

A statistical anomaly indicates symbiotic origins of eukaryotic membranes

Suneyna Bansal and Aditya Mittal

Kusuma School of Biological Sciences, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India

ABSTRACT Compositional analyses of nucleic acids and proteins have shed light on possible origins of living cells. In this work, rigorous compositional analyses of ~5000 plasma membrane lipid constituents of 273 species in the three life domains (archaea, eubacteria, and eukaryotes) revealed a remarkable statistical paradox, indicating symbiotic origins of eukaryotic cells involving eubacteria. For lipids common to plasma membranes of the three domains, the number of carbon atoms in eubacteria was found to be similar to that in eukaryotes. However, mutually exclusive subsets of same data show exactly the opposite—the number of carbon atoms in lipids of eukaryotes was higher than in eubacteria. This statistical paradox, called Simpson's paradox, was absent for lipids in archaea and for lipids not common to plasma membranes of the three domains. This indicates the presence of interaction(s) and/or association(s) in lipids forming plasma membranes of eubacteria and eukaryotes but not for those in archaea. Further inspection of membrane lipid structures affecting physicochemical properties of plasma membranes provides the first evidence (to our knowledge) on the symbiotic origins of eukaryotic cells based on the "third front" (i.e., lipids) in addition to the growing compositional data from nucleic acids and proteins.

Monitoring Editor

David G. Drubin
University of California,
Berkeley

Received: Jun 11, 2014

Revised: Dec 22, 2014

Accepted: Jan 22, 2015

INTRODUCTION

Over the years, the evolutionary origins of eukaryotic cells have been studied with respect to two key components in living systems—nucleic acids and proteins. While genes involved in replication, transcription, and translation indicate similarity between archaea and eukaryotes, metabolic genes indicate eukaryotes are closer to eubacteria (Rivera *et al.*, 1998). On the other hand, the universal tree of single-stranded rRNA clusters eukaryotes near to archaea (Woese, 2000). To date, there are a number (more than 20) different models to explain the origin of eukaryotic cells (Martin and Muller, 1998; Martin, 2005; Martin and Koonin 2006; Moreira and Lopez-Garcia, 1998; Vesteg and Krajcovic, 2006; Forterre, 2011; Baum and Baum, 2014). The majority of these models are based on the merger of two prokaryotes to form ancient eukaryotic cells.

There are two basic schools of thought: 1) One-merger theory: the merger of eubacteria and archaea led to the formation of the nucleus and mitochondria at the same time (around 2 giga-annum [Ga]; Hartman and Fedorov, 2002). 2) Two-merger theory: the merger of archaea and eubacteria first formed the nucleus of the eukaryotic cell around 2.7 Ga; another alpha-Proteobacteria was later engulfed to produce mitochondria around 2 Ga (Margulis *et al.*, 2006). Literature on the one-merger theory indicates that early eukaryotes (e.g., *Trichomonas foetus*) are not amitochondriates and contain organelles, called *hydrogenomas*, that are analogous to mitochondria (Lindmark and Muller, 1973). On the other hand, the two-merger theory supports the presence of amitochondriates (Fuerst and Webb, 1991). The majority of these studies are based on genomics and proteomics data. Only a few scattered reports comment on the evolution of eukaryotic membranes, and those are also based largely on membrane proteins and biosynthetic pathways in cells (Pereto *et al.*, 2004; Daiyasu *et al.*, 2005; Mulikidjanian *et al.*, 2008; Lombard and Moreira, 2011; Lombard *et al.*, 2012a). Thus, in spite of the importance of amphipathic lipid molecules in formation of compartments as whole cells and within cells, emphasis on the essential "third" component (i.e., amphipathic lipid molecules) has been limited. Self-aggregation/assembly of amphipathic molecules was a crucial step in the origin of life (Monnard and Deamer, 2002; Rasmussen *et al.*, 2009; Chen and Walde, 2010). Appearance of the

This article was published online ahead of print in MBoC in Press (<http://www.molbiolcell.org/cgi/doi/10.1091/mbc.E14-06-1078>) on January 28, 2015.

