Selection pressure drives the co-evolution of several lipid metabolism genes in mammals

LIN BenFu[1], YUAN LiHong[1,2*] & CHEN JinPing[1,2]

[1]*South China Institute of Endangered Animals, Guangzhou 510260, China*
[2]*Key Laboratory of Marine Bio-resources Sustainable Utilization (LMB); Key Laboratory of Applied Marine Biology of Guangdong Procince and Chinese Academy of Sciences (LAMB), South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, 510301, China*

Adipose tissue is an important endocrine organ and energy supplier. Its physiological effect on the regulation of the energy balance is considered an important factor underlying the evolution of mammals. To test whether the genes controlling lipid metabolism have undergone adaptive molecular change in the evolution of mammals, in this study, we used the orthologous gene sequences of 12 important lipid metabolism proteins (*leptin*, *OB-RL*, *RXRA*, *RXRB*, *RXRG*, *PPARA*, *PPARB/D*, *PPARG*, *PNLIP*, *ADIPOQ*, *LPL* and *UCP1*) from NCBI's databases. We found evidence that 4 of the corresponding genes (*leptin*, *ADIPOQ*, *PNLIP* and *PPARA*) have undergone positive selection in their evolutionary history and that most adaptive changes occurred during the evolution of the super-clades Laurasiatheria (placentals) and suborders within Euarchontoglires (primates and rodents). Comparisons across sets of genes showed that in a third of cases, bursts of positive selection, more than would be expected by chance, occurred on corresponding branches. We propose that the positive selection drives adaptive changes in some lipid metabolism genes in or within Laurasiatheria and Euarchontoglires clades. Along with evidence from earlier studies, our results show that co-evolution among interacting lipid metabolism proteins has taken place.

protein evolution; lipid metabolism; Darwinian selection, positive selection

The first applications of molecular data to understand co-evolution typically focused on parallel cladogenesis between parasites and their hosts [1]. The later proliferation of sequence data allowed this approach to be extended to the study of co-evolution between interacting functional genes and/or proteins [2,3]. Such examples of co-evolution include the proteins expressed on male and female gametes, the rapid evolution of which might result from sexual conflict and sexual selection [4,5]. By grouping proteins based on their expression profiles, Marcotte et al. [6] have examined the co-evolution of protein expression. Fraser et al. [7] quantified codon usage bias to infer evidence of co-evolution in gene expression levels based on the genomes of four yeast species, while Goh et al. [8] found correlations between pairwise sequence distances of two protein binding domains in lower organisms, and also between chemokines and their receptor in mammals. More recently, McPartland et al. [9] reported parallel phylogenies between the genes of cannabinoid receptors and their ligands, and also found correlations between the substitution rates of corresponding branches across phylogenies. Here, we test for co-adaptation among genes that encode a suite of interacting proteins involved in lipid metabolism. These proteins are considered to play an essential role in the maintenance of a constant body temperature [10–14], an important trait underpinning the evolution of mammals.

Homeostatic thermal control by the central nervous system and peripheral organs has allowed mammals to exploit cold environments by both promoting survival and allowing activity during periods of seasonal and nocturnal cold [15].

*Corresponding author (email: yuanlh@gdei.gd.cn)

Adipose tissue is a dynamic endocrine organ that plays a key role in energy storage and in the regulation of both body temperature and energy balance [16]. Mammals possess two different forms of adipose tissue, white and brown. The former acts more as a source of energy and provides thermal and physical insulation [17]. Brown adipose tissue functions in the conversion of energy to heat by non-shivering thermogenesis [18,19] and is particularly associated with cold-adapted species, neonates and hibernators [20]. It has been speculated that brown adipose tissue probably arose from a single evolutionary event in the early evolution of mammals [15]. Although morphologically differentiated, recent evidence has shown that under certain conditions white adipocytes can *trans*-differentiate into brown adipocytes, and *vice versa*, to meet the metabolic needs with changing of environment [21].

To examine whether lipometabolic genes have experienced adaptive evolution associated with the radiation of mammals, we selected 12 important lipid metabolic molecules that interact with each other in a common biochemical pathway and obtained the orthologous sequences of the corresponding genes in a large range of mammals from NCBI databases. The known interactions among the protein products of these genes, and their homeostatic effect on the body's energy supply and temperature, are summarized in Figure 1 and Table S1. Briefly, the protein hormone leptin, encoded by the obese (*ob*) gene, is an essential regulator of lipid metabolism, communicating with both the CNS and other hormones [22–24]. Its effect on energy homeostasis and thermogenesis is mainly mediated by binding with the leptin receptor (OB-R), the product of the diabetes (*db*) gene [25]. Studies of obese mice indicated that abnormal splicing of the leptin receptor gene splice variant OB-Rb is likely to lead to defective signal transduction and therefore will inhibit the normal regulation of lipid metabolism by leptin [26]. Peroxisome proliferator-activated receptors (PPARs) are members of the superfamily of nuclear hormone receptors and play essential roles in regulating lipid homeostasis in the body by forming heterodimers with retinoid X receptors (RXRs) [27]. Several studies have shown that PPARα can stimulate the β-oxidation of fatty acids [28], while PPARγ and PPARβ/δ appear to be mainly involved in regulating the process of adipocyte differentiation [29,30]. Lipoprotein lipase (LPL) and adiponectin (ADIPOQ) are both regulated by complexes of PPARs/RXRs and function in lipid metabolism [12,27]. Uncoupling protein 1 (UCP1) is expressed specifically in the brown adipose tissue of mammals, is involved in classic non-shivering thermogenesis, and might be regulated by leptin and complexes of PPARs/RXRs [10,15,22]. Finally, pancreatic lipase (PNLIP) is mainly involved in transforming dietary fat into fatty acids [31].

