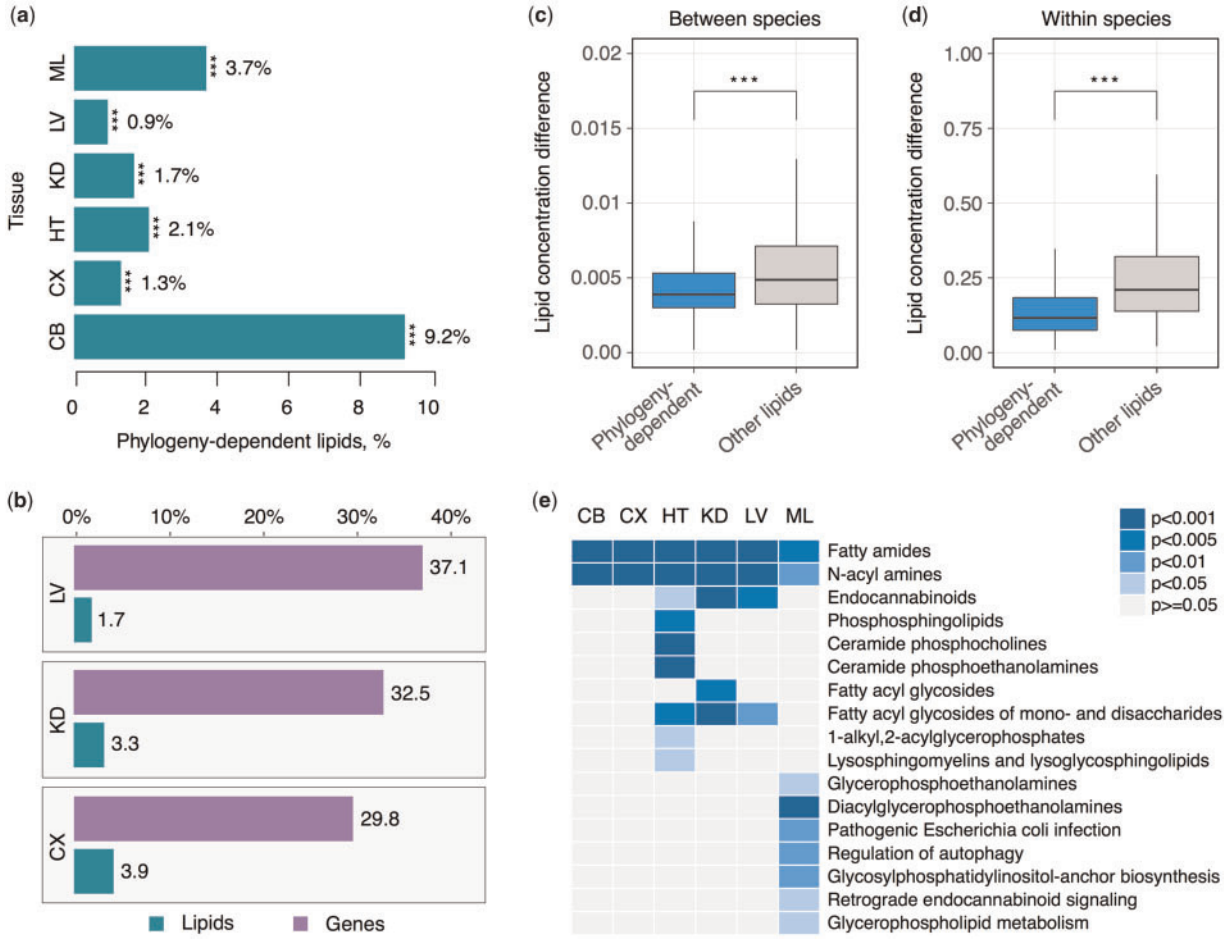

Fig. 2. Relationship between lipid concentrations and phylogenetic distances. (*a*) Percentages of lipids with concentration differences among species scaling with phylogenetic distances (phylogeny-dependent lipids) in each of six tissues. Stars indicate the significance of the difference between observed number distributions and random expectation (permutations of species labels, $P < 0.001$). (*b*) Percentages of phylogenetic genes (violet) and lipids (green) defined based on the same criteria in three tissues in a matching subset of eight species. The expression data was taken from Fushan et al. (2015). (*c*) Distribution of concentration differences measured in all pairwise comparisons between species for phylogeny-dependent lipids (blue) and the remaining lipids (light gray). Stars indicate the significance of the difference between the two distributions (two-sided *t*-test $P < 0.0001$, $n = 66{,}176$). (*d*) Distribution of concentration differences among individuals within species for phylogeny-dependent lipids (blue) and the remaining lipids (light gray). Stars indicate the significance of the difference between the two distributions (two-sided *t*-test $P < 0.0001$, $n = 3{,}916$). (*e*) Enrichment of phylogeny-dependent lipids in specific lipid classes and subclasses. Colors indicate BH-corrected enrichment *P* values.

On an average, phylogeny-dependent lipids showed fewer lipid concentration differences among species than the remaining lipids in all six tissues combined (two-sided *t*-test, $n = 66{,}176$, $P < 0.0001$, fig. 2c), as well as in each tissue separately, except the cerebellum (supplementary fig. 11, Supplementary Material online). Furthermore, phylogeny-dependent lipids showed reduced intraspecies variation (two-sided *t*-test, $n = 3{,}916$, $P < 0.0001$, fig. 2*d* and supplementary fig. 12, Supplementary Material online). These results were not caused by the difference in concentration levels between phylogeny-dependent and the remaining lipids or by outlier effects (see "Materials and Methods" and supplementary fig. 13, Supplementary Material online).

Remarkably, despite the relatively low numbers of phylogeny-dependent lipids annotated in each tissue (N_{min}=71, N_{median}=148), they showed unusually strong enrichment in the same lipid class, *fatty amides* and its subclass *N-acyl amides*, in all six tissues (hypergeometric test, $n = 3{,}134$, 3,979, 5,858, 5,977, 5,685, and 3,384 in CX, CB, HT, KD, LV, and ML, respectively, FDR < 0.01 in each tissue) (fig. 2*e*). Several other lipid classes, including *phosphosphingolipids*, and *fatty acyl glycosides*, were also enriched in phylogeny dependent lipids in one or several tissues (supplementary table 3, Supplementary Material online).