Address correspondence to: Aditya Mittal (amittal@bioschool.iitd.ac.in).

Abbreviations used: G-1-P, glycerol-1-phosphate; G-3-P, glycerol-3-phosphate; Ga, giga-annum.

© 2015 Bansal and Mittal. This article is distributed by The American Society for Cell Biology under license from the author(s). Two months after publication it is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

"ASCB®" "The American Society for Cell Biology®," and "Molecular Biology of the Cell®" are registered trademarks of The American Society for Cell Biology.

first single cell(s), unit(s) of life, required a boundary enclosing a “system” separated from the surrounding environment. Formation of this boundary, the plasma membrane of a cell, was driven by self-aggregation of amphipathic molecules that also subsequently led to intracellular compartment formation, resulting in a eukaryotic cell. In the course of evolution, cell–cell interactions, including possible exchange of membrane lipids (Reed, 1968; Pagano and Huang, 1975), might have allowed the emergence of a pool of shared membrane lipids in their plasma membranes. Simultaneously, direct interaction of these lipid boundaries with a dynamic outer environment could also have contributed to compositional modifications in plasma membranes (Goldfine, 2010; Koga, 2012; Oger and Cario, 2013). While these theories are certainly of importance, they provide limited insights into understanding the evolution of eukaryotic cells from the point of view of their plasma membranes. In view of this, and inspired by compositional approaches applied to genomics and proteomics, we initiated a study of plasma membranes of the three domains of life (archaea, eubacteria, and eukaryotes) to investigate the origin of life from the point of view of the “third front” (the first two “fronts” being nucleic acids and proteins, respectively). As a first step, we collected amphipathic (nonprotein) components of plasma membranes from 273 species in the three domains of life—this yielded more than 5000 membrane lipids. A comprehensive analysis involving chemical composition–based classification of these lipids revealed a remarkable statistical paradox. For lipids common to plasma membranes of the three domains, the number of carbon atoms in eubacteria was found to be similar to that in eukaryotes. However, mutually exclusive subsets of the same data individually show exactly the opposite, that is, the number of carbon atoms in lipids of eukaryotes was higher than in eubacteria. This statistical paradox, called the Simpson’s paradox (Wagner, 1982), was absent for lipids in archaea and for lipids not common to the plasma membranes of the three domains. This remarkable finding directly indicates (hidden) interaction(s) and/or association(s) (Simpson, 1951) in lipids forming plasma membranes of eubacteria and eukaryotes but not for those in archaea. Further exploration of lipid structures responsible for physicochemical properties of plasma membranes leads to the inference that the first compartmentalized cells, that is, eukaryotic single cells, had symbiotic origins with emphasis on eubacterial components as a source of their plasma membranes. In addition, membrane lipids of mitochondria and chloroplasts of eukaryotes suggest that membranes of these key organelles shared significant similarities with the eubacterial membrane lipids. Thus, through compositional analysis of lipidomics data and on the basis of the “third front” (i.e., lipids), we provide the first evidence (to our knowledge) of symbiotic origins of eukaryotic cells, with a larger role of eubacteria in formation of both plasma membranes and mitochondrial membranes.

RESULTS AND DISCUSSION

Common and uncommon membrane lipids

A comparative analysis of membrane lipids from each domain indicated three common classes that include 10 different compositions and a total of 3911 chemical structures of membrane lipids. We called them “common membrane lipids” (Figure 1). The remaining 20 classes (82 different compositions), which include a total of 1471 chemical structures of membrane lipids, were considered “uncommon membrane lipids.” Apart from membrane lipids common to all three domains, we found that eubacteria and eukaryotes alone shared 11 more lipid compositions, and archaea and eubacteria also uniquely shared three more membrane lipid compositions. These findings are consistent with the observation that eubacteria and eukaryotes share more similar physiological conditions than

archaea—more common lipid classes would be expected to be found between eubacteria and eukaryotes as compared with archaea and eukaryotes. In our data set, we found the maximum number of common lipids between eubacteria and eukaryotes (=21) was greater than those shared by eubacteria and archaea (=13) and eukaryotes and archaea (=10); we also found 21 membrane lipid types unique to archaea, 21 for eubacteria, and 26 for eukaryotes.