We constructed phylogenetic trees for each set of orthologs and derived maximum-likelihood estimates of substitution rates along corresponding branches. We tested for concordant bursts of adaptive change among different genes that might be expected if molecular adaptation in one protein has consequences for its interacting partners. This approach is broadly similar to that used by Fraser et al. [7], but includes more taxa. We also developed a randomization routine that accounts for false positives in the data.

1 Materials and methods

1.1 Genes collection

We identified 12 key genes involved in lipid metabolism using BLASTN, and downloaded their coding domain sequences from NCBI nucleotide collection (nr) and whole-genome shotgun (wgs) databases [32]. The number of taxa for which orthologous sequences were available ranged from 15 for $OB\text{-}R_L$ (the long form of the leptin receptor) to 28 for *leptin* (Table S2).

1.2 Phylogenetic tree reconstruction

For each gene, the nucleotide sequences were aligned using ClustalX 1.81 [33], and the alignment was used to generate amino acid alignment with MEGA3.1 [34]. Topology trees for the 12 selected genes were constructed based on the results of large-scale studies of mammal evolution [35,36] and used to analyze the maximum likelihood of codon substitution (Figure S1). For each tree, we calculated maximum-likelihood estimates of the number of synonymous and non-synonymous substitutions per site (dS and dN) and the ratio (ω) of these rates (dN/dS) using the PAML 3.15 software package [37]. Substitution rates were estimated for each branch in each tree under a free-ratio model in which the rates were assumed to be independent across branches. We compared the likelihood to a model in which the ratio was fixed across the tree (M0, one-ratio model) and, to test for significance, we obtained a log-likelihood ratio test
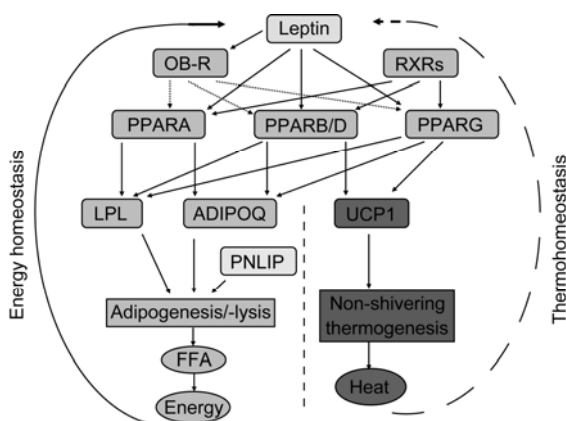

Figure 1 Schematic diagram representing the relationship between the lipid metabolic proteins that regulate the homeostasis of body temperature and energy. Arrows indicate stimulatory effects. Arrows with broken lines indicate stimulatory effects that are less well understood.

statistic ($2\Delta\ell$) which follows a χ^2 distribution.

1.3 Site models of molecular evolution

We tested for variable selective pressures among codons by assuming different classes of sites in the genes with different ratios of non-synonymous to synonymous substitution rates [38]. To examine the selective pressure acting on these genes, six different PAML site models were implemented. The null model (M0, one-ratio model) assumes the same ω ratio for all sites in the gene (and all branches in the phylogeny); the nearly neutral model (M1a) estimates two ω values ($0 < \omega_0 < 1$, $\omega_1 = 1$); the positive selection model (M2a) adds an additional ω value to M1a which allows ω_2 value, if present, greater than one; the discrete model (M3) extends from M0 and has two unconstrained ω values (ω_1 and ω_2); M8 (beta and ω model) and M8a (fixed $\omega = 1$) are extensions of M7 which constrains the ω ratio smaller than one following the β distribution and takes into account possible positively selected (PS) sites [37]. Three pairs of models were compared (M2a and M1a, M3 and M0, and M8 and M8a) and significant model improvement was assessed using the likelihood ratio test (LRT). Positive selection was inferred when the ω parameter of M3 or M8 was greater than one and the LRT was significant. To calculate the posterior probability of a site falling into each site class, a Markov Chain Monte Carlo (MCMC) analysis was undertaken in MrBayes program [39]; sites with high probabilities for $\omega > 1$ were considered to be likely under positive selection under the M3 and M8 models [40].

1.4 Tests for molecular co-evolution

To test for co-evolution between genes, we compared the distribution of substitution rates across species trees reconstructed for the same sets of taxa. Since the number of comparable gene trees (with the same taxa) will decrease as more taxa were included, we repeated the phylogenetic analyses for 9 different sets of taxa and genes (Table 1). The membership of each reduced dataset is described in Table S2. Maximum-likelihood estimates are known to rely on the phylogenetic relationship of the sequences under consideration, so repeating the analysis with several datasets provided an additional means of verifying our results.

We used a randomization procedure modified from that described by Rossiter et al. [41] to statistically test for evidence of co-evolution between the genes. For each tree, we first constructed an *m-x-n* matrix, where *m* was the number of genes used to construct the tree, and *n* was the number of branches in the tree, recording for each matrix entry the ω value for the *i*th gene and *j*th branch, $\omega_{i,j}$. In a matrix of this kind, co-evolution between genes is reflected in a tendency for the higher ω values for the different genes to occur in the same branch or branches increasing the variance (S^2) in

the sum of the ω values for each branch as shown below.

$$S^2 = \frac{1}{n-1}\left(\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{m}\omega_{i,j} - \sum_{i=1}^{m}\omega_{i,j}\right)^2.$$

To test whether S^2 was greater than that expected by chance, we randomized the $\omega_{i,j}$ value within genes which, in effect, randomly reassigning evidence for positive selection to different branches of the tree while preserving the distribution of ω values for each gene. We then recalculated S^2 for the new, randomized data sets 100000 times (S^2_{rand}), and compared the actual S^2 statistic to the random distribution. The *p* value associated with the test is simply the proportion of randomizations for which $S^2 > S^2_{rand}$.

