

# ANALYSIS OF A LIPID BIOSYNTHESIS PROTEIN FAMILY AND PHOSPHOLIPID STRUCTURAL VARIATIONS

MICHIHIRO TANAKA<sup>1</sup>  
mtanaka@kuicr.kyoto-u.ac.jp

YUKI MORIYA<sup>1</sup>  
moriya@kuicr.kyoto-u.ac.jp

SUSUMU GOTO<sup>1</sup>  
goto@kuicr.kyoto-u.ac.jp

MINORU KANEHISA<sup>1,2</sup>  
kanehisa@kuicr.kyoto-u.ac.jp

<sup>1</sup>*Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan*

<sup>2</sup>*Human Genome Center, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108-8639, Japan*

Glycerophospholipids are major structural lipids in cellular membrane systems and play key roles as suppliers of the first and second messengers in the signal transduction and molecular recognition processes. The distribution of lipid components differs among organelles and cells. The distribution is controlled by two pathways in lipid metabolism: *de novo* and remodeling pathways. Glycerophospholipids including arachidonic and stearic acids are mostly produced in the remodeling pathway, whereas lipid chains are reconstructed from those synthesized in the *de novo* pathway. Recently lysophospholipid acyltransferases have been isolated as key enzymes in the remodeling pathway, and the substrate specificity has been investigated in terms of the chemical substructures of glycerophospholipids, such as the type of head groups and the length of aliphatic chains. These experimental studies have been reported for specific organisms, and only two representative sequence motifs are known for acyltransferases: a general pattern and the pattern for membrane-bound *O*-acyltransferase (MBOAT). Here we attempt to correlate the sequence patterns and the substrate specificity of lysophospholipid acyltransferases in 89 eukaryotic genomes in order to understand the roles of this enzyme family and underlying glycerophospholipid structural variations. Using phylogenetic and domain analyses, the lysophospholipid acyltransferase family was divided into 18 subtypes. Furthermore, we examined the occurrence of identified subtypes in eukaryotic genomes, and found the expansion of these subtypes in vertebrates. These findings may provide clues to understanding structural variations and distributions of glycerophospholipids in different organisms.

*Keywords:* lysoacyltransferase; glycerophospholipids remodeling pathways.

## 1. Introduction

Glycerophospholipids are major structural lipids in cellular membrane systems and play key roles as suppliers of the first and second messengers in the signal transduction and molecular recognition processes [1]. They have several molecular variations stemming from the combination of three substructure components: a polar-head group and two aliphatic chains at *sn*-1 and *sn*-2 positions. Based on the polar-head group, glycerophospholipids are classified into several classes, such as phosphatidic acid (PA), phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylglycerol (PG), cardiolipin (CL), phosphatidylinositol (PI) and phosphatidylserine (PS). Combined with

the variation of two aliphatic chains, over 1,000 glycerophospholipids species have been reported so far [1].

The distribution of lipid species differs among organelles and cells. Glycerophospholipid biosynthesis is controlled by two lipid metabolism pathways: the *de novo* pathway [2] and the remodeling pathway [3]. Two acyltransferases, glycerol-phosphate acyltransferase (GPAT) and acylglycerol-phosphate acyltransferase (AGPAT), are responsible for acylation in the *de novo* pathway. Because these enzymes have no substrate specificities to aliphatic chains [4-7], the biased distributions of lipid species are not generated in the *de novo* pathway. The remodeling pathway uses phospholipase A<sub>2</sub> (PLA<sub>2</sub>) and lysophospholipid acyltransferases (LPLATs). PLA<sub>2</sub> hydrolyses the ester bond at *sn*-2 position of a glycerophospholipid. The produced free fatty acid and lysoacylglycerophospholipids are then used as precursors for various second messengers. LPLAT converts the remaining lysoacylglycerophospholipid into another glycerophospholipid, in which the aliphatic chain at *sn*-2 position may differ from that of the original one. For example, glycerophospholipids including arachidonic and stearic acids are mostly produced in the remodeling pathway [8]. Many members of the PLA<sub>2</sub> gene family have been cloned, and their substrate specificities for aliphatic chains have been well characterized, as well as the corresponding sequence motifs [9-12]. On the other hand, only several members from the mammalian LPLAT family have been isolated [13-16, 17-20], although their substrate specificity has been investigated in terms of the chemical substructures of glycerophospholipids, such as the type of head groups and the length of aliphatic chains [20]. However, these experimental studies have been performed only for certain organisms, and only two representative sequence motifs are currently registered in the Pfam motif database [21]: a general acyltransferase pattern (Acyltransferase) and the pattern for membrane-bound *O*-acyltransferase (MBOAT).

