The Repertoire of Desaturases for Unsaturated Fatty Acid Synthesis in 397 Genomes

Kosuke Hashimoto khashimo@kuicr.kyoto-u.ac.jp
Akiyasu C. Yoshizawa acyshzw@kuicr.kyoto-u.ac.jp
Koichi Saito saito@kuicr.kyoto-u.ac.jp
Takuji Yamada takuji@kuicr.kyoto-u.ac.jp
Minoru Kanehisa kanehisa@kuicr.kyoto-u.ac.jp
Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan

Abstract

Fatty acids are the major components of membrane molecules. The fact that unsaturated fatty acids play multiple important roles, physically and biologically, means that the ratio of unsaturated to saturated fatty acids in the membrane needs to be strictly regulated to maintain cellular homeostasis. The most ubiquitous and widespread modification to fatty acids, which results in a great diversity of different structures, is the insertion of double bonds. Fatty acid desaturases directly introduce regioselective double bonds into fatty acids. A phylogenetic analysis of desaturases suggests that the sequences of these proteins include a highly conserved domain, which determines the differences in specificity and regioselectivity found in these enzymes. In this study, we performed a systematic analysis of fatty acid desaturases found in the genomic data of 397 organisms. We obtained a set of desaturases clustered by regioselectivity using a hierarchal clustering analysis.

Keywords: fatty acid desaturase, MUFA, PUFA, comparative genomics, lipid metabolism

1 Introduction

Membrane lipid molecules are composed of hydrophobic and hydrophilic parts, the latter of which is generally used for lipid classification. However, the hydrophobic part, whose major component is fatty acids, also have a great diversity [9]. Most of the natural fatty acids include more than sixteen carbon atoms, which can be modified with methyl, hydroxyethyl, and other functional groups. However, the most ubiquitous and widespread modification to fatty acids is the insertion of double bonds, which dramatically increase the membrane fluidity due to the decrease in electrostatic interactions between lipid molecules. Recent reports proposed models that can control the ratio of unsaturated/saturated fatty acids in response to temperature signals [15, 25]. In addition to their biophysical characteristics, unsaturated fatty acids have crucial roles in many biological systems. For example, monounsaturated fatty acid (MUFA), which contains one double-bond in an acyl-chain, plays a significant role in the control of metabolism and can serve as a mediator in signal transduction [7, 14]. In mammals, metabolites of polyunsaturated fatty acids (PUFA) are also used as signaling molecules. Arachidonic acid (Figure 1) is abundantly stored within the cell membrane and is required as a substrate for eicosanoid synthesis [30]. Eicosapentanoic acid (20:5, see the nomenclature in Figure 1) is the precursor of prostaglandins, while docosahexaenoic acid (22:6) is essential for nervous system maintenance and development [16]. Because of these multiple roles of unsaturated fatty acids, the ratio of unsaturated fatty acids to saturated fatty acids is strictly regulated to maintain cellular homeostasis in each organism or tissue.

To date, two different pathways of unsaturated fatty acid biosynthesis are known. The first one, used in many bacteria, is to leave double bonds generated during the biosynthesis of the fatty acid.

The second, found in almost all eukaryotes and some bacteria, uses desaturase enzymes to directly introduce regioselective double bonds into fatty acids taken up through the diet or produced during *de novo* synthesis. Fatty acid desaturases can be classified into two phylogenetically unrelated groups: soluble desaturases and membrane-bound desaturases. The former is the acyl-acyl carrier protein (ACP) desaturase found so far only in plants, which contains two D/EXXH motifs [27]. The latter is ubiquitous in eukaryotes and bacteria, characterized by three histidine box motifs that contain eight histidine residues. A phylogenetic analysis of desaturases suggests that the sequences in each group are conserved, and reflect the differences in specificity and regioselectivity between the groups [29].

In this study, we focus on the repertoire of desaturases, which controls the possible fatty acid variation in eukaryotes and some bacteria. For example, three desaturases, $\Delta 9$, $\Delta 5$ and $\Delta 6$, are present in the *Homo sapiens* genome [19]. On the other hand, no desaturases are found that can introduce double bonds in the positions between the 9th carbon and the methyl end. *H. sapiens* thus synthesizes an arachidonic acid through fatty acids received in the diet. The repertoire of desaturases is known not only in *H. sapiens* but also in other model organisms such as *Mus musculus, Rattus norvegicus*, and *Caenorhabditis elegans* [10, 18]. Using desaturase sequences from these organisms as queries for a PSI-BLAST search and clustering analysis, we have comprehensively explored the desaturase-like proteins in the KEGG GENES database [12], which contains only complete and high quality draft genome sequences. To divide this set of genes by regioselectivity, we applied hierarchical clustering to the desaturase-like sequences. Furthermore, the set of desaturases in each organism is mapped onto the metabolic pathways of essential PUFAs as the first step to elucidate fatty acid variations.