Randomizations of this sort have the advantage that any imperfections in the original data set, for example, spuriously high values for ω, non-normal distributions or small sample sizes, are retained in each of the randomizations, as are the ω values for each gene. Thus, significant departures from the randomized distributions only occur when ω values are positively assorted within branches of the tree in question. In addition, for the significant datasets, the branch-specific model (two-ratio) and the branch-site model (Model A) were implemented to test for changes of selective pressure and the potential positive sites in specific lineages in the ancestor of Laurasiatheria and/or Euarchontoglires. The branch-specific model assigned two ω ratios for the foreground (ω_1) and background (ω_0) which allows for variable ω ratios among branches but invariable ω ratios in sites in the tree [42]. Model A assigns three ω ratios ($0 < \omega_0 < 1$, $\omega_1 = 1$, and ω_2 is variable among sites) and tests the positive sites along the fixed lineage [40]. Then, the branch-specific model was compared with M0, Model A vs. Model A′ (test II) and Model A vs. M1a (test I) and the LRT test was performed.

To further explore the patterns of co-evolution among genes, we undertook a principal component analysis (PCA) using the S-PLUS 2000 software (MathSoft Inc.). Components were generated based on estimates of ω under a free-ratio model for 12 sets of corresponding branches (variables) (dataset *b* in Table 1).

1.5 Protein structure analysis

We performed protein structure predictions for all the genes that were found to be under positive selection, using the *Homo sapiens* sequence as a reference in each case. The signal peptide was predicted using the SignalP tool (http: //www.cbs.dtu.dk/services/SignalP/), and the functional amino acid motifs were predicted using the MotifScan program in the PROSITE database (http: //www.expasy.org/ prosite) [11]. Sites that were predicted to have undergone positive selection were mapped onto the tertiary structure of the orthologous human proteins from the Protein Data Bank (http://www.rcsb.org/pdb/home/home.do).

Table 1 Randomization tests of concordant bursts in the nine datasets of gene trees[a]

Code for gene tree dataset	Number of genes	Number of taxa	Branches	*P* value
a	12	13	22	0.256
b	12	9	15	0.00009***
c	6	18	29	0.01045*
d	11	13	22	0.11423
e	10	14	24	0.06229
f	9	15	25	0.06718
g	6	18	29	0.53121
h	6	18	29	0.2911
i	4	19	30	0.01018*

a) For details of the taxa and genes present in each analysis see Table S2. * indicates $P < 0.05$; *** indicates $P < 0.001$.

2 Results

2.1 Concordant evolution analysis of lipid metabolic genes

We constructed separate phylogenetic trees for 12 key genes involved in lipid metabolism in 11 to 25 different mammal species (Tables 2 and S2) in which the ω ratio was allowed to vary among different branches provided a significantly better fit to the data than a one-ratio model (M0) in which the ω ratio was fixed (Table 2). This result revealed that selection pressure varied across the trees. Moreover, average ω values based on the one-ratio model ranged from 0.016 for the *RXRA* tree to 0.312 for the *OB-R$_L$* tree,

indicating that overall selection pressure during mammalian evolution also varied widely for these genes (Table 2).

As the number of comparable gene trees (with the same taxa) decreased as more taxa were included, we selected 9 sets of gene trees (those with the highest possible number of genes and taxa) to tested for evidence of co-evolution of the genes by comparing branch-specific estimates of ω across the trees. The number of gene trees within each set ranged from 4 to 12 and the number of taxa ranged from nine to 19 (Tables 1 and S2). Randomization tests provided evidence of concordant bursts in positive selection in three out of nine sets of gene trees (Table 1). In each of these, high ω estimates were found to be clustered on corresponding branches, leading to higher than expected variance among sums of branch-specific ω values (Figure 2 and Table S3).

In dataset *b* (12 genes, 9 taxa, 15 branches), parallel cases of $\omega > 1$ across the species tree were observed on the ancestral branch leading to the two ungulates (horse and cow) and on the ancestral branch leading to members of the Euarchontoglires (rodents and primates, Figure 2(a)). Dataset *c* (6 genes, 18 taxa, 29 branches) also revealed a parallel signature of positive selection leading to the Euarchontoglires as well as on the ancestral branch leading to members of the Laurasiatheria (ungulates, shrew, bat, dog and cat) and on separate branches leading to the orangutan and the chimpanzee (Figure 2(b)). Finally, dataset *i* (4 genes, 19 taxa and 30 branches) showed evidence of parallel positive selection across multiple gene trees in the ancestral

Table 2 Likelihood ratio test and parameters of branch and site-specific models

Gene	*N*[a]	*Lc*[b]	dN/dS[c]	Branch model tests[d] 2Δℓ free-ratio *vs* M0	Site model tests[d] 2Δℓ M2a *vs* M1a (df = 1)	2Δℓ M3 *vs* M0 (df = 4)	2Δℓ M8 *vs* M8a (df = 1)	ω_2 ratio (p_2) of M3
Leptin	25	164	0.307	129.67** (df = 45)	0	191.83**	7.30**	$\omega_2 = 1.523$ ($p_2 = 0.088$) $\omega_2 = 1.52325$
OB-R$_L$	11	1134	0.312	16.58 (df = 18)	0	193.14**	0	$\omega_2 = 0.829$ ($p_2 = 0.274$) $\omega_2 = 0.82848$
RXRA	13	371	0.016	161.80** (df = 22)	0	28.9**	16.59**	$\omega_2 = 0.064$ ($p_2 = 0.267$) $\omega_2 = 0.06419$
RXRB	15	382	0.023	28.26 (df = 26)	0	47.80**	5.16*	$\omega_2 = 0.849$ ($p_2 = 0.013$) $\omega_2 = 0.84912$
RXRG	13	397	0.040	39.29* (df = 22)	0	25.55**	0.08	$\omega_2 = 0.245$ ($p_2 = 0.099$) $\omega_2 = 0.24478$
PPARA	14	467	0.045	84.15** (df = 24)	0	161.35**	0.30	$\omega_2 = 0.874$ ($p_2 = 0.008$) $\omega_2 = 0.87403$
PPARG	21	361	0.025	36.09 (df = 38)	0	92.83**	0.15	$\omega_2 = 0.864$ ($p_2 = 0.0103$) 0.0103) $\omega_2 = 0.86391$
PARB/D	12	438	0.048	40.96** (df = 20)	0	96.69**	0.04	$\omega_2 = 0.246$ ($p_2 = 0.185$) $\omega_2 = 0.24606$
ADIPOQ	21	193	0.169	42.99 (df = 36)	3.13	378.82**	1.13	
UCP1	20	295	0.157	91.05** (df = 33)	0	297.52**	1.32	$\omega_2 = 0.611$ ($p_2 = 0.135$) $\omega_2 = 0.61091$
PNLIP	14	459	0.243	48.44** (df = 23)	0	749.21**	9.49**	$\omega_2 = 0.961$ ($p_2 = 0.225$)
LPL	16	463	0.073	64.07** (df = 28)	0	134.94**	0.13	$\omega_2 = 0.891$ ($p_2 = 0.018$) $\omega_2 = 0.89114$