Here we attempt to correlate the sequence patterns and the substrate specificity of LPLATs in 89 eukaryotic genomes in order to understand the roles of this enzyme family and underlying glycerophospholipid structural variations. Using phylogenetic and domain analyses, the LPLAT family was divided into 18 subtypes. Furthermore, we examined the occurrence of identified subtypes in the eukaryotic genomes and found that these subtypes are expanded in vertebrates. These findings may provide clues to understanding structural variations and distributions of glycerophospholipids in different organisms.

## 2. Materials and Methods

### 2.1. Experimentally characterized lysophospholipid acyltransferase sequences

We collected all possible LPLATs (lysophospholipid acyltransferases) from completely sequenced eukaryotes to analyze species variations of glycerophospholipids (Fig. 1). To provide initial query sequences, we retrieved experimentally characterized LPLAT encoding protein sequences from the UniProtKB database [22] using several keywords related to lysophospholipid acyltransferase followed by manual inspection for substrate specificities. We obtained 24 sequences, including 7 sequences from *H. sapiens* (Q6P1A2, A9EDR2, A9EDQ5, Q6UWP7, Q7L5N7, Q96N66 and Q8NF37), 6

sequences from *M. musculus* (Q91V01, Q3TFD2, Q8BH98, Q3UN02, Q8BYI6 and Q8R3I2), 3 sequences from *D. rerio* (Q6NYV8, Q502J0 and Q1LWG4), 2 sequences from *R. norvegicus* (Q1HAQ0 and Q5FVN0), and a sequence from each of *B. taurus* (Q3SZL3), *G. gallus* (Q5F3X0), *D. melanogaster* (Q0KHU5), *S. cerevisiae* (Q06510), *S. pombe* (O42916) and *D. discoideum* (Q54DX7).

## 2.2. PSI-BLAST Search for lysophospholipid acyltransferase candidates

We performed PSI-BLAST search [23] on eukaryotes in the KEGG GENES database [24] using each of the above 24 sequences as a seed. The following two conditions were used for each PSI-BLAST search:

- i) The E-value threshold was set to  $< 10^{-2}$  for creating the sequence set to generate a position-specific scoring matrix (PSSM).
- ii) The PSI-BLAST hit should include the seed sequence, and the maximum number of iteration cycles was set to 30.
- iii) All of the results were integrated to produce a list of non-redundant sequences.

## 2.3. Clustering analysis

We divided PSI-BLAST hits into several clusters to identify non-LPLAT sequences (false positives) in the following way:

- i) A Smith-Waterman score matrix was constructed using all-against-all SSEARCH [17] for the sequences in the list.
- ii) The complete linkage method was applied to the score matrix by using the R package [26] with the threshold Smith-Waterman score  $> 20$ .

LPLATs are known to have either of two Pfam motifs, PF01553 (Acyltransferase) or PF03062 (MBOAT). We considered the clusters that did not contain either of these motifs as false positives and did not use them for further analysis.

## 2.4. Phylogenetic analysis of each cluster

To determine the orthologous relationships among sequences, we applied the tree-reconciliation phylogenetic approach [27] to each of the remaining clusters.

- i) We used *E. coli* acyltransferase (eco:b2836) as the outgroup of lysophospholipid acyltransferase family to construct rooted gene trees.
- ii) For each cluster, a multiple sequence alignment was created using MAFFT [28] with default parameters except for max-iteration, which was set to 30.
- iii) If more than 10% of the sequences in an alignment region corresponded to gaps, the region was eliminated.
- iv) Neighbor-joining phylogenies were constructed using the NEIGHBOR program in the PHYLIP 3.6 package [29]. One hundred bootstrap replicates were generated to assess the variation in the data using the PHYLIP 3.6 SEQBOOT program.
- v) A species tree for the 89 eukaryotes was constructed based on KEGG taxonomy data [30] and a previously defined metazoan species tree [31]. The list of 89 eukaryotic species and taxonomic classification used in this study can be obtained from [http://mbi3.kuicr.kyoto-u.ac.jp/supp/mtanaka/taxonomy\\_ibsb2009.txt](http://mbi3.kuicr.kyoto-u.ac.jp/supp/mtanaka/taxonomy_ibsb2009.txt)