Species-Specific Lipidome Differences

To further assess the evolutionary dynamics of lipid concentration levels, we identified species-specific differences defined as significant lipid concentration differences between a given species and the others (fig. 3*a*). The analysis was based on lipid concentration measurements from three individuals per species, sampled randomly among all measured species' individuals. Among all 104 tissue and lineage combinations represented by at least three biological replicates, only three

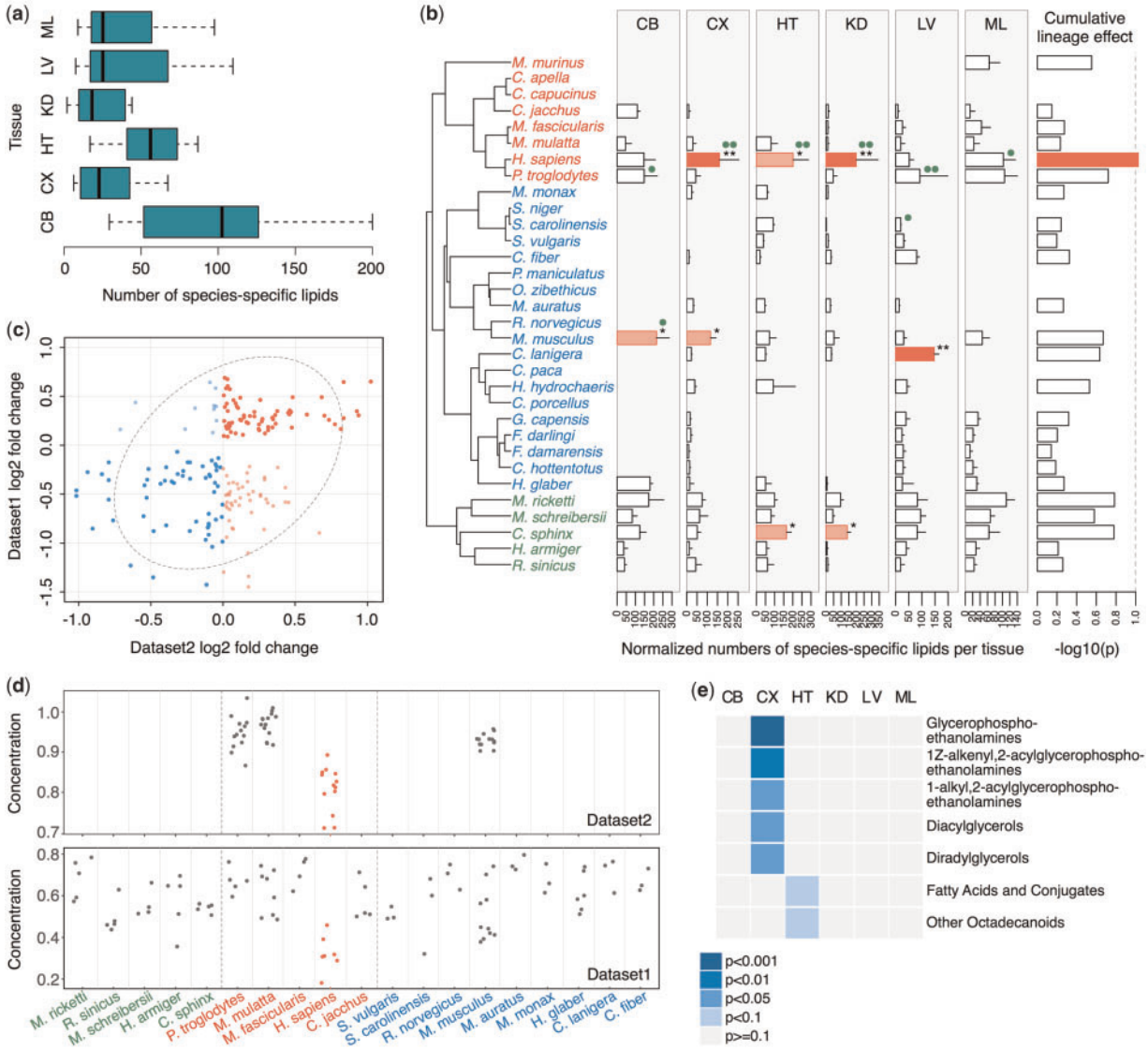

Fig. 3. Species-specific lipid concentration differences. (*a*) Numbers of lipids showing significant species-specific concentration differences. The distributions show the numbers of such lipids in each of 23 lineages represented by at least three biological replicates in a given tissue. (*b*) Numbers of lipids showing significant species-specific concentration differences in each of 104 tissue and lineage combinations (normalized by the phylogenetic distances). Error bars show variations of estimates calculated by way of the random sampling of three individuals per species. Stars and bar colors indicate the significance of the difference between observed number distributions and random expectation (permutations, ** and red—$P < 0.05$, * and pink—$P < 0.1$). Green circles indicate the significance of the difference between observed number distributions and random expectation according to the OU model (permutations, ^{oo}—$P < 0.05$, ^{o}—$P < 0.1$). The rightmost column shows the cumulative lineage effect calculated as an average -log10 *P* value of the difference between observed and chance numbers of species-specific lipids across tissues. (*c*) Lipid concentration differences between humans and the other three species (chimpanzee, macaque, and mouse) calculated as log2-transformed fold changes of the average values for 183 lipids showing a significant human-specific concentration difference in our data (Data Set 1) and detected in the published data set (Data Set 2) (Bozek et al. 2015). Colors indicate signs of log2-transformed fold changes in both data sets. The ellipse shows a 90% confidence interval. (*d*) An example of the concentrations in the kidney of one lipid (monogalactosyldiacylglycerol, LMGL05010014) shown in the panel (*c*). (*e*) Enrichment of lipids with human-specific concentration differences in specific lipid classes and subclasses. Colors indicate BH-corrected enrichment *P* values.

combinations showed significantly more species-specific lipid concentration differences than expected by chance (permutations, $P < 0.05$ at Wilcoxon test $P < 0.01$ cutoff) and five showed a marginally significant trend (permutations, $P < 0.1$ at Wilcoxon test $P < 0.01$ cutoff) (fig. 3*b*). Of these eight, three were located in the human lineage: two significant ones in the brain cortex and kidney and one marginally significant one in the heart. The use of an alternative approach, the Ornstein–Uhlenbeck (OU) model, confirmed three out of three combinations located in the human lineage (permutations, $P < 0.05$ at the OU model $P < 0.01$ threshold) (supplementary fig. 14, Supplementary Material online). Additionally, the human kidney showed a marginally significant excess of lineage-specific differences (permutations, $P < 0.1$ at the OU model $P < 0.01$ threshold). By contrast, an excess of species-specific lipidome differences in the other lineages was not

confirmed by the OU model, except the mouse cerebellum (fig. 3*b*). This result indicates the outstanding character of the lipidome evolution in the human lineage, while a robust detection of more subtle species-specific signals in the other lineages might require greater sample sizes. Consequently, the human evolutionary lineage stood out as having the greatest average lineage-specific lipid concentration divergence compared with the other examined mammalian lineages (fig. 3*b*).