a) *N*, number of sequences in the full gene tree. b) *Lc*, number of codons after alignment with the human reference sequence and after gap removal. c) dN/dS (ω), was estimated for model M0 (one-ratio) which assumes the same ω ratio for all branches in the full phylogeny gene tree and for all sites in the gene. d) Significance levels for likelihood ratio tests (2Δℓ); * indicates $P < 0.05$ and ** indicates $P < 0.01$.
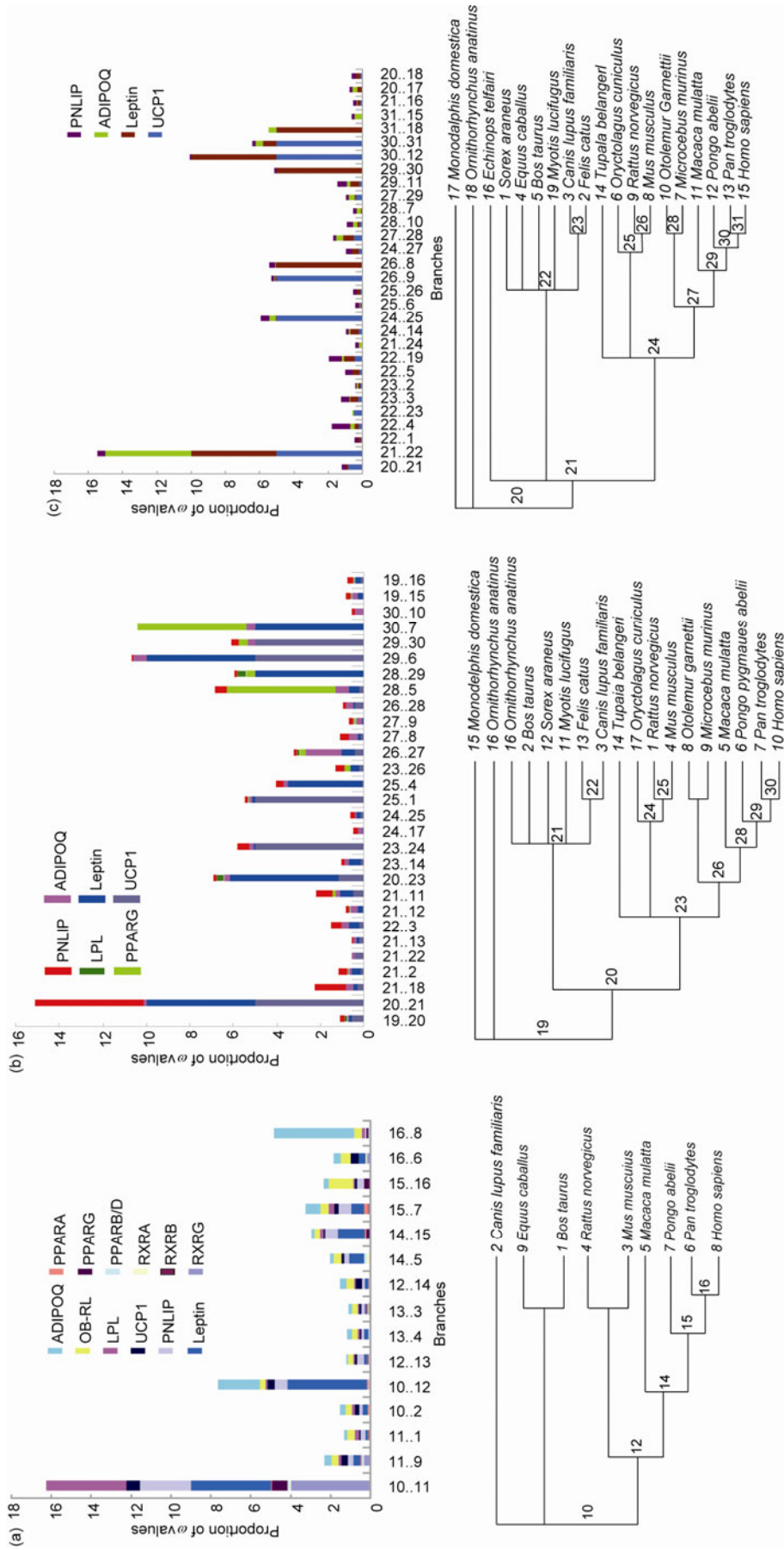
Figure 2 Concordant evolution analyses of lipid metabolic genes. (a) The 12 selected genes in nineteen species. (b) Six of the selected genes in nine mammalian species. (c) Four of the selected genes in eighteen species. The *X* axis indicates the branches of the phylogenic trees; the *Y* axis indicates the proportion of the ω value for each gene on the corresponding branches. More detailed information is available in Table S3.

branches leading to both the Laurasiatheria and to the orangutan (Figure 2(c)).