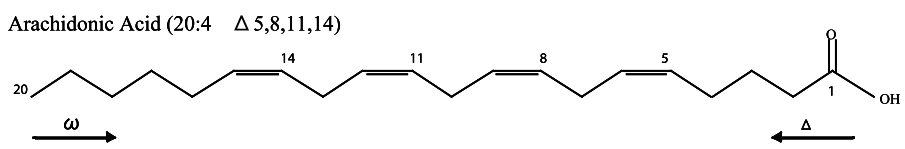

Figure 1: An example of the structure and nomenclature of unsaturated fatty acids [6]. The number before the colon denotes the number of carbon atoms and the number following refers to the number of double bonds. The delta (Δ) nomenclature is used to assign the position of an individual double bond counting from the carboxyl group carbon. In this case, arachidonic acid has twenty carbon atoms and four double bonds at the 5th, 8th, 11th and 14th carbons, so one simplified nomenclature is 20:4 Δ5, 8, 11, 14. The ω designation is also used to describe the position of a double bond from the methyl end.

2 Method

2.1 Genomic Data

We have used amino acid sequences from one of the following sources:

1. The complete genomic ORF sequences of 17 eukaryotes, 307 bacteria, 27 archaea, and draft-quality whole genomic ORF sequences of 38 eukaryotes were derived from KEGG GENES and DGENES [12].

2. The Ensembl ver38 (April 2006 version) database [5], which stores genomic sequences of 19 eukaryotes.

3. The six eukaryotic draft genomes and the gene prediction model of the rice genome (Table 1).

Table 1: The list of additional draft genomes.

Organism	Source	data version
Giardia lamblia	WGS(Genbank)	20030326
Gibberella zeae PH-1 11875 (*Fusarium graminearum*)	WGS(Genbank)	20040213
Medicago truncatula	TIGR	20060209
Oryza sativa japonica	TIGR	Rice Annotation Release 4 (January 12, 2006)
Plasmodium berghei	PlasmoDB	v4.4
Plasmodium chabaudi	PlasmoDB	v4.4
Ustilago maydis	WGS(Genbank)	20040402

2.2 Analyzed Enzyme Family

We searched for homologs of 59 fatty acid desaturase protein sequences. 54 of them are sequences from *M. musculus, Arabidopsis thaliana, Saccharomyces cerevisiae, Bacillus subtilis, Synechocystis* sp. PCC6803, *Thermosynechococcus elongatus, Gloeobacter violaceus, and Anabaena* sp. PCC7120 [1, 17] derived from the KEGG GENES database Release 38.0. Five of them, which are desaturases newly detected from non-complete genomes, are AF139720, AY278558, AY332747, AF489589, and AY493438 [31, 32] derived from Genbank database Release 153 [4] in GenomeNet.

2.3 Searching for Homologs – PSI-BLAST Search

The homologs were extracted using a similar method to that described in reference [33]. First, we collected desaturase-related (similar) proteins in the genomic dataset with PSI-BLAST [2], then discarded false-positive hits by hierarchical clustering analysis using the Smith-Waterman scores between sequences. The initial PSI-BLAST search (blastp2.2.10) against the genomic dataset was performed using the sequences as queries mentioned above.

By combining all the PSI-BLAST results whose E-values were below 0.01 into one file and removing duplicate hits, the initial data set was obtained. Then, the sequence similarity for each pair of sequences of the whole proteins in the initial dataset was examined with the SSEARCH 3.4t06 program [22], which is an implementation of the Smith-Waterman algorithm [28].

2.4 Searching for Homologs – Hierarchical Clustering Analysis

Hierarchical cluster analysis of these sequences was performed using the SSEARCH raw similarity scores with the complete linkage method. This allowed the removal the false positive hits from the initial data set and to classify the working data set. False positive clusters were defined as those clusters containing one or more proteins annotated as non-desaturases and no proteins annotated as desaturases. We defined the distance between the sequences as (distance)=1000/(Smith-Waterman score). The clustering analysis was performed with the R program package for statistical computing ver.1.7.1 [24, 36] and with the BioRuby library ver.1.0 [34].

The clustering results were confirmed manually using review articles [10, 19, 20] and the annotations for the sequence entries. We determined "putative desaturase" clusters as those that either contained (a) one or more sequences defined as known desaturases, or (b) no known non-desaturases (i.e. all sequences were functionally unannotated).