The excess of lipid concentration changes in the human linage was not caused by differences in the postmortem delay among samples, as estimated based on postmortem delay effects identified in Bozek et al. (2015). This excess was not caused by environmental effects either, as human-specific lipid concentration changes were enriched in the brain, yet most lipids composing brain tissue are synthesized in the brain and, therefore, are shielded from environmental and dietary changes by the blood–brain barrier (Sherman and Brophy 2005; Piomelli et al. 2007). Additionally, environmental exposure experiments (stress, exercise, and diet factors) conducted in macaques showed that each of the environmental perturbations induced substantially fewer lipid concentration changes compared with the lipidome differences observed between chimpanzees and humans (Bozek et al. 2015). Consistently, the largest effect of environmental factors was observed in the kidney, and the smallest—in the brain prefrontal cortex (Bozek et al. 2015). Moreover, human-specific differences detected in this study were in agreement with the differences calculated using a published lipidome data set (Bozek et al. 2015) (fig. 3*c* and *d*, Fisher's test, OR = 6.5, $P < 0.0001$).

Lipids showing human-specific concentration changes (HS-lipids) in the brain cortex were clustered in two specific lipid classes, *glycerophosphoethanolamines* and *diradylglycerols* and their three subclasses (fig. 3*e*). Similarly, enzymes linked to HS-lipids demonstrated an excess of human-specific expression changes in the cortex (two-sided *t*-test, $n = 257$, $P = 0.0004$) (fig. 4*a*), and were significantly overrepresented in 11 KEGG pathways (the Kyoto Encyclopedia of Genes and Genomes; Kanehisa et al. 2017) (hypergeometric test, $n = 511$, BH-corrected $P < 0.0001$, fig. 4*b* and supplementary table 4, Supplementary Material online). Of them, three interlinked pathways, glycerophospholipid metabolism, glycerolipid metabolism, and linoleic acid metabolism pathways, showed the strongest overrepresentation of both protein-coding genes and lipids showing human-specific expression and concentration levels (fig. 4*c*). Notably, enzymes linked to the brain-specific lipids showed a conservation signature compatible with increased purifying selection pressure (fig. 5*a* and *b*), suggesting their functional importance, despite a reduction of the expression level in brain (Wilcoxon test, $n = 131$, $P < 0.0001$, fig. 5*c*).

Discussion

Recent technical advances in measuring lipidome composition in multiple species allow us to assess the evolution at a new level of molecular phenotype: the level of lipid concentrations. Our analysis based on lipidome measurements conducted in six tissues of 32 mammalian species resulted in several remarkable observations. First, in contrast to genetic and gene expression data, where 84.8% and 25.6% of differences among species were proportional to phylogenetic distances, only 1.9% of all lipids' concentrations scaled with the phylogeny (fig. 6*a*). Moreover, unlike genetic and gene expression changes, lipid concentration changes that followed the phylogeny did not show characteristic features of a neutral evolutionary model. Specifically, while much more abundant genetic and gene expression changes did not cluster in any functional categories (hypergeometric test, $n = 5{,}206$, 5,219, 5,120, 5,206, and 5,062 in CX, CB, HT, KD, and LV, respectively, BH-corrected $P > 0.1$, supplementary table 5, Supplementary Material online), the few lipids that followed the phylogeny were grouped in the same particular lipid class, *fatty amides*, in all six tissues. These lipids showed smaller amplitude of concentration differences among species and varied less within species than the rest of the lipidome. Taken together these two observations, the remarkable agreement between tissues, which otherwise have very different lipidome compositions, and reduced concentration variation, suggests the importance of fatty amides in as yet unexplained functions related to an evolutionary "clock-like" lipidome divergence in all tissues. The review of the known features of fatty amides revealed that they are indeed found in all tissues of mammal species (Zoerner et al. 2011). However, the evolution of fatty amides and, in particular, of their most abundant subclass *endocannabinoids*, represented by 15% of all fatty amides detected in six tissues, has been studied at the genetic level only (McPartland et al. 2006; Elphick 2012). Like the pharmacologically active compounds in marijuana or cannabis, endocannabinoids exert their effects by binding to and activating specific cannabinoid receptors CB1 and CB2. The endocannabinoids produce neurobehavioral effects and have key neurotransmitter roles in the central nervous system, especially in the perception of pain, stress, and anxiety, in energy balance and in appetite control (Piomelli 2003). Moreover, endocannabinoids have anti-inflammatory and anticancer properties (De Petrocellis et al. 2000).

We further assessed the possible relationship between the lipidome and species' phenotypes by identifying lipid concentration changes unique to each species. Even though the number of individuals per species was limited, we were able to examine 23 lineages represented by at least three biological replicates per tissue, and detected a significant excess of lipidome changes in three lineages—the human, mouse, and chinchilla ones—as well as a marginally significant trend in the two bat lineages. Notably, the human lineage stood out among the rest by showing the greatest cumulative species-specific divergence. Due to the small numbers of individuals examined in our study, it is likely that many species-specific lipid concentration changes remained undetected. Still, it is noteworthy that among 23 lineages used in the analysis, the human one showed the most pronounced excess of lipid concentration changes. Furthermore, of all species, only humans showed a significant excess of lipid concentration