Variation in the total carbon content of membrane lipids of each domain

Inspired by compositional analysis of nucleic acids and proteins (Nussinov, 1984; Knight *et al.*, 2001; Kanhere and Bansal, 2005; Mittal *et al.*, 2010; Mittal and Jayaram, 2011, 2012; Mittal and Acharya, 2013), we carried out a compositional analysis of chemical structures of these membrane lipids. For simplicity, we counted the total number of carbon atoms in each membrane lipid and the number of carbon atoms in the hydrophobic tails of each membrane lipid. Ten concentric circles, each circle representing an evolutionary time point in Ga units, were drawn to represent the total number of carbon atoms (Figure 2, A–C) and tail carbon atoms (Figure 3, A–C) in membrane lipids of each domain across evolutionary time; we show these separately for total (Figures 2A and 3A), common (Figures 2B and 3B), and uncommon membrane lipids (Figures 2C and 3C). To compare the average size and hydrophobic tail lengths of membrane lipids, we calculated the mean value of carbon atoms for each domain separately for each group of total, common, and uncommon lipids. Then we further divided the total number of carbon atoms in each group into two mutually exclusive subsets; smaller-sized membrane lipids with a number of carbon atoms less than and equal to 60 and larger membrane lipids with more than 60 carbon atoms in their structures. Similarly, for tail carbon atoms in lipids, two subsets were formed in such a way that one includes those with structures with a number of tail carbon atoms less than and equal to 50, and the other includes those with tail carbons numbering more than 50. Mean values of carbon atoms were also computed for these two mutually exclusive subsets (Figures 2, D–F, and 3, D–F). Three independent *t* tests (two-sample, assuming equal or unequal variances depending on the *F* test for homo- or heteroscedasticity of the data sets) with $\alpha = 0.05$ were applied on each of the above data sets. For lipids common to plasma membranes of the three domains, the number of carbon atoms in eubacteria was found to be similar to that in eukaryotes (extreme left black and gray bars in Figures 2E and 3E) for both total carbon atoms and carbon atoms in only the fatty acid tails. However, mutually exclusive subsets of the same data showed exactly the opposite trend, that is, the number of carbon atoms in lipids of eukaryotes was higher than in eubacteria (middle and extreme right black and gray bars in Figures 2E and 3E).

Simpson’s paradox supports the symbiotic origin of eukaryotes involving eubacteria

The remarkable and complete reversal of association between the average tail lengths of eubacterial and eukaryotic common membrane lipids with their mutually exclusive subsets, shown in Figures 2E and 3E, presented a fascinating statistical anomaly to us. After a substantial literature review, we found that this statistical anomaly, called “Simpson’s paradox” (Wagner, 1982), has been reported as a rare phenomenon in biological data, especially clinical data (Tu *et al.*, 2005; Rucker and Schumacher, 2008). To our knowledge, our findings in this work are the first example of Simpson’s paradox in biological data at a molecular level through the compositional analysis of chemical structures. Simpson (Simpson, 1951) established