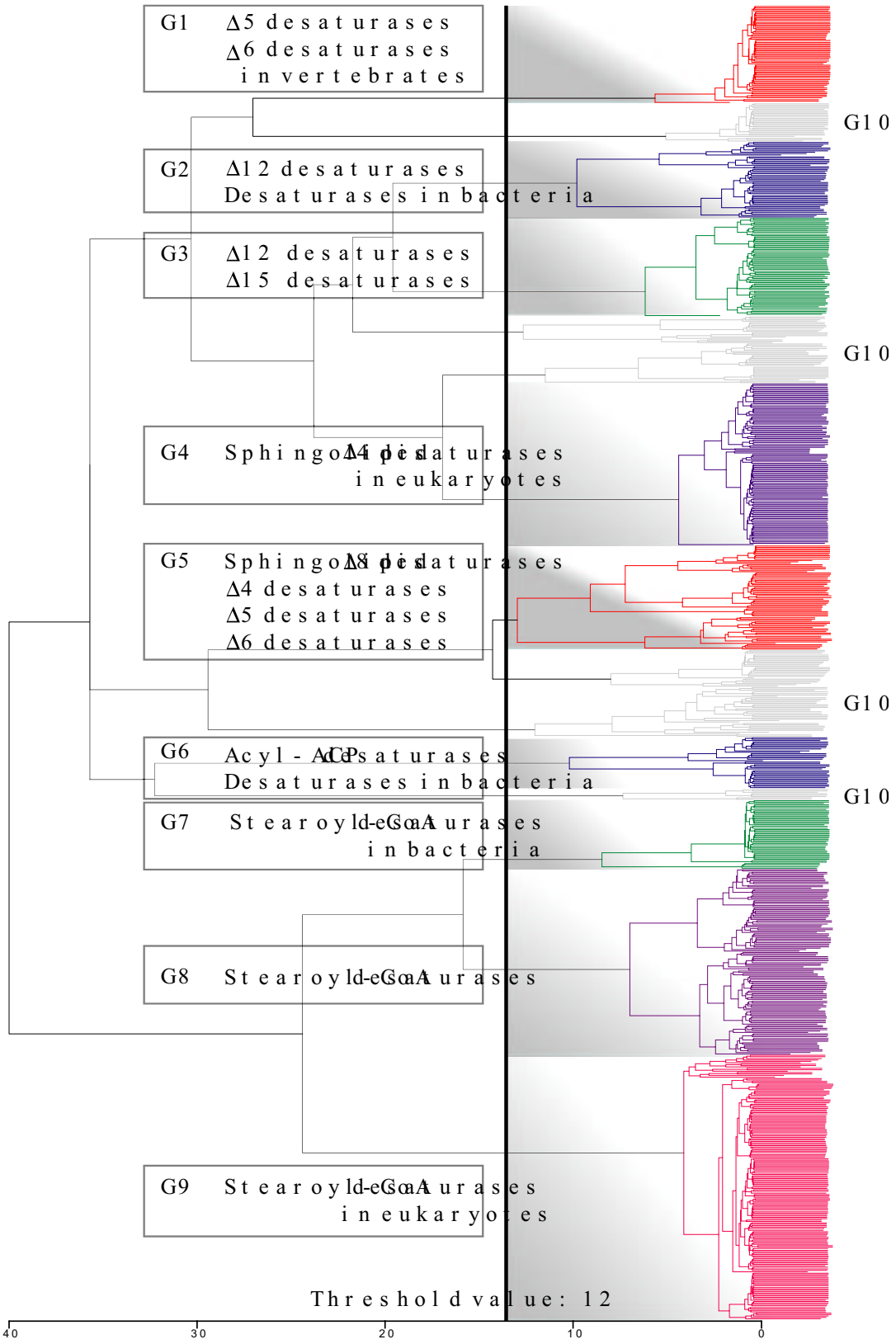

Figure 2: Clustering result of desaturase related proteins.

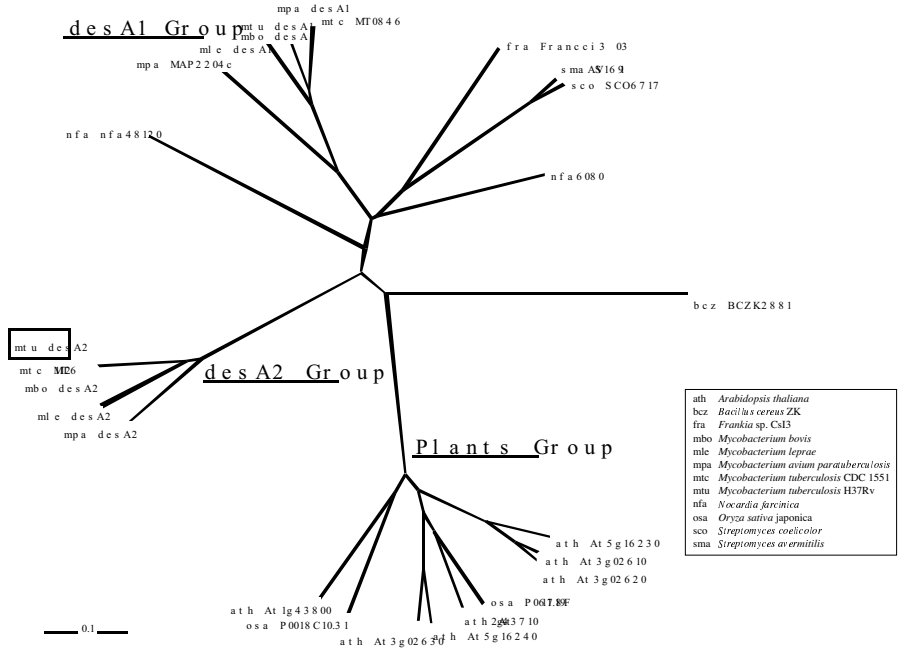

Figure 3: Neighbor-Joining tree of representative sequences in G6 using the programs MAFFT ver.5.6 [13] and Tree View [21]. Names of sequences are the species codes and entry names in KEGG GENES.

2.5 Pfam Motif Search

The Pfam profiles PF00487 and PF07238 [3] were used for the genomic dataset searches using the HMMER package ver.2.3.2 [35] with default parameters.