A closer examination revealed that the different genes varied markedly in their contribution to these cases of branch-specific signatures of concordant positive selection (Figure 2). For example, *leptin* contributed to both cases evident in dataset *b*. Similarly, in dataset *c*, *leptin*, *UCP1*, *PNLIP* and *PPARG* contributed to 4, 3, 1 and 1 cases, respectively, whereas in dataset *i*, both *leptin* and *UCP1* contributed to each case, and *ADIPOQ* contributed for one case. The branch-specific model showed that three cases along the ungulate lineage in dataset b, two cases in the ancestor of Euarchontoglires lineage in dataset c, and three cases in the ancestor of Laurasiatheria in dataset i with $\omega > 1$, are consistent with the randomization test. Potential positive sites were identified by Bayes Empirical Bayes (BEB) method of Model A in all of 3 datasets, although the LRT tests (test I and test II) were not significant (Table S4).

A principal component analysis (PCA) using the ω values from dataset *b* provided additional insights into the contribution of the different lipid metabolism genes to signatures of selection and co-evolution (Figure 3). The plotted position of *leptin*, for example, is midway between two vector loadings for the two ancestral branches (ungulates and Euarchontoglires) where it shows evidence of positive selection. The PCA also revealed the close association of some lipid protein genes such as *PNLIP* and *ADIPOQ* based on correlated branch-specific ω values, even for $\omega < 1$.

2.2 Site models of selection

Site models for individual genes provided further evidence of heterogeneous selective pressure. Likelihood ratio tests constructed for two model comparisons were significant for 4 of the 12 genes (Table 2). Specifically, M8 (beta + ω) fitted the data significantly better than M8a (fixed for $\omega = 1$) for the genes *leptin* ($\omega_2 = 2.46$), *RXRA* ($\omega_2 = 1.00$), *RXRB* ($\omega_2 = 1.00$), and *PNLIP* ($\omega_2 = 1.93$). A third LRT compared

Figure 3   A principal component analysis based on the ω estimates for 14 sets of corresponding branches for the 12 genes and 9 mammalian taxa.

the discrete model (M3) with two classes of freely estimated ω values to the one-ratio model (M0) and was significant in all genes. Moreover, M3 was associated with a freely estimated $\omega_2 > 1$, indicating potential positive selection for *leptin* and *ADIPOQ*. In *leptin*, 13 sites were identified as being potentially under positive selection based on BEB thresholds, comprising three sites with $P > 0.95$ (8G, 25Q, 119W), seven sites with $0.95 > P > 0.8$ (17F, 19V, 43N, 49S, 90R, 130G, 156W) and three sites with $0.8 > P > 0.65$ (4G, 15Y, 93I). In *ADIPOQ*, we identified 19 sites as potentially under positive selection, comprising 5 sites with $P > 0.95$ (8H, 13T, 16L, 59I, 192V), 9 sites with $0.95 > P > 0.8$ (11T, 12T, 14Q, 56P, 68T, 82I, 92G, 107Y, 153M) and 5 sites with $0.8 > P > 0.57$ (7G, 15P, 22T, 130G, 167M). In 10 of the lipid metabolism genes, *leptin*, *ADIPOQ*, *OB-R$_L$*, *PNLIP*, *LPL*, *PARB/D*, *PPARG*, *RXRB*, *UCP1*, and *PPARA*, 10, 18, 14, 12, 7, 2, 2, 2, 1 and 1 positive selected sites respectively, were identified by M8. All sites inferred to be under positive selection are listed in Table S5: Summary of positive selected sites. The M3 results for all 12 genes were verified using the MCMC method and 4 genes, *leptin*, *ADIPOQ*, *PNLIP* and *PPARA*, were identified as having positive selection sites. The corresponding parameters are given in Tables 3 and S5: Summary of positive selected sites. Positive sites based on the M3 model are shown on the tertiary structures of the human leptin and ADIPOQ proteins (Figure 4).

3   Discussion

The evolution of mammals has been partly attributed to the evolution and metabolism of adipose tissue which provides a means of homeostatic control and thus survival in the face of cold climates and food shortages [43]. To determine whether such genes have undergone adaptive evolution, both independently and in parallel with each other, we compared patterns of selection evolution across 12 genes involved in mammalian lipid metabolism.
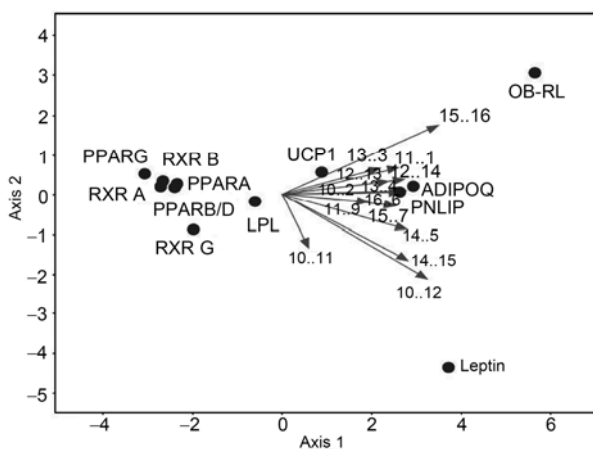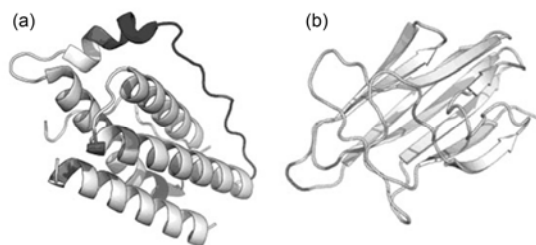

Figure 4   Predicted tertiary structure of the proteins (a) Leptin and (b) ADIPOQ, constructed using the human sequence as a reference and with the peptide signal removed. (a) The predicted structure of leptin is based on the human leptin crystal structure (pdb: 1AX8) [54]; positively selected sites are shown in red and the segments SCHLPWASGLETLDSL and QLDLS are shown in blue (see Discussion for details). (b) The predicted structure of ADIPOQ is also based on the human ADIPOQ ortholog. The shown structure is the C-terminal domain from residue 107 to the end.