To detect duplication and speciation events on the gene trees, we used a phylogeny-based algorithm [27]. This algorithm fits a given gene tree to its corresponding species tree and then infers the minimum set of duplications necessary to explain the topology of the gene tree. We call the derived tree with the duplication events a reconciled tree. We manually investigated the reconciled tree to infer the orthologous and paralogous relationships for sequence pairs in a given cluster.

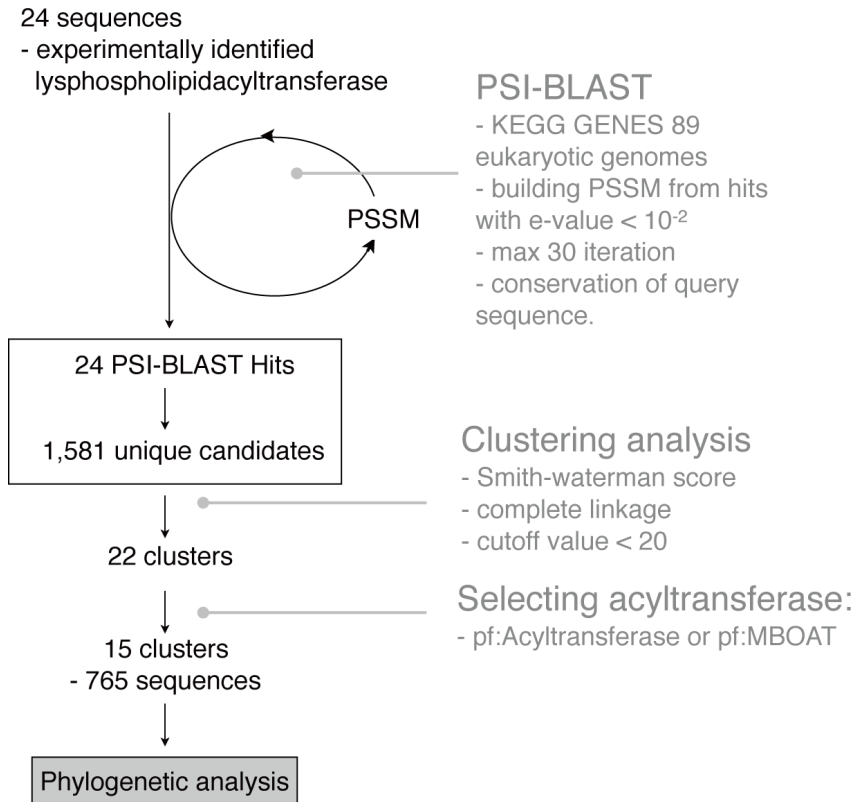


Figure 1. The protocol for detecting lysophospholipid acyltransferase candidates.

### 3. Results

#### 3.1. Identification of lysophospholipid acyltransferase sequences

We obtained 1,581 LPLAT candidates among 89 eukaryotes through PSI-BLAST. The successive clustering analysis divided them into 22 clusters. After examining the occurrence of two Pfam motifs (pf:Acyltransferase and pf:MBOAT) for each of 22 clusters, 7 clusters were considered as false positives and excluded from further analyses. Of the remaining 15 clusters, 9 clusters (429 sequences from 75 species) contained

sequences with the Acyltransferase motif and the other 6 clusters (336 sequences from 75 species) contained sequences with the MBOAT motif. In total, we obtained 765 LPLAT sequences in 89 eukaryotic genomes in the 15 clusters.