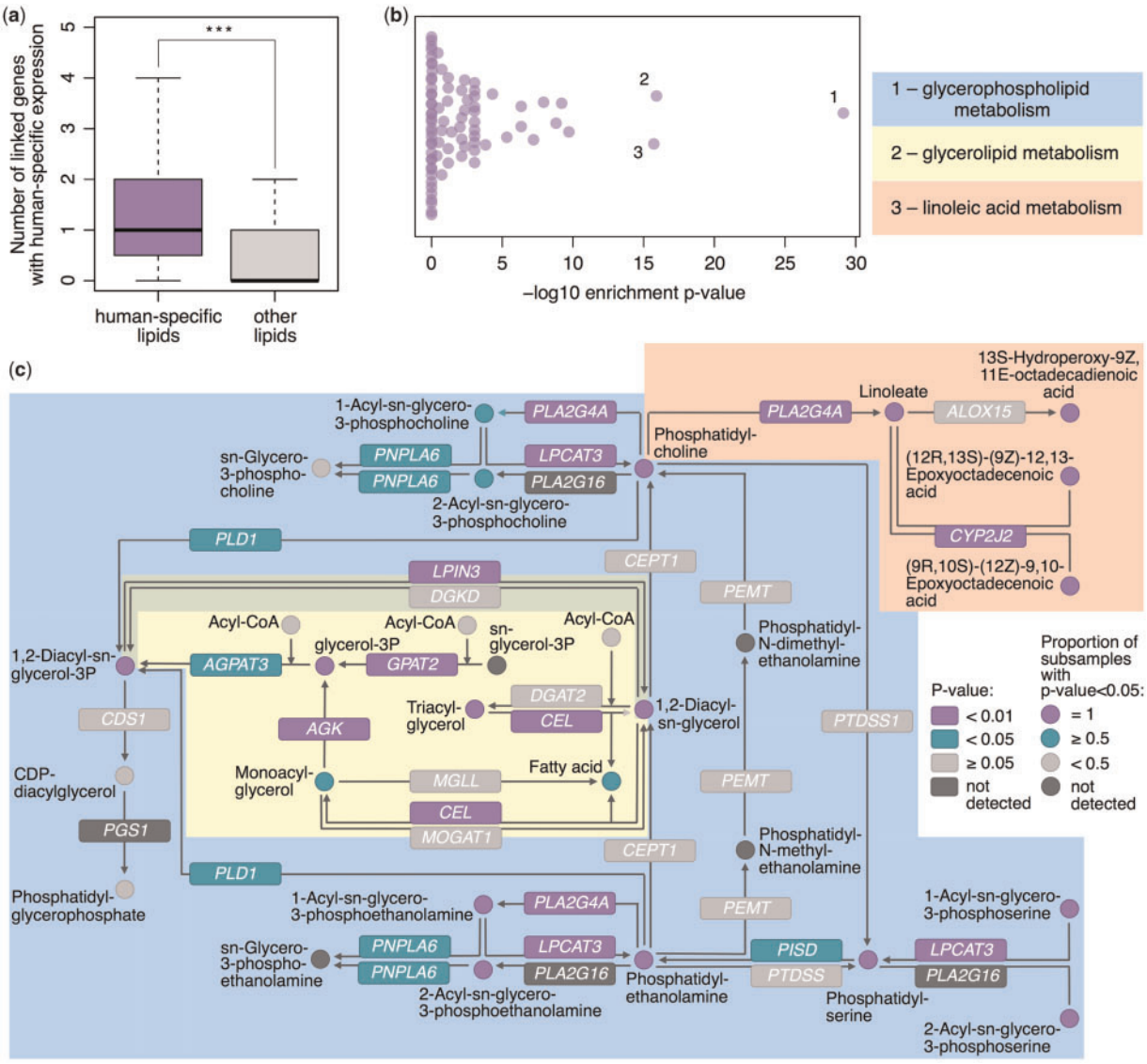

Fig. 4. Characterization of lipids showing human-specific concentration levels (HS-lipids) in cortex. (*a*) Distribution showing numbers of protein-coding genes with human-specific expression in cortex linked to HS-lipids (violet) and control lipids (gray). Stars indicate the significance of the difference between the two distributions (two-tailed *t*-test $P < 0.001$, $n = 257$). (*b*) Enrichment of protein-coding genes linked to HS-lipids in KEGG pathways. Symbols represent pathways. The numbers next to the symbols and the legend above the panel show the top three enriched pathways. (*c*) The simplified schematic representation of the top three enriched KEGG pathways showing HS-lipids and their linked genes. The background colors indicate the pathways, as in panel (*b*) legend.

changes in the brain cortex. This result agrees well with observations of accelerated lipidome evolution in the human brain (Bozek et al. 2015; Li et al. 2017).

Human-specific lipidome changes were most pronounced in the brain cortex and kidney, but only cortical changes clustered in distinct functional pathways, particularly in glycerolipid, glycerophospholipid, and linoleic acid metabolism. These pathways were implicated in a number of neurodegenerative disorders such as Alzheimer's disease (Snowden et al. 2017), Parkinson's disease (Cheng et al. 2011), and neurodegeneration with brain iron accumulation (Morgan et al. 2006), as well as other nervous system disorders: hereditary spastic paraplegia, congenital myasthenic syndrome, Fabry disease, pyridoxine-dependent epilepsy, and Sjögren–Larsson syndrome (Kanehisa et al. 2017).

To conclude, our observations indicate that lipid concentrations evolve differently compared with genome sequences and gene expression levels. We speculate that lipid evolution represents phenotypic differences between species more closely than genomic and gene expression differences do, with a greater proportion of differences accumulating among species representing functional changes shared among tissues (fig. 6*b*), while tissue-specific lipidome differences might represent species-specific adaptations.

Materials and Methods

Source of Samples

Human samples were obtained from the NICHD Brain and Tissue Bank for Developmental Disorders at the University of

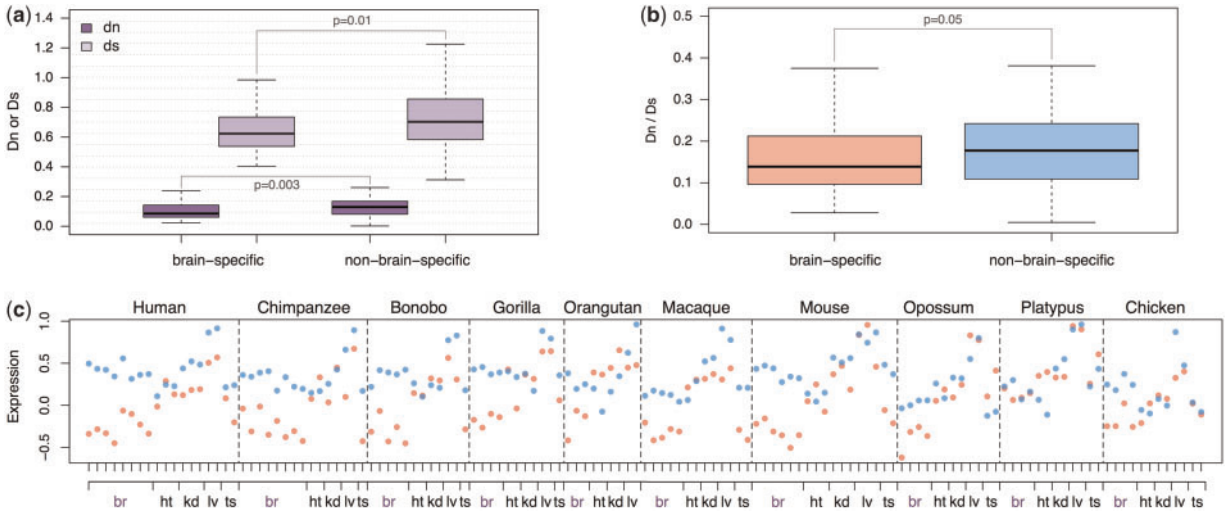

Fig. 5. Evolution of two mutually exclusive groups of lipid-metabolizing enzymes: 1) enzymes linked to brain-specific lipids and 2) enzymes linked to nonbrain-specific lipids. Brain-specific lipids were defined as the 20% of lipids with the highest log2-transformed fold-change of the average concentration difference between brain and the other tissues. Nonbrain-specific lipids were defined as 20% of lipids with the lowest log2-transformed fold-change in the same comparison. (*a*) Dn and Ds values of the two groups of lipid-metabolizing enzymes. (*b*) Dn/Ds ratios of the two groups of lipid-metabolizing enzymes. (*c*) Average expression levels of the two groups of lipid-metabolizing enzymes. The colors indicate the enzymes linked to brain-specific lipids (red) and the enzymes linked to nonbrain-specific lipids (blue).