Table 3 Posterior probability for the selected sites identified by the MCMC analysis[a]

Gene	N	Amino acid	Posterior probability
Leptin	25	8G	0.95*
		17F	0.73
		25Q	0.99**
		93I	0.77
		119W	0.81
ADIPOQ	21	11T	0.99**
PNLIP	14	33T	0.98*
		71S	0.96*
		126G	0.87
		151A	0.82
		332D	0.83
		438K	0.90
PPARA	14	96G	0.82

a) Each position (numbered using the human orthologs of the protein as the reference with the gaps removed) with a significant posterior probability is shown here. In each case, posterior distributions were obtained using MCMC sampling. Samples were drawn every 20 steps from a total of 1000000 steps following a discarded burn-in of 100000 steps. Acceptable mixing and convergence to the stationary distribution were checked by inspection and by plotting of posterior samples. N, number of sequences; * indicates $P > 0.95$; ** indicates $P > 0.99$.

Evidence of parallel positive selection was particularly apparent between *leptin* and *PNLIP* in the Laurasiatheria (datasets *b* and *c*, respectively), between *leptin* and *ADIPOQ* in the ancestral branch of the superorders Euarchontoglires and Laurasiatheria (datasets *b* and *i*), and between *leptin* and *UCP1* in the same branches (datasets *c* and *i*, respectively). Many previous studies have shown that positive selection drove the adaptive evolution of *leptin* in many lineages in Haplorrhini [44–47] and of *UCP1* in Eutheria [48,49]. These results are similar to our findings that positive selection occurred in the superorders of Laurasiatheria and Euarchontoglires and in the suborder of Haplorrhini. Matsubara et al. [50] reported a significant negative correlation between *in vivo* plasma concentrations of ADIPOQ and leptin in humans, suggesting that the expression of these proteins is inversely related. Similarly, Cancello et al. [51] found that, at mRNA level, *leptin* and *UCP1* were reciprocally expressed in the brown adipose tissue of mice, and that their regulation was related to the thermogenic activity of brown adipocytes. However, we found no clear relationship between the co-evolution of genes and the direct interaction of their protein products during lipid metabolism (Figure 1). For example, leptin showed no parallel signature of selection with its receptor *OB-R*, which are known to interact to modulate energy homeostasis and thermogenesis by influencing food intake and fat metabolism [52,53]. This is different from the results of a study of cannabinoid receptors and their ligands, which do show evidence of parallel selection [9]. Moreover, as the limited data that was available, we only examined the co-evolution of lipid genes in mammals; with the enrichment of the databases, it should become possible to investigate the evolutionary role of the lipid genes in the acquisition of warm-bloodedness.

In addition to evidence of co-evolution, the majority of genes independently showed positive selection based on site models. BEB and MCMC analyses revealed that most amino acid residues likely to be under positive selection occur in leptin and ADIPOQ and, to a lesser extent, in PNLIP and PPARA (Table S5).

Zhang et al. [54] reported that the crystal structure of a human mutant leptin protein comprised 4 anti-parallel α-helices (A, B, C and D) connected by two long crossover links (AB and CD) and one short loop (BC) (Figure 4(a)). Five positive selection sites were located in the signal peptide sequence (Table S5). Structural analysis has shown that the 4 α-helices (A, B, C and D) form a large hydrophobic cylindrical core that may help to maintain the structural integrity of the protein. The α-helices also contain the major and minor interfaces (A and C helixes) for the interaction of leptin with its receptor (OB-R). Thus, the positively selected sites 25Q and 43N in helix A, and 90R and 93I in helix C, might play a role in leptin/OB-R binding. Indeed, given that the main function of the minor interface seems to stabilize the leptin/OB-R complex [26,55], positive selection in the A and C helices could support the ability of leptin to interact with multiple receptors thereby regulating metabolism. The structural analysis has also shown that a kinked 3_{10} helix formed by 5 residues (QLDLS) at the end of helix D is important for protein folding and receptor binding [54] and thus site 156W may help leptin/receptor binding and stabilize the conformation of the complex. Other sites were found in the CD loop (119W and 130G), the SCHLPWASGLETLDSL segment of which is known to regulate energy metabolism by reducing food intake in ob/ob mice [56]. A previous study of *leptin* gene sequences in several species of Ochotonidae (pika) (order Lagomorpha) also found evidence of positive selection, and it was suggested that non-synonymous substitution rates were ecological adaptations linked to cold weather [11]. Although the LRT statistics of two model comparisons (M1a/Model A and M3/Model B) proved nonsignificant, 20 sites were identified as being under positive selection in the pika lineage, two of them (Q25W and S49A) were also identified in the present study (Table S5).

Pancreatic lipase (PNLIP) has 2 functional domains and is mainly involved in transforming dietary fat into fatty acids. We found 6 sites that were under positive selection in *PNLIP* based on MCMC analysis, 5 of them occurred in the N-terminal domain and one in the C-terminal domain (Tables 3 and S5). Three of the N-terminal sites (126G, 151A and 332D) were covered by two loops and a lid-domain which are known to have important functions in supporting the interaction of the protein with its substrate (lipase and serine) [57]. Another site (438K) in a PLAT domain might also help to regulate the lipo-metabolic function of PNLIP (Table S5).