### 3.2. Phylogenetic analysis for LPLAT sequences

The taxonomic distribution of the LPLATs in the 15 clusters is shown in Table 1. Note that genes belonging to a cluster are sometimes highly duplicated in a specific taxonomy. For example, Deuterostomia has 3.94 sequences per species in cluster A, which is the largest cluster within the Acyltransferase family, and other taxonomies (from Protostomia to Plants) have 2.06 to 3.33 sequences per species in the cluster. Another example is cluster J, which is the largest cluster within the MBOAT family. Deuterostomia has 2.56 sequences per species, but other taxonomies have 1.00 to 1.88 sequences per species in this cluster. These results suggest that the Metazoa-specific and Deuterostomia-specific gene expansions have occurred within the Acyltransferase family and MBOAT family, respectively.

Table 1. Taxonomic distribution of the LPLAT gene families

Cluster	# of sequences	# of subtypes	Deuterostomia(18)	Protostomia(19)	Cnidarians(1)	Placozoans(1)	Plants(6)	Fungi(26)	Protists(18)
<b>Acyltransferase</b>									
-A	144	3	<b>3.94</b> (71/18)	<b>2.06</b> (37/18)	<b>3.00</b> (3/1)	<b>3.00</b> (3/1)	<b>3.33</b> (10/3)	1.15 (15/13)	1.67 (5/3)
-B	92	3	<b>2.00</b> (34/17)	1.63 (31/19)	1.00 (1/1)	<b>2.00</b> (2/1)	<b>3.33</b> (20/6)	-	0.75 (3/4)
-C	67	4	<b>2.76</b> (47/17)	1.00 (16/16)	<b>2.00</b> (2/1)	<b>2.00</b> (2/1)	-	-	-
-D	53	1	1.13 (18/16)	1.00 (19/19)	1.00 (1/1)	1.00 (1/1)	1.83 (11/6)	1.00 (2/2)	1.00 (1/1)
-E	23	1	-	-	-	-	-	1.05 (22/21)	1.00 (1/1)
-F	22	1	1.00 (7/7)	1.00 (12/12)	-	-	-	1.00 (3/3)	-
-G	15	1	1.08 (14/13)	1.00 (1/1)	-	-	-	-	-
-H	7	1	-	-	-	-	-	1.00 (7/7)	-
-I	6	1	<b>4.00</b> (4/1)	1.00 (3/3)	-	-	-	1.00 (2/2)	-
<b>MBOAT</b>									
-J	98	2	<b>2.56</b> (41/16)	1.88 (30/16)	1.00 (1/1)	-	1.00 (1/1)	1.60 (24/15)	1.00 (1/1)
-K	80	5	<b>2.77</b> (61/22)	1.00 (19/19)	1.00 (1/1)	1.00 (1/1)	-	-	<b>3.50</b> (7/2)
-L	69	1	1.00 (3/3)	1.16 (22/19)	<b>2.00</b> (2/1)	-	1.60 (8/5)	1.08 (27/25)	<b>2.33</b> (7/3)
-M	43	1	1.31 (21/16)	1.18 (20/17)	-	<b>2.00</b> (2/1)	-	-	-
-N	31	1	1.00 (8/8)	-	-	-	-	1.10 (23/21)	-
-O	15	1	1.21 (17/14)	1.00 (1/1)	-	-	-	-	-

Bold numbers show taxonomies with  $\geq 2$  sequences per species in the cluster.

Two numbers in the parenthesis show the number of sequences and species, respectively.