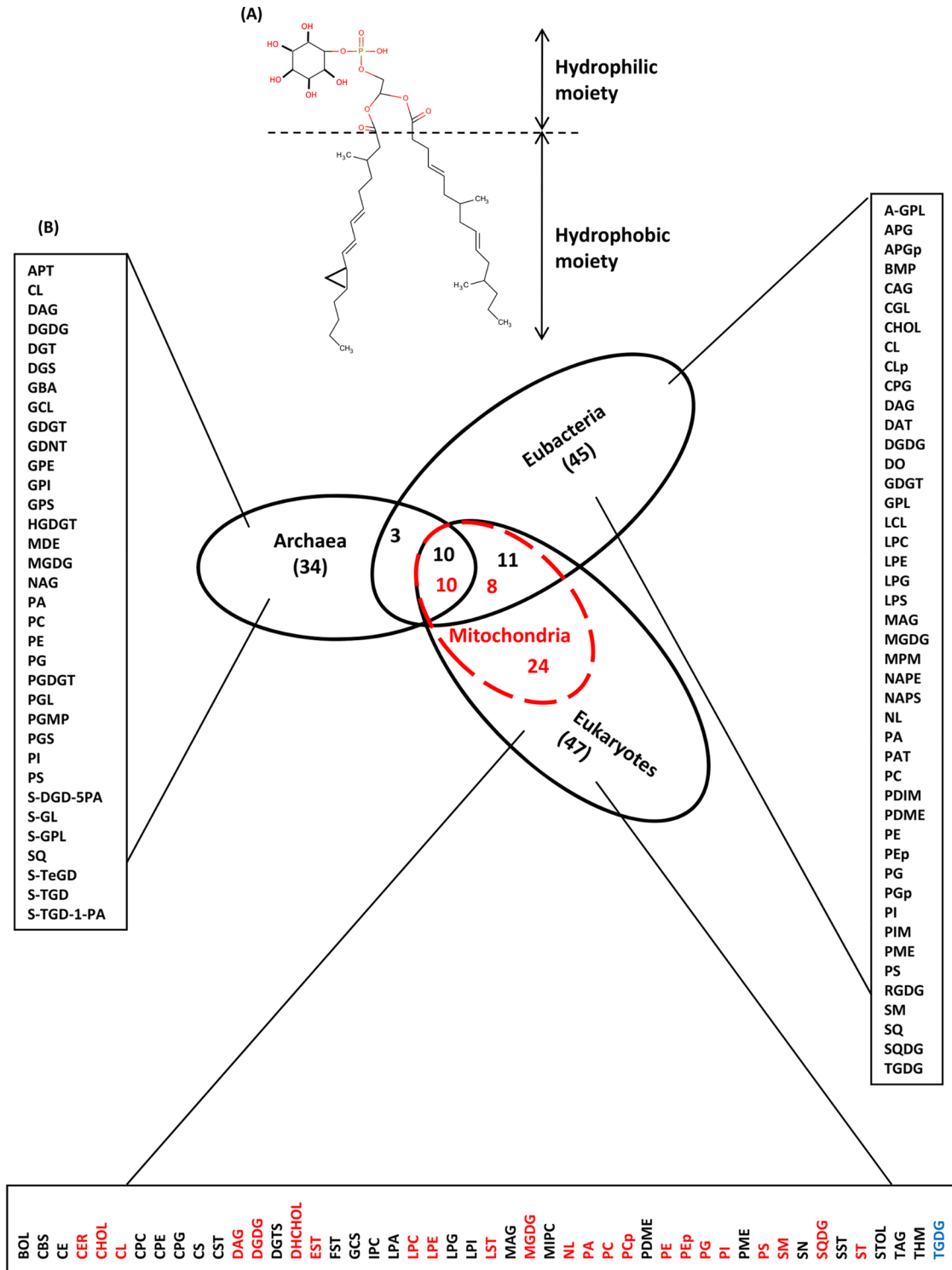


FIGURE 1: Lipid “class diagram” for all three domains. (A) Representation of a generic chemical structure for a membrane lipid composed of two regions: a hydrophilic part made up of polar atoms and a hydrophobic part composed of nonpolar atoms or carbon tails. Membrane lipids usually consist of head rings that are generally made up of sugar moieties, and hydrocarbon tail regions with multiple bonds, branched methyl groups, and cyclopropane rings. However, it is important to note here that archaeal lipids contain ether linkages instead of ester bonds, and their tetraethers usually contain cyclopentane rings instead of cyclopropane rings in their tail region. (B) Different membrane lipids of all three domains were collected and represented using a Venn diagram. The intersection of two circles represents the number of common lipids in two respective domains, and the common intersection area of all three circles represents the common lipids shared by all three domains of life. Red numbers represent organellar membrane lipids. Red text represents lipids common to eukaryotic plasma membranes and organellar membranes. Blue text represents lipids found in the membranes of mitochondria but absent in the plasma membranes of eukaryotes. Abbreviations used in this figure: A-GPL, amino glycopospholipid; APG, acyl phosphatidylglycerol; APGp,

that this statistical anomaly, called a paradox due to reversal in association of total data when divided into mutually exclusive data sets, is caused by hidden interactions within the studied variables. However, the exact nature of the hidden interactions is not extractable from this analysis—it simply shows in a clinical setting that the variables are related either through interactions, such as possible exchanges and/or common lineages/origins arising directly out of a common ancestral population (Lindley and Novick, 1981).