A similar distribution of positive selected sites was found

in ADIPOQ, which is regulated by the PPARs/RXR complex and functions in lipolysis [12,27]. As a hormone specifically secreted by adipocytes, ADIPOQ helps to maintain the homeostasis of lipid and glucose [58]. Of 19 sites identified as being under positive selection, 9 sites occur in the N-terminal domain and 5 are in the C-terminal domain (Figures 4(b) and S5). Regions within the N-terminal are known to have roles in the assembly of higher order oligomeric structures (including homotrimers) [59], and regions in the C-terminal are involved in correct folding and alignment for protein-protein recognition [60] (Table S5).

4 Conclusion

Mammals are found on all continents and are able to survive a range of extreme environments. Our finding that several mammalian lipid metabolism genes show signatures of adaptive selection in important functional domains appears to support the view that the evolution of mammals is partly due to their ability to regulate lipid metabolism according to their energy and thermal needs. More generally, however, cases of positive selection detected in this larger scale study do not appear to be related to particular aspects of ecology of extant taxa. Instead, most evidence corresponds to ancestral branches in the radiations of several major clades. More comparative data are needed on patterns of lipid metabolism from representatives from these clades to assess whether bursts of selection reported here are related to divergence in the metabolic pathways of these groups. Alternatively, changes in selection pressure early in the evolution of these mammalian groups could reflect features about the ecology of the ancestral form.

1 Hafner M S, Nadler S A. Phylogenetic trees support the coevolution of parasites and their hosts. Nature, 1988, 332: 258–259
2 Fryxell K J. The coevolution of gene family trees. Trends Genet, 1996, 12: 364–369
3 Pellegrini M, Marcotte E M, Thompson M J, et al. Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. Proc Natl Acad Sci USA, 1999, 96: 4285–4288
4 Swanson W J, Yang Z, Wolfner M F, et al. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proc Natl Acad Sci USA, 2001, 98: 2509–2514
5 Panhuis T M, Clark N L, Swanson W J. Rapid evolution of reproductive proteins in abalone and *Drosophila*. Philos Trans R Soc Lond B Biol Sci, 2006, 361: 261–268
6 Marcotte E M, Pellegrini M, Thompson M J, et al. A combined algorithm for genome-wide prediction of protein function. Nature, 1999,
402: 83–86
7 Fraser H B, Hirsh A E, Wall D P, et al. Coevolution of gene expression among interacting proteins. Proc Natl Acad Sci USA, 2004, 101: 9033–9038
8 Goh C S, Bogan A A, Joachimiak M, et al. Co-evolution of proteins with their interaction partners. J Mol Biol, 2000, 299: 283–293
9 McPartland J M, Norris R W, Kilpatrick C W. Coevolution between cannabinoid receptors and endocannabinoid ligands. Gene, 2007, 397: 126–135
10 Golozoubova V, Hohtola E, Matthias A, et al. Only UCP1 can mediate adaptive nonshivering thermogenesis in the cold. Faseb J, 2001, 15: 2048–2050
11 Yang J, Wang Z L, Zhao X Q, et al. Natural selection and adaptive evolution of leptin in the ochotona family driven by the cold environmental stress. PLoS ONE, 2008, 3: e1472
12 Takada I, Kato S. PPARs target genes. Nippon Rinsho, 2005, 63: 573–577
13 Qian H, Hausman G J, Compton M M, et al. Leptin regulation of peroxisome proliferator-activated receptor-gamma, tumor necrosis factor, and uncoupling protein-2 expression in adipose tissues. Biochem Biophys Res Commun, 1998, 246: 660–667
14 Scarpace P J, Matheny M. Leptin induction of *UCP1* gene expression is dependent on sympathetic innervation. Am J Physiol, 1998, 275(2 Pt 1): E259–264
15 Cannon B, Nedergaard J. Brown adipose tissue: Function and physiological significance. Physiol Rev, 2004, 84: 277–359
16 Kershaw E E, Flier J S. Adipose tissue as an endocrine organ. J Clin Endocrinol Metab, 2004, 89: 2548–2556
17 Trayhurn P, Beattie J H. Physiological role of adipose tissue: white adipose tissue as an endocrine and secretory organ. Proc Nutr Soc, 2001, 60: 329–339
18 Himms-Hagen J. Brown adipose tissue thermogenesis: Interdisciplinary studies. Faseb J, 1990, 4: 2890–2898
19 Jansky L. Humoral thermogenesis and its role in maintaining energy balance. Physiol Rev, 1995, 75: 237–259
20 Rothwell N J, Stock M J. Biological distribution and significance of brown adipose tissue. Comp Biochem Physiol A, 1985, 82: 745–751
21 Cinti S. Adipocyte differentiation and transdifferentiation: Plasticity of the adipose organ. J Endocrinol Invest, 2002, 25: 823–835
22 Reidy S P, Weber J. Leptin: An essential regulator of lipid metabolism. Comp Biochem Physiol A Mol Integr Physiol, 2000, 125: 285–298
23 Zhang Y, Proenca R, Maffei M, et al. Positional cloning of the mouse obese gene and its human homologue. Nature, 1994, 372: 425–432
24 Campfield L A, Smith F J, Guisez Y, et al. Recombinant mouse OB protein: Evidence for a peripheral signal linking adiposity and central neural networks. Science, 1995, 269: 546–549
25 Chen H, Charlat O, Tartaglia L A, et al. Evidence that the diabetes gene encodes the leptin receptor: Identification of a mutation in the leptin receptor gene in *db/db* mice. Cell, 1996, 84: 491–495
26 Lee G H, Proenca R, Montez J M, et al. Abnormal splicing of the leptin receptor in diabetic mice. Nature, 1996, 379: 632–635
27 Fujii H. PPARs-mediated intracellular signal transduction. Nippon Rinsho, 2005, 63: 565–571
28 Dreyer C, Krey G, Keller H, et al. Control of the peroxisomal beta-oxidation pathway by a novel family of nuclear hormone receptors. Cell, 1992, 68: 879–887
29 Kliewer S A, Forman B M, Blumberg B, et al. Differential expression and activation of a family of murine peroxisome proliferator-activated receptors. Proc Natl Acad Sci USA, 1994, 91: 7355–7359
30 Tontonoz P, Hu E, Devine J, et al. PPAR gamma 2 regulates adipose expression of the phosphoenolpyruvate carboxykinase gene. Mol Cell Biol, 1995, 15: 351–357
31 Carriere F, Barrowman J A, Verger R, et al. Secretion and contribution to lipolysis of gastric and pancreatic lipases during a test meal in humans. Gastroenterology, 1993, 105: 876–888
32 Pruitt K D, Tatusova T, Maglott D R. NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res, 2005, 33(Database issue):