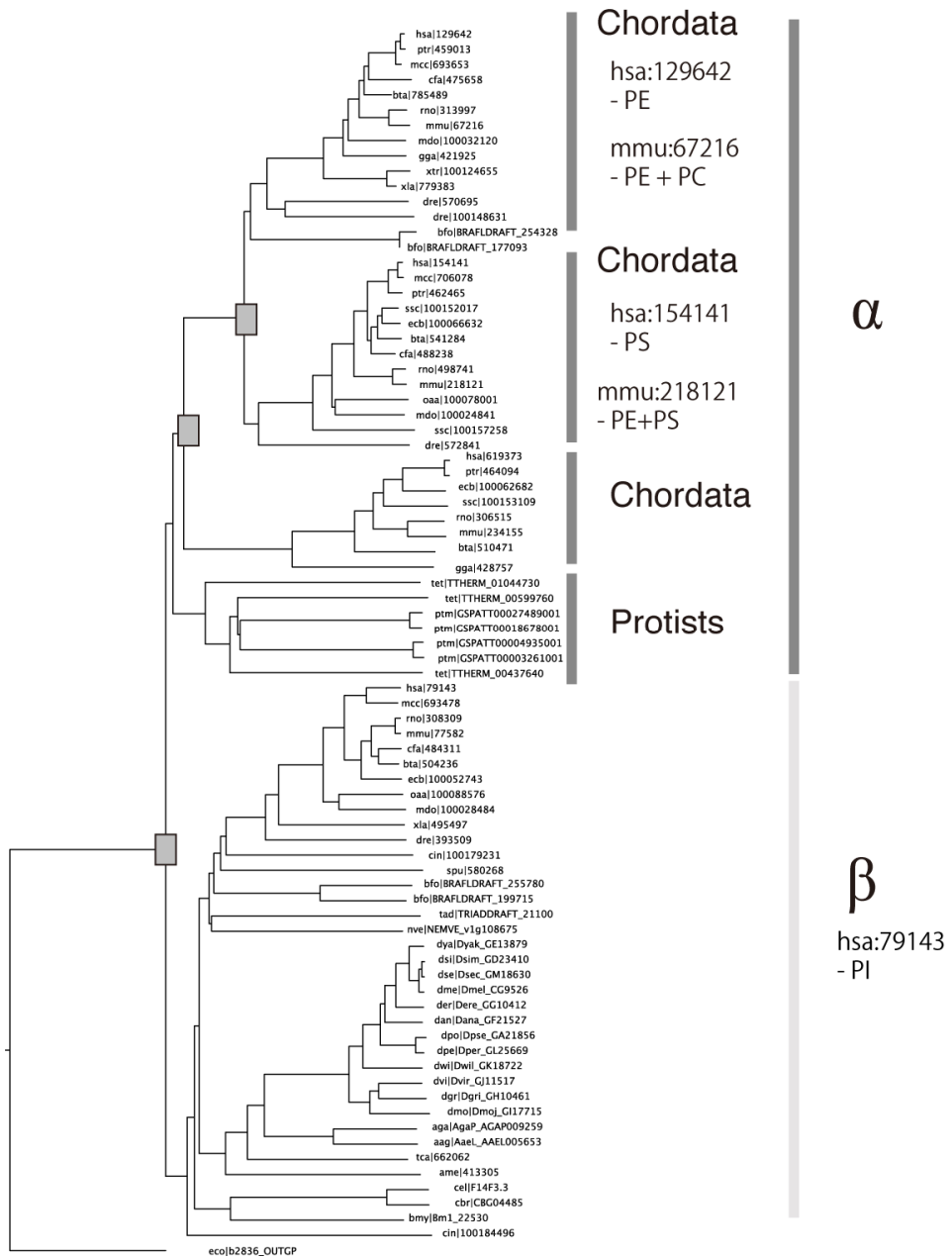


Figure 2. Phylogenetic tree representing the evolutionary relationships among sequences in cluster K. The branches representing duplication events are marked as grey squares.

In contrast, several clusters show a tandem distribution for a number of genes among different species. Deuterostomia, Protostomia and Fungi each have 1.00 sequence per species in clusters F and G, while other taxonomies have no sequences in these clusters. Similar distributions can be observed in clusters N and M. They also show sparse distributions along the wide range of taxonomies, suggesting that gene loss events have occurred frequently in these clusters.

To clarify the evolutionary process within the 15 clusters, we further examined and classified orthologous and paralogous relationships of sequence pairs by manually investigating the reconciled trees that derived from the gene trees. The gene trees for the 15 clusters were supported by bootstrap values > 45%. The number of subtypes in each cluster by this classification is shown in Table 1. Figure 2 shows the phylogenetic relationships of cluster K. With this result, we estimated that cluster K was divided into two subtypes ( $\alpha$  and  $\beta$ ), and one of the subtypes was further divided into three Chordata subtypes and a Protist subtype. Therefore, this reconciled tree indicates that several gene duplication events have occurred in the Chordata lineage, and each subtype may have evolved to have the substrate specificity of LPLATs.

### ***3.3. Applications of orthologous relationships to inferring the function of LPLATs***

The branching of gene trees emerges from two events. One is the speciation event. If a branching point between genes in a given gene tree derived from a speciation event, we consider this as an orthologous relationship. The function is usually conserved between orthologous genes. The other type of event is the ancestral gene duplication event that leads to a paralogous relationship between the duplicated genes. Paralogous genes can evolve different functions. To infer the function of a given gene from the other functionally characterized genes using phylogeny-based approach, we should correctly distinguish between orthologous and paralogous relationships for each branching point in a given tree.