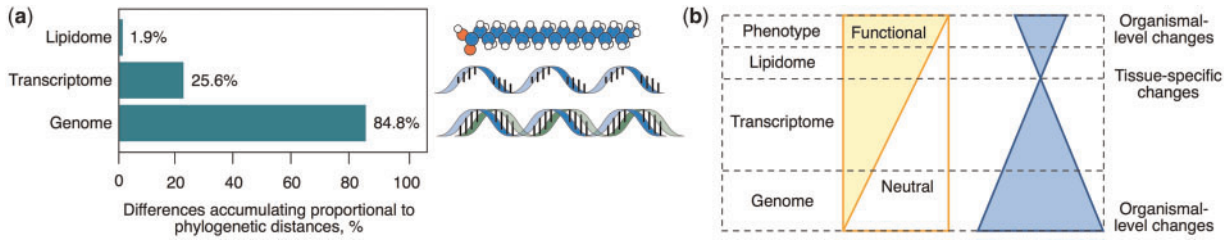

Fig. 6. Evolution at lipidome, transcriptome and genome levels. (*a*) Percentages of lipids, protein-coding transcripts and genes with differences among species accumulating proportionally to phylogenetic distances. (*b*) Schematic representation of the suggested lipidome evolution properties in comparison to the other levels of molecular and organismal phenotype. Left: relative proportions of neutral and functional differences among species. Right: proportion of evolutionary differences shared among tissues. The vertical axis shows the relevant levels of organismal organization. The inversion of this proportion between the transcriptome and lipidome levels reflects the notion that ubiquitously present transcriptome differences are mainly neutral and lipidome ones are mainly functional.

Maryland, the Netherlands Brain Bank, and the Chinese Brain Bank Center (CBBC, http://cbbc.scuec.edu.cn, Wuhan, China). In accordance with the protocols of these institutions, specific permission for brain autopsies and the use of brain tissue for research purposes was given by the donors or their relatives. The use of human autopsy tissue is considered nonhuman subject research and is IRB exempt under NIH guidelines. All subjects were defined as healthy with respect to the sampled tissue by forensic pathologists at the corresponding tissue bank. All subjects suffered sudden death with no prolonged agony state.

Primate samples were obtained from Simian Laboratory Europe (SILABE) in Strasburg, research unit CNRS-MNHM 717 in Brunoy, France, the German Primate Center (DPZ) in Goettingen, the Max Planck Institute for Anthropology in Leipzig, Germany, the Suzhou Experimental Animal Center in China, the Anthropological Institute and Museum of the University of Zürich-Irchel, Switzerland, and the Biomedical Primate Research Centre in the Netherlands. All nonhuman primates used in this study suffered sudden deaths for reasons other than their participation in this study and without any relation to the tissue used.

Rodent samples were obtained from the University of Rochester Biology Department, MDC Berlin, Department of Zoology and Entomology, University of Pretoria, and the animal center at the Shanghai Institute for Biological Sciences. All mice were from the C57/BL6 strain with no genetic modifications. Bat samples were obtained from Kunming Institute of Zoology of the Chinese Academy of Sciences. The use and care of the animals in this research was reviewed and approved by the Biological Research Ethics Committee, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences. All samples were lawfully acquired and their retention and use were in every case in compliance with national and local laws and regulations. Additionally, all nonhuman samples acquisition, retention and use were in every case in accordance with the Institute for Laboratory Animal Research (ILAR) Guide for the Care and Use of Laboratory Animals.

Normalization of Concentrations

To limit potential technical artifacts, mass spectrometric peaks that had zero values in >50% of individuals, were removed. All peak concentration values were then normalized by the internal standard (IS) concentration levels and log2 transformed. Before the log2 transformation, we added 1 to the concentration values in order to avoid infinite values. Missing concentration values were ignored in further analyses. Normalized concentrations were transformed into *Z*-scores for each peak.

Peak Annotation

Peaks were matched to the LIPID MAPS database as described in Fahy et al. (2009). Peaks showing a correlation >0.7 and a difference in RT of <0.05 that matched the same hydrophobic compound in the database search were merged together. This procedure resulted in the merging of 427–880 peaks in different tissues into composite peaks of size of up to six. As a result, the data sets contained between 5,313 and 13,067 compounds in different tissues and ionization modes and between 13,089 (cerebellum) and 20,669 (heart) compounds in both ionization modes together (fig. 1*b*).

We used LIPID MAPS (Fahy et al. 2009), HMDB (Wishart et al. 2009), and KEGG (Kanehisa et al. 2017) database annotations to assign each identified compound to its lipid class, subclass, and pathway. Unannotated peaks were excluded from further analysis (supplementary table 2, Supplementary Material online). About 1,231 peaks were detected and annotated across all six tissues, and were further used to estimate the relationship among samples with a multidimensional scaling algorithm, after quantile normalization of concentration values (fig. 1*c*).

We tested the overrepresentation of lipid classes, subclasses, and pathways in a given group of peaks (lipids) using a hypergeometric test corrected for multiple testing by the Benjamini and Hochberg method (figs. 2*e* and 3*e*).

Identification of Lipids and Genes That Follow Phylogeny

To estimate lipid concentration changes during evolution, we fitted a linear model for each lipid using concentration differences between any two samples as dependent variables and phylogenetic distances for the same samples as independent variables. Phylogenetic distances were obtained from the phylogenetic tree http://www.timetree.org/ (Kumar et al. 2017). Based on the linear models, for each lipid we estimated a coefficient of determination (R^2), which in this case is a measure of the relationship between concentration differences and phylogenetic distances (supplementary fig. 2, Supplementary Material online). Additionally, we quantified the phylogenetic signal in lipid evolution using Blomberg's K as described in Ma et al. (2015). As both approaches yielded similar results (supplementary figs. 3 and 4, Supplementary Material online), we defined phylogeny-dependent lipids as the ones with $R^2 \geq 0.1$ (fig. 2*a*). To determine the validity of this cutoff, we calculated Spearman correlations between concentration differences and phylogenetic distances. All lipids with concentration differences between species following phylogenetic distances demonstrated positive correlation coefficients. Moreover, we obtained qualitatively similar results at a different R^2 cutoff ($R^2 \geq 0.2$, supplementary fig. 5, Supplementary Material online).