Remarkably, Simpson's paradox was observed only in the data of lipids common to eubacteria and eukaryotes and was absent from the data of uncommon lipids between these two domains—the only exception was seen in the total carbon content of their uncommon lipids (Figure 2F). However, this paradox was completely absent from the data of the total carbon content of eubacterial and eukaryotic uncommon lipids after removal of 11 more uniquely shared common lipid types between eubacteria and eukaryotes (inset in Figure 2F). Interestingly, the data from lipids of archaea was independent of this paradox. This significant finding clearly indicates (hidden) interaction(s) and/or association(s) among only the common membrane lipids in plasma membranes of eubacteria and eukaryotes but not those in archaea (regardless of whether they are common or uncommon). In further support of our findings, *t* tests showed *p* values > 0.05 for the total carbon and tail carbon content of membrane lipids of eubacteria and eukaryotes only, which also indicates the shared similarities between the plasma membranes of these two domains (Figures 2, D–F, and 3, D–F).

In addition to the above, it is important to consider that the relative abundance of each lipid varies in the plasma membranes of different cells. To account for this, we collected available mole fractions of more than 2000 membrane lipids of archaea, eubacteria, and eukaryotes and reanalyzed the data by computing for the weighted means (i.e., normalized by their relative abundance) of total carbon atoms and total tail carbon atoms in their membrane lipids. Remarkably, we found a similar paradoxical trend in weighted mean values, as observed in the data sets discussed above, using equal weights for each membrane lipid (see Supplemental Figure S1). Here Simpson's paradox was also observed only in the com-

mon lipid types of eubacteria and eukaryotes and was absent in the data of uncommon lipids between these two domains and in archaea. Student's *t* tests on the weighted means of the total carbon atoms and total tail carbon atoms also showed *p* values > 0.05 for the membrane lipids of eubacteria and eukaryotes only (see Supplemental Figure S1). Thus our data clearly indicate that the first eukaryotic single cells had symbiotic origins, with emphasis on eubacterial lipid components as a source of their plasma membranes. To further explore the nature of the symbiotic origins, we extended our analyses to more structural and compositional parameters of the membrane lipids.

Structural variations responsible for physicochemical properties of plasma membranes reveal a possible exchange of common membrane lipids among domains

We counted the total number of multiple bonds, branched methyl groups, and rings in the tail region of each membrane lipid. Ten concentric circles, each circle representing an evolutionary time point in Ga units, were drawn to represent unsaturation, branched methyl groups, and cyclic rings in the tail regions of membrane lipids of each domain across evolutionary time; we show these separately for total (Figure 4A), common (Figure 4B), and uncommon (Figure 4C) membrane lipids. Further, the mean values of multiple bonds, branching, and tail rings were calculated separately for each domain for each group of total, common, and uncommon lipids (Figure 4, D–F). From archaea to eukaryotes, we observed the increasing mean value of multiple bonds for both common and uncommon lipids, which indicates enhanced lipid mobility, membrane disorder, and fluidity from archaea to eukaryotes (Figure 4D). It is also important to observe here that only common membrane lipids of archaea showed the presence of unsaturation, which is completely absent in their uncommon lipids. It is also interesting that the appearance of oxygen (~2 Ga) might have aided the introduction of unsaturation in the tails of membrane lipids as the presence of multiple bonds was seen in all domains only after ~2 Ga (see Figure 4, A–C). As a lipid's tail branching provides structural stability at high temperatures, resistance to oxidation, reduction of permeability to