To infer functions of LPLATs of genes with unknown function or substrate specificity, we used orthologous relationships of the 24 experimentally characterized LPLATs that were in 7 clusters among the 15 clusters. The 7 clusters included 4 clusters from Acyltransferase families (clusters A, C, D, E) and 3 clusters from MBOAT families (clusters K, L, M), and 479 LPLAT sequences in total. Table 2 shows the phylogenetic distributions of the LPLAT subtypes for which polar-head group specificity can be inferred from the experimentally characterized LPLATs.

## **4. Discussion**

In this work, we classified lysophospholipid acyltransferase sequences hierarchically and also examined the relationship between subtypes and polar-head group specificity. However, we could not identify LPLAT subtypes for PS, PG, PI and CL in plants and fungi. This may be due to 24 biased query sequences that include only two fungi and no plant sequences. Therefore, we need to find such LPLAT sequences. Because we did not use several clusters in Table 1 for the substrate specificity analysis, these sequences may be classified in the uninvestigated clusters. Further investigation of the sequences in the

subtypes in Table 2 may reveal novel features for the substrate specificities, which can be applied to the prediction of the function of LPLAT in those clusters.

It is difficult to elucidate the variation of glycerophospholipids only by the results for LPLAT analysis in this study. For example, Substrate specificities of LPLAT have been reported in terms of the length of aliphatic chains [20]. However, we have not investigated them in the current study, and to examine them in detail is one of our future works.

To better understand variation of the synthesis mechanisms of glycerophospholipids, we need to consider at least four other enzymes: elongases [32], desaturases [33], acyl-CoA synthetases [34] and phospholipases [35]. Desaturases and elongases play important roles in *de novo* fatty acid synthesis. Hashimoto *et al* reported that fatty acid variations could be predicted from the combination of subfamilies for desaturases and elongases [36, 37]. An important step in transferring fatty acids to the glycerol backbone is acyl-CoA synthesis. Variation in the acyl-CoA synthetase family has been reported based on the substrate specificity of the chain length of fatty acids [9-11]. Phospholipases are also classified into over 20 subfamilies whose acyl chain selectivities are already known. For example, cPLA2  $\alpha$  has selectivity for arachidonic acid, and iPLA2  $\beta$  has no selectivity for specific acyl chains (a general purpose lipase). We must comprehensively understand the relationship between these enzymes and genetic variation among their encoding genes to elucidate the variation of glycerophospholipids in cellular membranes.

Table 2. Phylogenetic distribution of LPLAT subtypes based on head-group specificity of the substrate glycerophospholipids. Subtypes consist of a cluster ID and a subtype number. For example, C1 represents the first subtype of the cluster C. PC: Phosphatidylcholine, PE: Phosphatidylethanolamine, PS: Phosphatidylserine, PG: Phosphatidylglycerol, PI: Phosphatidylinositol, CL: Cardiolipin

Head group	Class	Vertebrates	Protochordata	Echinodermata	Protostomia	Cnidarians	Placozoans	Plants	Fungi	Protists
<b>PC</b>										
- C1		+	+	+	+	+				
- C2		+								
- D		+	+	+	+	+	+	+	+	+
- E									+	+
- K1		+	+							
- L		+	+	+	+	+		+	+	+
- M		+	+	+	+		+			
<b>PE</b>										
- K1		+	+							
- K2		+								
- K3		+								
- K4										+
- L		+	+	+	+	+		+	+	+
<b>PS</b>										
- K2		+								
- M		+	+	+	+		+			
<b>PG</b>										
A3		+	+	+	+	+	+			+
<b>PI</b>										
- A3		+	+	+	+	+	+			+
- K5		+	+	+	+	+	+			
<b>CL</b>										
A3		+	+	+	+	+	+			+

In conclusion, we have classified 765 lysophospholipid acyltransferase candidates into 15 clusters and defined 18 lysophospholipid acyltransferase subtypes for 7 of 15 clusters through phylogenetic analysis. Further analyses of these results may provide new insights into mechanisms contributing to the variation of glycerophospholipids.