To construct the background distribution of R^2 values, we performed 1,000 permutations of species labels across samples within clades, and for all samples as well, and repeated R^2 calculation procedure described above for each permutation (fig. 2*a*). The same procedure was carried out in Blomberg's K distribution analysis (supplementary fig. 4, Supplementary Material online).

Additionally, we performed permutations for individual lipids. Specifically, in each tissue, we performed 1,000 permutations of species labels, while preserving data structure, that is, all individuals representing one species were switched to a new species label together. We then defined the permutation *P* value as the proportion of permutations where the coefficient of determination (R^2) or Blomberg's K was higher than or equal to the original one (supplementary figs. 6 and 7, Supplementary Material online). All lipids defined as phylogeny-dependent by either R^2 or Blomberg's K passed the permutation *P* value cutoff of 0.05, and >93% passed the permutation *P* value cutoff of 0.01 (supplementary figs. 8 and 9, Supplementary Material online).

The exact same analysis was performed for the gene expression data set GSE43013 (Fushan et al. 2015). For the data set preprocessing, genome annotations were obtained from Ensembl, release 90. Paired-end reads were downloaded from SRA and mapped using STAR v2.5.3a with default settings. Average gene expression levels were calculated as FPKM using RSEM v.1.2.31. Orthologs were obtained from Ensembl biomart, release 90. To eliminate unambiguity, genes that had more than one ortholog in the same species were removed. Further, FPKM values were log2 transformed, and quantile normalization was applied. Genes that had missing values in ≥50% of individuals were removed. To compare percentages of phylogeny-dependent protein-coding genes and lipids, we analyzed transcriptome and lipidome data in a matching subset of three tissues of eight species: *H. sapiens*, *P. troglodytes*, *M. mulatta*, *M. musculus*, *R. norvegicus*, *M. auratus*, *C. porcellus*, *H. glaber* (fig. 2*b*).

Additionally, we repeated the same analysis for the gene expression data set GSE30352 (Brawand et al. 2011). To compare percentages of phylogeny-dependent protein-coding genes and lipids, we analyzed transcriptome and lipidome data in a matching subset of four tissues of eight species: *P. troglodytes*, *M. mulatta*, *H. sapiens*, *C. capucinus*, *C. jacchus*, *M. musculus*, *H. glaber*, and *H. armiger* for lipidome data and *P. troglodytes*, *M. mulatta*, *H. sapiens*, *M. musculus*, *G. gorilla*, *P. pygmaeus*, *M. domestica*, and *O. anatinus* for transcriptome data (supplementary fig. 10, Supplementary Material online).

Identification of Protein Coding Sequences That Follow Phylogeny

Orthologous coding sequences for mammal species were obtained from the OrthoMaM database (v9) (Douzery et al. 2014). We estimated pairwise distances between two orthologous sequences using the Jenson–Shannon divergence

measure implemented in *Spaced words* software (Horwege et al. 2014). To fit linear models, the obtained Jenson–Shannon divergences were used as dependent variables, and phylogenetic distances between the same species were used as independent variables. Based on the linear models, for each protein sequence we estimated a coefficient of determination (R^2). Further steps were performed exactly as for lipids (see above, fig. 6*a*).

Analysis of Lipid Concentration Differences within and between Species

To construct the control lipid group, we selected a subset of remaining lipids so as to ensure its size and concentration distribution were the same as those of phylogeny-dependent lipids.

To estimate intraspecies differences of lipid concentrations among individuals, we calculated pairwise Spearman correlations of concentrations between samples derived from the same species (fig. 2*d* and supplementary fig. 12, Supplementary Material online). To estimate interspecies differences of lipid concentrations among individuals, we calculated pairwise Spearman correlations of concentrations between samples derived from different species (fig. 2*c* and supplementary fig. 11, Supplementary Material online). Since evolutionarily distant species might yield lower interspecies Spearman's correlation coefficients than the closely related ones, we additionally normalized the obtained values by the average phylogenetic distances between this species and the rest. For all lipids, we defined intraspecies concentration differences as 1 minus the intraspecies Spearman correlation and interspecies concentration differences as 1 minus the interspecies Spearman correlation (supplementary fig. 15, Supplementary Material online). All calculations were performed for each tissue separately.

To confirm our observations, we repeated the analysis for phylogeny-dependent lipids defined using Blomberg's K as described in Ma et al. (2015) at $K \geq 1.2$ cutoff (supplementary fig. 16, Supplementary Material online).

Identification of Species-Specific Lipids

For each lipid, we compared concentrations in one species against all other species using the Wilcoxon test. If the Wilcoxon test *P* value was lower than the chosen cutoff (0.01), the lipid was defined as species-specific (fig. 3*a*). To balance the number of samples per species, we randomly selected three samples in each species, repeated the procedure 100 times, and calculated the sum number of species-specific lipids among 100 iterations for each species. Because the analysis was applied to each species and each tissue separately, it was limited to 23 out of 32 lineages containing at least three biological replicates in a given tissue. Since evolutionarily distant species might yield higher numbers of species-specific lipids than the closely related ones, we normalized the obtained numbers of species-specific lipids by the average phylogenetic distances between this species and the rest. To test whether the observed number of species-specific lipids was higher than expected by chance, we permuted species labels between samples 100 times, each time calculating the number of species-specific lipids as described earlier. For each species, we then calculated the permutation *P* value as the proportion of permutations resulting in an equal or greater number of significant lipids at the chosen nominal significance cutoff (0.01, fig. 3*b*). The analysis was performed for lipids measured in the negative ionization mode only, which included the majority of detected lipids.

To confirm our observations with an alternative approach, we estimated the number of species-specific lipids using the Ornstein–Uhlenbeck (OU) model, as implemented in R package *ouch* (Butler and King 2004). To find whether changes in optimal lipid concentration have occurred at a particular branch of the phylogenetic tree, we tested the null hypothesis that all branches share the same optimum parameter against the alternative hypothesis that there is a distinct optimum at a different phylogenetic branch corresponding to a particular species. For each lipid, we built an S_0 model (one global optimum across all species) and S_1, S_2, . . ., S_x models (different optimums for particular species *x*). Based on these models, we used the likelihood ratio test between a species-specific selective regime S_x model and the S_0 model. The likelihood ratio between S_x and S_0 had an asymptotic $\chi 2$ distribution (with one degree of freedom) from which a *P* value could be calculated. If the *P* value was lower than the chosen cutoff (0.01), the lipid was defined as species-specific (supplementary fig. 14, Supplementary Material online). To balance the number of samples per species, we randomly selected three samples in each species, repeated the procedure 100 times, and calculated the sum number of species-specific lipids among 100 iterations for each species. To test whether the observed number of species-specific lipids was higher than expected by chance, we performed permutations exactly as described earlier for species-specific lipids defined using the Wilcoxon test.