3 Results

3.1 Clustering Desaturase Sequences into Annotated Relevant Groups

We collected a complete set of desaturase-related proteins based on amino acid sequence similarity using a PSI-BLAST search and clustering analysis with a wide variety of known desaturases as queries. The number of sequences obtained in our results is 710, and 652 of them contain fatty acid desaturase motifs contained in Pfam (PF00487 and PF07238). These desaturases are clustered into eight annotated groups (G1-G4, G6-G9 in Figure 2), one complex group (G5), and four hypothetical groups (G10). The groups from G1 to G3, listed in Figure 2, are desaturases required for the biosynthesis of the PUFA, whilst the groups from G7 to G9 are required for the biosynthesis of MUFA. We checked these results in detail, by comparing the clustering results with known experimental data, and found that obviously unrelated sequences did not fall into the same group except for the case of G5. In groups G2 and G6, two separate groups of enzymes (bacterial desaturases and $\Delta 12$ desaturases) form a single cluster, these are mentioned below.

The largest cluster is the stearoyl-CoA desaturase group, which is separated into three groups by the organisms in which they are found: those only found in bacteria (G7), only in eukaryote (G9), and those found in both (G8). Almost all eukaryotes have proteins from G8 or G9, and many bacteria have proteins in G7 or G8.

Other noteworthy groups are G2 and G6, where some bacterial proteins fall near the cluster of known plant desaturases. Figure 3 shows the Neighbor-Joining (NJ) tree composed of all 17 bac-

Table 2: Number of desaturase related proteins in eukaryotes.

	organism	total	ACP	Δ12/15	Δ5/6	Δ9	Δ4s	com	other
Mammal	*Homo sapiens*	8	0	0	3	2	2	0	1
	Pan troglodytes	7	0	0	2	2	2	0	1
	Macaca mulatta (d)	13	0	0	4	5	2	0	2
	Mus musculus	16	0	0	3	7	5	0	1
	Rattus norvegicus	10	0	0	3	3	3	0	1
	Canis familiaris	9	0	0	3	2	2	0	2
	Bos taurus (p)	11	0	0	6	2	2	0	1
	Sus scrofa (p)	1	0	0	0	1	0	0	0
	Monodelphis domestica (d)	8	0	0	1	2	3	0	2
Bird	*Gallus gallus*	9	0	0	4	2	2	0	1
Amphibian	*Xenopus laevis* (p)	6	0	0	2	1	3	0	0
	Xenopus tropicalis (p)	3	0	0	0	1	2	0	0
Fish	*Danio rerio* (d)	7	0	0	1	3	2	0	1
	Fugu rubripes (d)	5	0	0	0	2	2	0	1
	Tetraodon nigroviridis (d)	5	0	0	0	2	1	0	2
Ascidian	*Ciona intestinalis* (d)	4	0	0	0	1	2	1	0
Insect	*Drosophila melanogaster*	7	0	0	0	6	1	0	0
	Drosophila pseudoobscura (d)	7	0	0	0	6	1	0	0
	Anopheles gambiae (d)	13	0	0	0	11	2	0	0
	Apis mellifera (d)	25	0	0	0	25	0	0	0
	Bombyx mori (d)	9	0	0	0	8	1	0	0
Nematode	*Caenorhabditis elegans*	9	0	2	0	3	2	2	0
	Caenorhabditis briggsae (d)	10	0	2	0	3	2	3	0
Plant	*Arabidopsis thaliana*	26	8	6	0	9	1	2	0
	Medicago truncatula (d)	5	0	2	0	1	1	1	0
	Oryza sativa japonica (d)	22	9	11	0	0	1	1	0
	Cyanidioschyzon merolae	4	0	1	0	2	1	0	0

	organism	total	ACP	Δ12/15	Δ5/6	Δ9	Δ4s	com	other
Fungi	*Saccharomyces cerevisiae*	1	0	0	0	1	0	0	0
	Saccharomyces paradoxus (d)	1	0	0	0	1	0	0	0
	Saccharomyces mikatae (d)	1	0	0	0	1	0	0	0
	Saccharomyces bayanus (d)	1	0	0	0	1	0	0	0
	Ashbya gossypii	5	0	1	0	1	1	1	1
	Khyveromyces lactis (d)	6	0	2	0	1	1	1	1
	Khyveromyces waltii (d)	6	0	2	0	1	1	1	1
	Debaryomyces hansenii (d)	6	0	2	0	1	1	1	1
	Candida albicans	7	0	2	0	1	1	1	2
	Candida glabrata (d)	1	0	0	0	1	0	0	0
	Yarrowia lipolytica (d)	4	0	1	0	1	1	1	0
	Schizosaccharomyces pombe	2	0	0	0	1	1	0	0
	Neurospora crassa (d)	6	0	2	0	1	1	1	1
	Magnaporthe grisea (d)	6	0	2	0	1	1	1	1
	Gibberella zeae (d)	8	0	2	0	1	1	2	2
	Aspergillus nidulans	8	0	2	0	2	1	1	2
	Aspergillus fumigatus	8	0	2	0	2	1	2	1
	Aspergillus oryzae	11	0	2	0	4	1	3	1
	Cryptococcus neoformans JEC21	4	0	1	0	1	1	1	0
	Cryptococcus neoformans B-3501A (d)	4	0	1	0	1	1	1	0
	Ustilago maydis (d)	4	0	1	0	1	1	1	0
	Encephalitozoon cuniculi	0	0	0	0	0	0	0	0
Protist	*Dictyostelium discoideum*	7	0	0	0	3	1	3	0
	Plasmodium falciparum	1	0	0	0	1	0	0	0
	Plasmodium yoelii (d)	1	0	0	0	1	0	0	0
	Plasmodium berghei (d)	2	0	0	0	2	0	0	0
	Plasmodium chabaudi (d)	1	0	0	0	1	0	0	0
	Cryptosporidium parvum	0	0	0	0	0	0	0	0
	Cryptosporidium hominis	0	0	0	0	0	0	0	0
	Theileria annulata	0	0	0	0	0	0	0	0
	Theileria parva	0	0	0	0	0	0	0	0
	Trypanosoma brucei	4	0	1	0	1	1	1	0
	Trypanosoma cruzi	9	0	2	0	1	2	2	2
	Leishmania major	12	0	3	0	2	4	3	0
	Entamoeba histolytica	0	0	0	0	0	0	0	0
	Giardia lamblia (d)	0	0	0	0	0	0	0	0