### Acknowledgments

We thank Dr. Kosuke Hashimoto and Dr. Nelson Hayes for critical reading of our manuscript. This work was supported in part by the grants from the Ministry of Education, Culture, Sports, Science and Technology. The computational resource was provided by the Bioinformatics Center, Institute for Chemical Research, Kyoto University.

### References

- [1] van Meer, G., Voelker, D.R., Feigenson, G.W., Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol*, 9(2):112-124, 2008.
- [2] Kennedy, E.P., Weiss, S.B., The function of cytidine coenzymes in the biosynthesis of phospholipides. *J Biol Chem*, 222(1):193-214, 1956.
- [3] Lands, W.E., Metabolism of glycerolipides; a comparison of lecithin and triglyceride synthesis. *J Biol Chem*, 231(2):883-888, 1958.
- [4] Holub, B.J., Kuksis, A., Metabolism of molecular species of diacylglycerophospholipids. *Adv Lipid Res*, 16:1-125, 1978.
- [5] Miki, Y., Hosaka, K., Yamashita, S., Handa, H., Numa, S., Acyl-acceptor specificities of 1-acylglycerolphosphate acyltransferase and 1-acylglycerophosphorylcholine acyltransferase resolved from rat liver microsomes. *Eur J Biochem*, 81(3):433-441, 1977.
- [6] Waku, K., Origins and fates of fatty acyl-CoA esters. *Biochim Biophys Acta*, 1124(2):101-111, 1992.
- [7] Nakagawa, Y., Rustow, B., Rabe, H., Kunze, D., Waku, K., The de novo synthesis of molecular species of phosphatidylinositol from endogenously labeled CDP diacylglycerol in alveolar macrophage microsomes. *Arch Biochem Biophys*, 268(2):559-566, 1989.
- [8] Sugiura, T., Masuzawa, Y., Nakagawa, Y., Waku, K., Transacylation of lyso platelet-activating factor and other lysophospholipids by macrophage microsomes. Distinct donor and acceptor selectivities. *J Biol Chem*, 262(3):1199-1205, 1987.
- [9] Soupene, E., Kuypers, F.A., Mammalian long-chain acyl-CoA synthetases. *Exp Biol Med* (Maywood), 233(5):507-521, 2008.
- [10] Fujino, T., Takei, Y.A., Sone, H., Ioka, R.X., Kamataki, A., Magoori, K., Takahashi, S., Sakai, J., Yamamoto, T.T., Molecular identification and characterization of two medium-chain acyl-CoA synthetases, MACS1 and the Sa gene product. *J Biol Chem*, 276(38):35961-35966, 2001.
- [11] Watkins, P.A., Very-long-chain acyl-CoA synthetases. *J Biol Chem*, 283(4):1773-1777, 2008.