amino phosphatidyl glycerol plasmalogens; APT, aminopentane tetrol; BMP, bis monoacyl glycerol phosphate; BOL, brassicasterol; CAG, cholesteryl acyl glucoside; CBS, cerebroside; CE, cholesteryl ester; CER, ceramides; CGL, cholesteryl glucoside; CHOL, cholesterol; CHPG, cholesteryl phosphatidyl glucoside; CL, cardiolipin; CLp, cardiolipin plasmalogen; CPC, ceramide phosphorylcholine; CPE, ceramide phosphorylethanolamine; CPG, phosphatidylglycerol ceramides; CS, cholesterol sulfate; CST, campesterol; DAG, diacylglycerol; DAT, 2,3-di-O-acyltrehalose; DGDG, digalactosyl diacylglycerol; DGS, diglycosyl sulfate; DGT, digalactofuranosyl caldearcheol; DGTS, diacylglycerol trimethylhomoserine; DHCHOL, dehydrocholesterol; DO, diols; EST, ergosterol; FST, fucosterol; GBA, gentiobiosyl archeol; GCL, glycardiolipin; GCS, galactosyl ceramide sulfate; GDGT, glycerol dialkyl glycerol tetraether; GDNT, glycerol dialkyl nonitol tetraether; GPE, glyco-phosphatidyl ethanolamine; GPI, glyco-phosphatidyl inositol; GPL, glycophospholipid; GPS, glyco-phosphatidyl serine; HGDGT, H-shaped glycerol dialkyl glycerol tetraether; IPC, inositol phosphoryl ceramide; LCL, lysyl cardiolipin; LPA, lyso-phosphatidic acid; LPC lyso-phosphatidyl choline; LPE, lyso-phosphatidyl ethanolamine; LPG, lyso-phosphatidyl glycerol; LPI, lyso-phosphatidyl inositol; LPS, lyso-phosphatidyl serine; LST, lanosterol; MAG, monoglycerides; MDE, macrocyclic diether; MGDG, monogalactosyl diacylglycerol; MIPC, mannosyl phosphatidyl inositol ceramide; MPM, mannosyl- β -1-phosphomycoketides; NAG, *N*-acetyl glucosamine; NAPE, *N*-acylphosphatidyl ethanolamine; NAPS, *N*-acylphosphatidyl serine; NL, neutral fatty acids; PA, phosphatidic acid; PAT, polyacyltrehalose; PC, phosphatidylcholine; PDIM, phthiocerol dimycocerosates; PDME, phosphatidyl dimethyl ethanolamine; PE, phosphatidyl ethanolamine; PEp, phosphatidyl ethanolamine plasmalogens; PG, phosphatidyl glycerol; PGDGT, polar glycerol dialkyl glycerol tetraether; PGL, glucopyranosyl galactofuranosyl; PGMP, phosphatidyl glycerol methyl phosphate; PGp, phosphatidyl glycerol plasmalogens; PGS, phosphatidyl glycerol sulfate; PI, phosphatidyl inositol; PIM, mannosides acyls phosphatidyl inositol; PME, phosphatidyl-*N*-methylethanolamine; PS, phosphatidyl serine; RGDG, rhamnosyl galactosyl diacylglycerol; S-DGD-5PA, sulfated diglycosyl diether phosphatidic acid; S-GL, sulfated glycolipids; S-GPL, sulfated glycophospholipids; SM, sphingomyelin; SN, sphinganine; SQ, squalene; SQDG, sulfoquinovosyl diglycerol; SST, stigmasterol; ST, sterol; S-TeGD, sulfated tetraglycosyl diether; S-TGD, sulfated triglycosyl diether; S-TGD-1-PA, sulfated triglycosyl diether phosphatidic acid; STOL, β -sitosterol; TAG, triglycerides; TGDG, trigalactosyl diacylglycerol; THM, tetrahymanol.