ACP : acyl-ACP desaturases in G6 group
Δ12/15 : Δ12 desaturases or Δ15 desaturases in G2 or G3 groups
Δ5/6 : Δ5 desaturases or Δ6 desaturases in G1 group
Δ9 : Stearyol-CoA desaturases in G8 or G9 groups
Δ4s : Δ4 sphingolipid desaturases in G4 group
com : variety of desaturases in G5 group
other : hypothetical proteins in G10 group
(d) : draft genome
(p) : partial genome
Non-zero values are colored for convenience

terial sequences and nine representative sequences from plants, (after removing very closely related sequences), found in G6. While plant sequences form a single clustered group, the bacterial sequences spread out of the plant branch. In the case of G2, phylogenetic trees indicated a similar result.

On the other hand, it is very natural that Δ12 and Δ15 desaturases, which are involved in PUFA synthesis in plants, cyanobacteria, nematodes, fungi, and protists, form a single cluster: G3. We examined the G3 cluster and confirmed that Δ12 and Δ15 desaturases form separate clusters within the group except for the nematode sequences which do cluster, but with a higher threshold. In the same manner, Δ5 and Δ6 desaturases used in PUFA synthesis in vertebrates also cluster into the same group: G1. These two proteins could not be separated using the sequence similarity measure.

Although the average number of bacterial desaturase-related proteins, whose function is still unknown, is smaller than the number of eukaryotic ones, they are found not only in G2 and G6, mentioned above, but also in G5, G7, G8, and G10. No desaturase-related proteins of archaea were found in this study.

3.2 The Repertoire of Fatty Acid Desaturases in Eukaryotes

By examination of the clustering results, the number of desaturases, along with their regioselectivity is obtained for all eukaryotes (Table 2). *Encephalitozoon cuniculi*, which as a parasite strongly depends on its host organism, is the only organism, outside the protists, that does not contain any desaturases. A

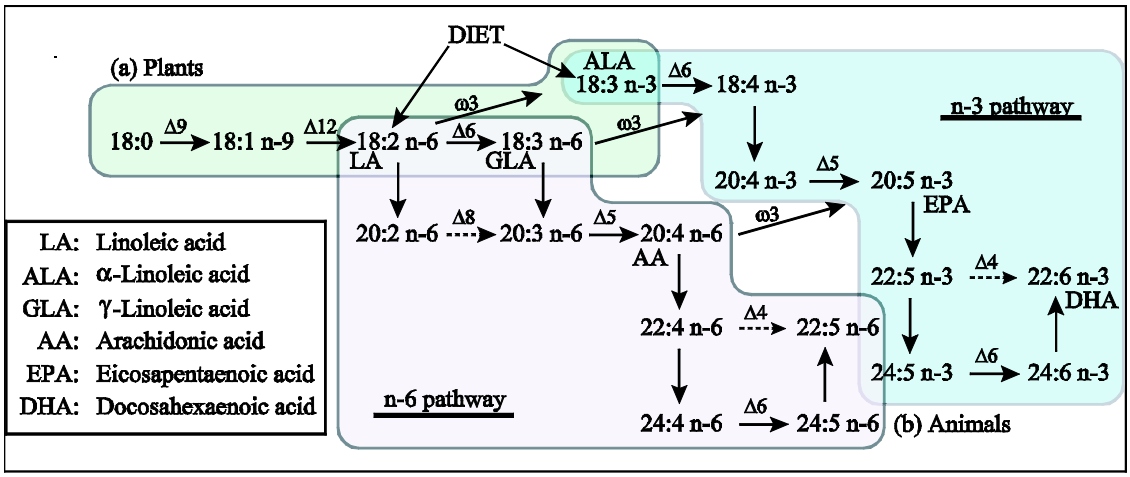

Figure 4: The reference pathway of major PUFA synthesis: To designate an individual fatty acid within a family of structurally related acids, the n- nomenclature is used [6]. Here, the position of the first double bond from the methyl end is described e.g. 18:3 n-6 (GLA) indicates that the double bond closest to the methyl end is 6 carbons from the methyl group and in the Δ12 position. The n-6 pathway and the n-6 family are composed of n-6 fatty acids. (a) Plants generally synthesize the LA and GLA using Δ9, Δ12, and Δ15 (ω3) desaturases. (b) Animals generally synthesize AA, EPA, and DHA with Δ5 and Δ6 desaturases starting from LA and ALA obtained from the diet. Dashed lines indicate alternative pathways with Δ4 or Δ8 desaturases.