- [12] Kratzer, S., Schuller, H.J., Carbon source-dependent regulation of the acetyl-coenzyme A synthetase-encoding gene ACS1 from *Saccharomyces cerevisiae*. *Gene*, 161(1):75-79, 1995.
- [13] Kume, K., Shimizu, T., cDNA cloning and expression of murine 1-acyl-sn-glycerol-3-phosphate acyltransferase. *Biochem Biophys Res Commun*, 237(3):663-666, 1997.
- [14] Eberhardt, C., Gray, P.W., Tjoelker, L.W., Human lysophosphatidic acid acyltransferase. cDNA cloning, expression, and localization to chromosome 9q34.3. *J Biol Chem*, 272(32):20299-20305, 1997.
- [15] Hollenback, D., Bonham, L., Law, L., Rossnagle, E., Romero, L., Carew, H., Tompkins, C.K., Leung, D.W., Singer, J.W., White, T., Substrate specificity of lysophosphatidic acid acyltransferase beta - evidence from membrane and whole cell assays. *J Lipid Res*, 47(3):593-604, 2006.
- [16] Yang, Y., Cao, J., Shi, Y., Identification and characterization of a gene encoding human LPGAT1, an endoplasmic reticulum-associated lysophosphatidylglycerol acyltransferase. *J Biol Chem*, 279(53):55866-55874, 2004.
- [17] Nakanishi, H., Shindou, H., Hishikawa, D., Harayama, T., Ogasawara, R., Suwabe, A., Taguchi, R., Shimizu, T., Cloning and characterization of mouse lung-type acyl-CoA:lysophosphatidylcholine acyltransferase 1 (LPCAT1). Expression in alveolar type II cells and possible involvement in surfactant production. *J Biol Chem*, 281(29):20140-20147, 2006.
- [18] Shindou, H., Hishikawa, D., Nakanishi, H., Harayama, T., Ishii, S., Taguchi, R., Shimizu, T., A single enzyme catalyzes both platelet-activating factor production and membrane biogenesis of inflammatory cells. Cloning and characterization of acetyl-CoA:LYSO-PAF acetyltransferase. *J Biol Chem*, 282(9):6532-6539, 2007.
- [19] Chen, X., Hyatt, B.A., Mucenski, M.L., Mason, R.J., Shannon, J.M., Identification and characterization of a lysophosphatidylcholine acyltransferase in alveolar type II cells. *Proc Natl Acad Sci U S A*, 103(31):11724-11729, 2006.
- [20] Hishikawa, D., Shindou, H., Kobayashi, S., Nakanishi, H., Taguchi, R., Shimizu, T., Discovery of a lysophospholipid acyltransferase family essential for membrane asymmetry and diversity. *Proc Natl Acad Sci U S A*, 105(8):2830-2835, 2008.
- [21] Finn, R.D., Tate, J., Mistry, J., Coghill, P.C., Sammut, S.J., Hotz, H.R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L., et al., The Pfam protein families database. *Nucleic Acids Res*, 36(Database issue):D281-288, 2008.
- [22] <http://www.uniprot.org/help/uniprotkb>.
- [23] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25(17):3389-3402, 1997.
- [24] Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., et al., KEGG for linking genomes to life and the environment. *Nucleic Acids Res*, 36(Database issue):D480-484, 2008.
- [25] Smith, T.F., Waterman, M.S., Identification of common molecular subsequences. *J Mol Biol*, 147(1):195-197, 1981.
- [26] Team RDC: R: a language and environment for statistical computing Vienna, Austria: R Foundation for Statistical Computing 2003.

- [27] Goodman, M.C.J., Moore, G.M., Romero-Herrera, A.E., Matsuda, G., Fitting the gene lineage into its species lineage, a parsimony strategy illustrated by cladograms constructed from globin sequences. *Syst Zool*, 28:132-163, 1979.
- [28] Katoh, K., Kuma, K., Toh, H., Miyata, T., MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*, 33(2):511-518, 2005.
- [29] Felsenstein, J., PHYLIP: Phylogenetic Inference Package. Release 3.6. Department of Genome Sciences, University of Washington, Seattle. 2005.
- [30] <ftp://ftp.genome.jp/pub/kegg/genes/taxonomy>.
- [31] Ponting, C.P., The functional repertoires of metazoan genomes. *Nat Rev Genet*, 9(9):689-698, 2008.
- [32] Leonard, A.E., Pereira, S.L., Sprecher, H., Huang, Y.S., Elongation of long-chain fatty acids. *Prog Lipid Res*, 43(1):36-54, 2004.
- [33] Sperling, P., Ternes, P., Zank, T.K., Heinz, E., The evolution of desaturases. *Prostaglandins Leukot Essent Fatty Acids*, 68(2):73-95, 2003.
- [34] Watkins, P.A., Fatty acid activation. *Prog Lipid Res*, 36(1):55-83, 1997.
- [35] Burke, J.E., Dennis, E.A., Phospholipase A2 structure/function, mechanism, and signaling. *J Lipid Res*, 50 Suppl:S237-242, 2009.
- [36] Hashimoto, K., Yoshizawa, A.C., Okuda, S., Kuma, K., Goto, S., Kanehisa, M., The repertoire of desaturases and elongases reveals fatty acid variations in 56 eukaryotic genomes. *J Lipid Res*, 49(1):183-191, 2008.
- [37] Hashimoto, K., Yoshizawa, A.C., Saito, K., Yamada, T., Kanehisa, M., The repertoire of desaturases for unsaturated fatty acid synthesis in 397 genomes. *Genome Inform*, 17(1):173-183, 2006.