Comparison with Published Data Set

We defined lipids with human-specific concentration changes as lipids demonstrating Wilcoxon test *P* values (for human-vs.-other-species comparisons) of <0.05 in 95% of bootstraps in one tissue. Bootstraps were performed via the random selection of three individuals per species, repeated 100 times. Lipids with missing values in $\geq 50\%$ of permutations were removed. To compare the remaining lipids with a published data set (Bozek et al. 2015) we selected lipids with a common annotation in both data sets. In cases when multiple LIPIDMAPS IDs were assigned to each lipid, we selected lipids that shared the greatest number of LIPIDMAPS IDs in the two data sets. We plotted concentration distributions for these lipids (fig. 3*d*). Additionally, we calculated log2-fold changes between human lipid concentrations and average lipid concentrations of chimpanzees, macaques, and mice (fig. 3*c*). To check if log2-fold change values were in agreement between data sets, we performed the Fisher's test.

Human-Specific Gene Expression Changes

We defined lipids with human-specific concentration changes as lipids demonstrating Wilcoxon test *P* values (for human-vs.-other-species comparisons) of <0.01 in 75% of bootstraps

(three random individuals per species) in one tissue. We selected enzymes directly linked to the lipids according to the KEGG database (Kanehisa et al. 2017), and defined protein-coding genes with human-specific expression as those showing Wilcoxon test P values of <0.05 for humans vs. other primate comparisons in one tissue, according to gene expression values in the public data set GSE30352 (Brawand et al. 2011) (fig. 4*a*).

We tested the overrepresentation of KEGG pathways in each gene group using a hypergeometric test corrected for multiple testing by the Benjamini and Hochberg method (fig. 4*b* and *c*).

Evolution of Lipid-Metabolic Enzymes

We defined brain-specific lipids as 20% of lipids with the highest log2-fold changes of average concentrations in brain tissues (CB and CX) compared with the remaining tissues ($n = 1504$), and nonbrain-specific lipids as 20% of lipids with the lowest log2-fold changes ($n = 1504$). Then we selected enzymes directly linked to the lipids according to the KEGG database (Kanehisa et al. 2017), and divided them into three groups: 1) enzymes linked to the brain-specific lipids but not to the nonbrain-specific lipids ($n = 65$); 2) enzymes linked to the nonbrain-specific lipids but not to the brain-specific lipids ($n = 117$); and 3) the remaining enzymes linked to the lipids ($n = 85$). The numbers of nonsynonymous substitutions (Dn) and synonymous substitutions (Ds) for these enzymes were obtained from Ensembl Biomart v. 67 (Flicek et al. 2012) (fig. 5*a* and *b*). Expression levels (RPKM) for these enzymes were obtained from Brawand et al. (2011) (fig. 5*c*). RPKM values were log2 transformed, then the quantile normalization was applied, and the values were Z-transformed.

Code Accessibility

The Perl and R source code implementing the analytical procedures described earlier is available at http://cb.skoltech.ru/~khrameeva/evolipids/code.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank I. Burke for her helpful comments on the manuscript. This work was supported by the National Natural Science Foundation of China (grant numbers 31420103920 and 91331203); the Strategic Priority Research Program of the Chinese Academy of Sciences (grant number XDB13010200); the National One Thousand Foreign Experts Plan (grant number WQ20123100078); the Bureau of International Cooperation, the Chinese Academy of Sciences (grant number GJHZ201313); the Russian Science Foundation (grant number 16-14-00220); and the Skoltech Systems Biology Fellowship awarded to E.K. Funding for open access charge: Skolkovo Institute of Science and Technology internal funding.

References

Adibhatla RM, Hatcher JF, Dempsey RJ. 2006. Lipids and lipidomics in brain injury and diseases. *AAPS J* 8(2):E314–E321.

Ariga T, McDonald MP, Yu RK. 2008. Role of ganglioside metabolism in the pathogenesis of Alzheimer's disease—a review. *J Lipid Res.* 49(6):1157–1175.

Bozek K, Khrameeva EE, Reznick J, Omerbašić D, Bennett NC, Lewin GR, Azpurua J, Gorbunova V, Seluanov A, Regnard P, et al. 2017. Lipidome determinants of maximal lifespan in mammals. *Sci Rep.* 7(1):5.

Bozek K, Wei Y, Yan Z, Liu X, Xiong J, Sugimoto M, Tomita M, Pääbo S, Sherwood CC, Hof Patrick R, et al. 2015. Organization and evolution of brain lipidome revealed by large-scale analysis of human, chimpanzee, macaque, and mouse tissues. *Neuron* 85(4):695–702.

Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, et al. 2011. The evolution of gene expression levels in mammalian organs. *Nature* 478(7369):343–348.

Butler MA, King AA. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *Am Nat.* 164(6):683–695.

Cheng D, Jenner AM, Shui G, Cheong WF, Mitchell TW, Nealon JR, Kim WS, McCann H, Wenk MR, Halliday GM, et al. 2011. Lipid pathway alterations in Parkinson's disease primary visual cortex. *PLoS One* 6(2):e17299.

Colsch B, Afonso C, Turpin JC, Portoukalian J, Tabet JC, Baumann N. 2008. Sulfogalactosylceramides in motor and psycho-cognitive adult metachromatic leukodystrophy: relations between clinical, biochemical analysis and molecular aspects. *Biochim Biophys Acta* 1780(3):434–440.

Cutler RG. 1975. Evolution of human longevity and the genetic complexity governing aging rate. *Proc Natl Acad Sci U S A.* 72(11):4664–4668.

De Magalhães JP, Toussaint O. 2002. The evolution of mammalian aging. *Exp Gerontol.* 37(6):769–775.

De Petrocellis L, Melck D, Bisogno T, Di Marzo V. 2000. Endocannabinoids and fatty acid amides in cancer, inflammation and related disorders. *Chem Phys Lipids.* 108(1–2):191–209.

Douzery EJP, Scornavacca C, Romiguier J, Belkhir K, Galtier N, Delsuc F, Ranwez V. 2014. OrthoMaM v8: a database of orthologous exons and coding sequences for comparative genomics in mammals. *Mol Biol Evol.* 31(7):1923–1928.

Elphick MR. 2012. The evolution and comparative neurobiology of endocannabinoid signalling. *Philos Trans R Soc B Biol Sci.* 367(1607):3201–3215.

Eyre-Walker A. 2006. The genomic rate of adaptive evolution. *Trends Ecol Evol.* 21(10):569–575.