desaturase homolog was identified in *Plasmodium falciparum*, which is also an intracellular pathogen exhibiting a high mutation rate, hence it is more likely that *E. cuniculi* possesses no desaturase homologs rather than we failed to detected it. Stearoyl-CoA desaturases are also found in a wide range of eukaryotic organisms, although many protists do not possess them. On the other hand, higher plants and vertebrates dominate the acyl-ACP desaturases and Δ5 and Δ6 desaturases, respectively. Δ12 and Δ15 desaturases, for the insertion of second and third double bonds, belong to nematodes, plants, fungi and three protists (*Trypanosoma brucei, Trypanosoma cruzi, Leishmania major*). These protists are euglenozoa, whose repertoire of desaturases is similar to plants.

4 Discussion

4.1 Biosynthetic Pathway of PUFA in Eukaryotes

The Acyl-ACP, Δ12, Δ15, Δ5, and Δ6 desaturases listed in Table 2 are required for the synthesis of major PUFAs, such as linoleic acids, arachidonic acids, docosahexaenoic acids. The range of biosynthesis of fatty acids that each organism is capable of is different in each organism; hence, to investigate the fatty acid variation in each organism, we constructed a reference pathway of fatty acid biosynthesis, which is integrated over multiple species (Figure 4). In plants, acyl-ACP, Δ12, and Δ15 desaturases introduce double bonds into the 9th, the 12th, and the 15th carbon, respectively [20]. This is followed by the synthesis of linoleic acids and α-linoleic acids (Figure 4 (a)). Similarly to previous studies, our result supports both acyl-ACP and Δ12/Δ15 desaturases being present in higher plants (Table 2). In general, animals take linoleic acids and α-linoleic acids synthesized by plants in the diet, and then Δ5/Δ6 desaturases introduce double bonds into them to synthesize more multivalent

unsaturated fatty acids. Almost all vertebrates used in this analysis, retain multiple homologous proteins in the $\Delta 5/\Delta 6$ desaturases category. Proteins in this group, such as hsa:9415 and hsa:3995 from *H. sapiens*, mmu:76267 and mmu:56473 from *M. musculus*, and rno:84575 and rno:83512 from *R. norvegicus* are experimentally confirmed to work as $\Delta 5$ and $\Delta 6$ desaturases, respectively [10]. Only one homolog dre:140615 from *Danio rerio* in the group is confirmed to show the activity of both $\Delta 5$ and $\Delta 6$ desaturases [10]. Our clustering result suggests that other vertebrates also have both desaturases.

Although all sequences in G1 belong to vertebrates, a few non-vertebrate organisms have been reported to be able to synthesize EPA or DHA, because they possess an alternative pathway. For example, *Acanthamoeba castellanii* uses $\Delta 8$ desaturase [26], and *Thraustochytrium* sp ATCC21685. uses $\Delta 4$ desaturase to synthesize docosahexaenoic acids [23]. In our analysis, these proteins fall into the complex group (G5), and do not make their own functional cluster. Hence, some proteins, which have a similar function to these desaturases, may be included in this group. Although *C. elegans* belongs to the animal kingdom, it can, like plants, synthesize linoleic acids and α-linoleic acids using $\Delta 12$ and $\Delta 15$ desaturases [18]. Additionally, it also has the ability to synthesize arachidonic acid with $\Delta 5/\Delta 6$ desaturases, which unexpectedly fall into the complex G5 group, not G1, in our clustering result. This suggests that the *C. elegans* $\Delta 5$ desaturase was developed later by gene duplication from the original $\Delta 6$ desaturase. There is still not enough information to determine whether the $\Delta 12$ and $\Delta 5$ desaturases in nematodes are due to horizontal gene transfer or that other animals originally contained these enzymes, but subsequently lost them.

Generally, while the repertoire of desaturases amongst organisms in each kingdom is similar to each other, the number of paralogs is not identical even among closely-related organisms, such as mammals. This may be due to differences in organism specific reactions. However, not enough data are available to examine this hypothesis by sequence analysis at this time.

We have only discussed the clustering and pathway mapping of desaturases, however, the actual biosynthetic pathways these enzymes are part of, includes elongases, which we did not examine in this work. These enzymes have a specificity dependent on the acyl-chain length and the position of pre-existing double bonds [11]. In the next step of the study of the fatty acid variation, elongases should be mapped onto the pathway as well as desaturases.