D501–D504

33   Thompson J D, Gibson T J, Plewniak F, et al. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res, 1997, 25: 4876–4882

34   Kumar S, Tamura K, Nei M. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform, 2004, 5: 150–163

35   Springer M S, Burk-Herrick A, Meredith R, et al. The adequacy of morphology for reconstructing the early history of placental mammals. Syst Biol, 2007, 56: 673–684

36   Murphy W J, Eizirik E, Johnson W E, et al. Molecular phylogenetics and the origins of placental mammals. Nature, 2001, 409: 614–618

37   Yang Z. PAML: A program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci, 1997, 13: 555–556

38   Nielsen R, Yang Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics, 1998, 148: 929–936

39   Huelsenbeck J P, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics, 2001, 17: 754–755

40   Yang Z, Wong W S, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol, 2005, 22: 1107–1118

41   Rossiter S J, Ransome R D, Faulkes C G, et al. Mate fidelity and intra-lineage polygyny in greater horseshoe bats. Nature, 2005, 437: 408–411

42   Yang Z, Nielsen R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol, 2002, 19: 908–917

43   MacDougald O A, Lane M D. Transcriptional regulation of gene expression during adipocyte differentiation. Annu Rev Biochem, 1995, 64: 345–373

44   Berglund A C, Wallner B, Elofsson A, et al. Tertiary windowing to detect positive diversifying selection. J Mol Evol, 2005, 60: 499–504

45   Benner S A, Caraco M D, Thomson J M, et al. Planetary biology--paleontological, geological, and molecular histories of life. Science, 2002, 296: 864–868

46   Gaucher E A, Miyamoto M M, Benner S A. Evolutionary, structural and biochemical evidence for a new interaction site of the leptin obesity protein. Genetics, 2003, 163: 1549–1553

47   Benner S A, Trabesinger N, Schreiber D. Post-genomic science: converting primary structure into physiological function. Adv Enzyme Regul, 1998, 38: 155–180

48   Hughes D A, Jastroch M, Stoneking M, et al. Molecular evolution of UCP1 and the evolutionary history of mammalian non-shivering thermogenesis. BMC Evol Biol, 2009, 9: 4

49   Saito S, Saito C T, Shingai R. Adaptive evolution of the uncoupling protein 1 gene contributed to the acquisition of novel nonshivering thermogenesis in ancestral eutherian mammals. Gene, 2008, 408: 37–44

50   Matsubara M, Maruoka S, Katayose S. Inverse relationship between plasma adiponectin and leptin concentrations in normal-weight and obese women. Eur J Endocrinol, 2002, 147: 173–180

51   Cancello R, Zingaretti M C, Sarzani R, et al. Leptin and *UCP1* genes are reciprocally regulated in brown adipose tissue. Endocrinology, 1998, 139: 4747–4750

52   Ahima R S, Prabakaran D, Mantzoros C, et al. Role of leptin in the neuroendocrine response to fasting. Nature, 1996, 382: 250–252

53   Caro J F, Sinha M K, Kolaczynski J W, et al. Leptin: the tale of an obesity gene. Diabetes, 1996, 45: 1455–1462

54   Zhang F, Basinski M B, Beals J M, et al. Crystal structure of the obese protein leptin-E100. Nature, 1997, 387: 206–209

55   Hiroike T, Higo J, Jingami H, et al. Homology modeling of human leptin/leptin receptor complex. Biochem Biophys Res Commun, 2000, 275: 154–158

56   Grasso P, White D W, Tartaglia L A, et al. Inhibitory effects of leptin-related synthetic peptide 116-130 on food intake and body weight gain in female C57BL/6J ob/ob mice may not be mediated by peptide activation of the long isoform of the leptin receptor. Diabetes, 1999, 48: 2204–2209

57   Winkler F K, D'Arcy A, Hunziker W. Structure of human pancreatic lipase. Nature, 1990, 343: 771–774

58   Schmitt A, Gutierrez G J, Lenart P, et al. Histone H3 phosphorylation during *Xenopus oocyte* maturation: Regulation by the MAP kinase/p90Rsk pathway and uncoupling from DNA condensation. FEBS Lett, 2002, 518: 23–28

59   Shapiro L, Scherer P E. The crystal structure of a complement-1q family protein suggests an evolutionary link to tumor necrosis factor. Curr Biol, 1998, 8: 335–338

60   Engel J, Prockop D J. The zipper-like folding of collagen triple helices and the effects of mutations that disrupt the zipper. Annu Rev Biophys Biophys Chem, 1991, 20: 137–152

Supporting Information

Table S1   Summary of genes studied with their protein product name and function

Table S2   Accession numbers of orthologous sequences.

Table S3   Summary of likelihood ratio parameters and the phylogenetic trees of the nine datasets (*a–i*)

Table S4   Likelihood ratio test and parameters

Table S5   Summary of positive selected sites

Figure S1   Phylogenetic trees of the 12 lipid metabolism genes

The supporting information is available online at csb.scichina.com and www.springerlink.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.