Fahy E, Subramaniam S, Murphy RC, Nishijima M, Raetz CRH, Shimizu T, Spener F, van Meer G, Wakelam MJO, Dennis EA, et al. 2009. Update of the LIPID MAPS comprehensive classification system for lipids. *J Lipid Res.* 50(Suppl):S9–S14.

Finch CE, Stanford CB. 2004. Meat-adaptive genes and the evolution of slower aging in humans. *Q Rev Biol.* 79(1):3–50.

Finch CE. 2010. Evolution in health and medicine Sackler colloquium: evolution of the human lifespan and diseases of aging: roles of infection, inflammation, and nutrition. *Proc Natl Acad Sci U S A.* 107(Suppl_1):1718–1724.

Flicek P, Amode M, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S, Fitzgerald S, et al. 2012. Ensembl 2012. *Nucleic Acids Res.* 40(Database issue):D84–D90.

Fushan AA, Turanov AA, Lee S-G, Kim E, Lobanov AV, Yim S, Buffenstein R, Lee S-R, Chang K-T, Rhee H, et al. 2015. Gene expression defines natural changes in mammalian lifespan. *Aging Cell* 14(3):352–365.

Hahn MW. 2008. Toward a selection theory of molecular evolution. *Evolution* 62(2):255–265.

Han X, M Holtzman D, McKeel DW, Kelley J, Morris JC. 2002. Substantial sulfatide deficiency and ceramide elevation in very early Alzheimer's disease: potential role in disease pathogenesis. *J Neurochem.* 82(4):809–818.

Haughey NJ, Bandaru VVR, Bae M, Mattson MP. 2010. Roles for dysfunctional sphingolipid metabolism in Alzheimer's disease neuropathogenesis. *Biochim Biophys Acta* 1801(8):878–886.

Horwege S, Lindner S, Boden M, Hatje K, Kollmar M, Leimeister C-A, Morgenstern Burkhard. 2014. Spaced words and kmacs: fast alignment-free sequence comparison based on inexact word matches. *Nucleic Acids Res.* 42(Web Server issue):W7–W11.

Jordan IK, Mariño-Ramírez L, Koonin EV. 2005. Evolutionary significance of gene expression divergence. *Gene* 345(1):119–126.

Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45(D1):D353–D361.

Kritchevsky D. 1979. Diet, lipid metabolism, and aging. *Fed Proc*. 38(6):2001–2006.

Kumar S, Stecher G, Suleski M, Hedges B. 2017. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol Biol Evol*. 34(7):1812–1819.

Kuraku S, Feiner N, Keeley SD, Hara Y. 2016. Incorporating tree-thinking and evolutionary time scale into developmental biology. *Dev Growth Differ*. 58(1):131–142.

Lamari F, Mochel F, Sedel F, Saudubray JM. 2013. Disorders of phospholipids, sphingolipids and fatty acids biosynthesis: toward a new category of inherited metabolic diseases. *J Inherit Metab Dis*. 36(3):411–425.

Lanfear R, Welch JJ, Bromham L. 2010. Watching the clock: studying variation in rates of molecular evolution between species. *Trends Ecol Evol.* 25(9):495–503.

Li Q, Bozek K, Xu C, Guo Y, Sun J, Pääbo S, Sherwood C, Hof P, Ely J, Li Y, et al. 2017. Changes in lipidome composition during brain development in humans, chimpanzees, and macaque monkeys. *Mol Biol Evol.* 34(5):1155–1166.

Ma S, Yim S, Lee S-G, Kim E, Lee S-R, Chang K-T, Buffenstein R, Lewis KN, Park T, Miller R, et al. 2015. Organization of the mammalian metabolome according to organ function, lineage specialization, and longevity. *Cell Metab.* 22(2):332–343.

McPartland JM, Matias I, Di Marzo V, Glass M. 2006. Evolutionary origins of the endocannabinoid system. *Gene* 370:64–74.

Morgan NV, Westaway SK, Morton Jenny EV, Gregory A, Gissen P, Sonek S, Cangul H, Coryell J, Canham N, Nardocci N, et al. 2006. PLA2G6, encoding a phospholipase A2, is mutated in neurodegenerative disorders with high brain iron. *Nat Genet.* 38(8):957–754.

Nei M. 2007. The new mutation theory of phenotypic evolution. *Proc Natl Acad Sci U S A*. 104(30):12235–12242.

Orr HA. 2009. Fitness and its role in evolutionary genetics. *Nat Rev Genet*. 10(8):531–539.

Otto SP. 2000. Detecting the form of selection from DNA sequence data. *Trends Genet*. 16(12):526–529.

Piomelli D. 2003. The molecular logic of endocannabinoid signalling. *Nat Rev Neurosci*. 4:873–884.

Piomelli D, Astarita G, Rapaka R. 2007. A neuroscientist's guide to lipidomics. *Nat Rev Neurosci*. 8(10):743–754.

Sherman DL, Brophy PJ. 2005. Mechanisms of axon ensheathment and myelin growth. *Nat Rev Neurosci*. 6(9):683–690.

Simons K, Toomre D. 2000. Lipid rafts and signal transduction. *Nat Rev Mol Cell Biol*. 1(1):31–39.

Snowden SG, Ebshiana AA, Hye A, An Y, Pletnikova O, O'Brien R, Troncoso J, Legido-Quigley C, Thambisetty M. 2017. Association between fatty acid metabolism in the brain and Alzheimer disease neuropathology and cognitive performance: a nontargeted metabolomic study. *PLoS Med.* 14(3):e1002266.

Steiner UK, Tuljapurkar S. 2012. Neutral theory for life histories and individual variability in fitness components. *Proc Natl Acad Sci U S A*. 109(12):4684–4689.

Stern DL, Orgogozo V. 2008. The loci of evolution: how predictable is genetic evolution? *Evolution* 62(9):2155–2177.

Wenk MR. 2005. The emerging field of lipidomics. *Nat Rev Drug Discov*. 4(7):594–610.

Wishart DS, Knox C, Guo A, Eisner R, Young N, Gautam B, Hau DD, Psychogios N, Dong E, Bouatra S, et al. 2009. HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res*. 37(Database issue):D603–D610.

Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano Laura A. 2003. The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol*. 20(9):1377–1419.

Zhao G, Guo S, Somel M, Khaitovich P. 2014. Evolution of human longevity uncoupled from caloric restriction mechanisms. *PLoS One* 9(1):e84117.

Zoerner AA, Gutzki F-M, Batkai S, May M, Rakers C, Engeli S, Jordan J, Tsikas Dimitrios. 2011. Quantification of endocannabinoids in biological systems by chromatography and mass spectrometry: a comprehensive review from an analytical and biological perspective. *Biochim Biophys Acta* 1811(11):706–723.