4.2 Bacterial Fatty Acid Desaturases

Bacteria have an alternative pathway to produce PUFAs without desaturases and some recent reports have clarified the functions and the regulations of the desaturases involved in MUFA synthesis. For example, it experimentally verified that the desaturase from *B. subtilis* has regioselectivity of $\Delta 5$ [1]. Recently, a model of the regulatory pathway of involving a desaturase was proposed to adapt to temperature shifts by inserting double bonds into membrane acyl-chains [15]. In our results, the G2 group, composed of desaturases from plants and cyanobacteria, contains this desaturase along with other bacterial sequences, such as *Thermus thermophilus, Sinorhizobium meliloti*, and *Rhodopseudomonas palustris*. Despite the lack of experimental proof, the fact that these proteins show sequence similarities to known proteins in the same group and have the histidine motif (Pfam: PF00487) suggests that they are desaturase homologs and desaturases are utilized in bacteria more extensively than originally thought.

Another example is two desaturase-like proteins of *Mycobacterium tuberculosis* , which fall into the acyl-ACP group (G6). These proteins have been suggested to be desaturases for the synthesis of a complex fatty acid called mycolic acid, which forms the cell wall and guards against macrophage attack. A recent study determined the 3D structures of these two proteins [8], and, while one of them (desA2) could be folded, the other (desA1) could not. Hence, desA1 may not be functional. We confirmed that almost all the sequences in G6, including desA1 and desA2, contain the domain (Pfam: PF07238). In Figure 3, the desA2 group is separated from both the desA1 group and the plants group, indicating that the proteins of *Mycobacterium bovis, Mycobacterium avium paratuberculosis,*

Mycobacterium leprae, Mycobacterium tuberculosis CDC1551 in the desA2 group are good candidates to be desaturases.

Acknowledgments

We thank Alex Gutteridge for critical reading of our manuscript. This work was supported by grants from Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists, the Ministry of Education, Culture, Sports, Science and Technology of Japan, and Japan Science and Technology Agency. The computational resource was provided by the Bioinformatics Center, Institute for Chemical Research, Kyoto University.

Abbreviations

AA, arachidonic acid; ACP, acyl carrier protein; ALA, α-linoleic acid; DHA, docosahexaenoic acid; EPA, eicosapentaenoic acid; GLA, γ-linoleic acid; LA, linoleic acid; MUFA, monounsaturated fatty acid; NJ, Neighbor-Joining; PUFA, polyunsaturated fatty acid

References

[1] Altabe, S.G., Aguilar, P., Caballero, G.M., and de Mendoza, D., The *Bacillus subtilis* acyl lipid desaturase is a delta5 desaturase, *J. Bacteriol.*, 185:3228–3231, 2003.

[2] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.*, 25:3389–3402, 1997.

[3] Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., Studholme, D.J., Yeats, C., and Eddy, S.R., The Pfam protein families database, *Nucleic Acids Res.*, 32:D138–141, 2004.

[4] Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L., GenBank, *Nucleic Acids Res.*, 34:D16–20, 2006.

[5] Birney, E., Andrews, D., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cox, T., Cunningham, F., Curwen, V., Cutts, T., Down, T., Durbin, R., Fernandez-Suarez, X.M., Flicek, P., Graf, S., Hammond, M., Herrero, J., Howe, K., Iyer, V., Jekosch, K., Kahari, A., Kasprzyk, A., Keefe, D., Kokocinski, F., Kulesha, E., London, D., Longden, I., Melsopp, C., Meidl, P., Overduin, B., Parker, A., Proctor, G., Prlic, A., Rae, M., Rios, D., Redmond, S., Schuster, M., Sealy, I., Searle, S., Severin, J., Slater, G., Smedley, D., Smith, J., Stabenau, A., Stalker, J., Trevanion, S., Ureta-Vidal, A., Vogel, J., White, S., Woodwark, C., and Hubbard, T.J., Ensembl 2006, *Nucleic Acids Res.*, 34:D556–561, 2006.

[6] Dennis, E.V. and Jean, E.V. Biochemistry of lipids, lipoproteins and membranes, Elsevier Science, 2002.

[7] Dobrzyn, A. and Ntambi, J.M., The role of stearoyl-CoA desaturase in the control of metabolism, *Prostaglandins Leukot. Essent. Fatty Acids*, 73:35–41, 2005.

[8] Dyer, D.H., Lyle, K.S., Rayment, I., and Fox, B.G., X-ray structure of putative acyl-ACP desaturase DesA2 from *Mycobacterium tuberculosis* H37Rv, *Protein Sci.*, 14:1508–1517, 2005.

[9] Fahy, E., Subramaniam, S., Brown, H.A., Glass, C.K., Merrill, A.H. Jr., Murphy, R.C., Raetz, C.R., Russell, D.W., Seyama, Y., Shaw, W., Shimizu, T., Spener, F., van Meer, G., VanNieuwenhze, M.S., White, S.H., Witztum, J.L., and Dennis, E.A., A comprehensive classification system for lipids, *J. Lipid Res.*, 46:839–861, 2005.

[10] Hastings, N., Agaba, M., Tocher, D.R., Leaver, M.J., Dick, J.R., Sargent, J.R., and Teale, A.J., A vertebrate fatty acid desaturase with Delta 5 and Delta 6 activities, *Proc. Natl. Acad. Sci. USA*, 98:14304–14309, 2001.

[11] Jakobsson, A., Westerberg, R., and Jacobsson, A., Fatty acid elongases in mammals: Their regulation and roles in metabolism, *Prog. Lipid Res.*, 45:237–249, 2006.

[12] Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M., From genomics to chemical genomics: new developments in KEGG, *Nucleic Acids Res.*, 34:D354–D357, 2006.

[13] Katoh, K., Kuma, K., Toh, H., and Miyata, T., MAFFT version 5: improvement in accuracy of multiple sequence alignment, *Nucleic Acids Res.*, 33:511–518, 2005.

[14] Kim, J.H., Kim, Y., Lee, S.D., Lopez, I., Arnold, R.S., Lambeth, J.D., Suh, P.G., and Ryu, S.H., Selective activation of phospholipase D2 by unsaturated fatty acid, *FEBS Lett.*, 454:42–46, 1999.

[15] Mansilla, M.C. and de Mendoza, D., The *Bacillus subtilis* desaturase: a model to understand phospholipid modification and temperature sensing, *Arch. Microbiol.*, 183:229–235, 2005.

[16] Marszalek, J.R. and Lodish, H.F., Docosahexaenoic acid, fatty acid-interacting proteins, and neuronal function: breastmilk and fish are good for you, *Annu. Rev. Cell Dev. Biol.*, 21:633–657, 2005.

[17] Martin, C.E., Oh, C.S., Kandasamy, P., Chellapa, R., and Vemula, M., Yeast desaturases, *Biochem. Soc. Trans.*, 30:1080–1082, 2002.

[18] Morimoto, K.C., Van Eenennaam, A.L., DePeters, E.J., and Medrano, J.F., Endogenous production of n-3 and n-6 fatty acids in mammalian cells, *J. Dairy Sci.*, 88:1142–1146, 2005.

[19] Nakamura, M.T. and Nara, T.Y., Structure, function, and dietary regulation of delta6, delta5, and delta9 desaturases, *Annu. Rev. Nutr.*, 24:345–376, 2004.

[20] Napier, J.A., Michaelson, L.V., and Stobart, A.K., Plant desaturases: harvesting the fat of the land, *Curr. Opin. Plant Biol.*, 2:123–127, 1999.

[21] Page, R.D., TreeView: an application to display phylogenetic trees on personal computers, *Comput. Appl. Biosci.*, 12:357–358, 1996.

[22] Pearson, W.R., Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms, *Genomics*, 11:635–650, 1991.

[23] Qiu, X., Hong, H., and MacKenzie, S.L., Identification of a Delta 4 fatty acid desaturase from *Thraustochytrium* sp. involved in the biosynthesis of docosahexanoic acid by heterologous expression in *Saccharomyces cerevisiae* and *Brassica juncea*, *J. Biol. Chem.*, 276:31561–31566, 2001.

[24] R Development Core Team, R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2003.

[25] Saito, T., Kato, A., Ochiai, H., and Morita, N., Temperature adaptation in *Dictyostelium*: role of Delta5 fatty acid desaturase, *Microbiology*, 151:113–119, 2005.

[26] Sayanova, O., Haslam, R., Qi, B., Lazarus, C.M., and Napier, J.A., The alternative pathway C20 Delta8-desaturase from the non-photosynthetic organism *Acanthamoeba castellanii* is an atypical cytochrome b5-fusion desaturase, *FEBS Lett.*, 580:1946–1952, 2006.

[27] Shanklin, J. and Cahoon, E.B., Desaturation and related modifications of fatty acids1, *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 49:611–641, 1998.

[28] Smith, T.F. and Waterman, M.S., Identification of common molecular subsequences, *J. Mol. Biol.*, 147:195–197, 1981.

[29] Sperling, P., Ternes, P., Zank, T.K., and Heinz, E., The evolution of desaturases, *Prostaglandins Leukot. Essent. Fatty Acids*, 68:73–95, 2003.

[30] Sprecher, H., Biochemistry of essential fatty acids, *Prog. Lipid Res.*, 20:13–22, 1981.

[31] Tonon, T., Harvey, D., Larson, T.R., and Graham, I.A., Identification of a very long chain polyunsaturated fatty acid Delta4-desaturase from the microalga *Pavlova lutheri*, *FEBS Lett.*, 553:440–444, 2003.

[32] Tonon, T., Harvey, D., Qing, R., Li, Y., Larson, T.R., and Graham, I.A., Identification of a fatty acid Delta11-desaturase from the microalga *Thalassiosira pseudonana*, *FEBS Lett.*, 563:28–34, 2004.

[33] Yoshizawa, A.C., Kawashima, S., Okuda, S., Fujita, M., Itoh, M., Moriya, Y., Hattori, M., and Kanehisa, M., Extracting sequence motifs and the phylogenetic features of SNARE-dependent membrane traffic, *Traffic*, 7:1104–1118, 2006.

[34] `http://bioruby.org/`.

[35] `http://hmmer.wustl.edu/`.

[36] `http://www.r-project.org/`.