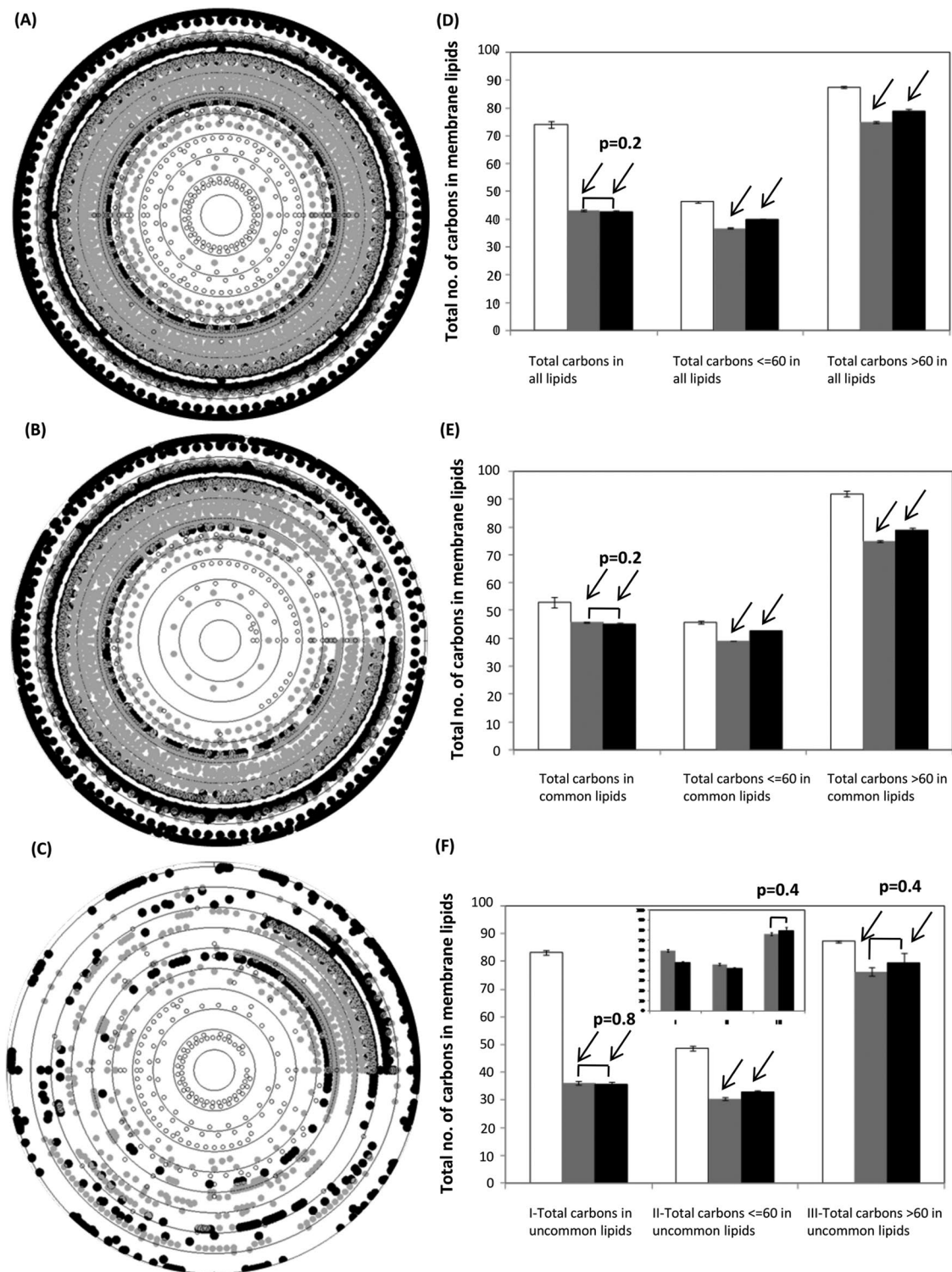


FIGURE 2: Variation in the total carbon content of membrane lipids of each domain. Total number of carbons in (A) all membrane lipid types (5382 lipid structures), (B) 10 common membrane lipid types (3911 lipid structures), and (C) 82 uncommon membrane lipid types (1471 lipid structures) were plotted against the time of evolution of their respective species. Ten concentric circles represent the evolutionary timescale; the innermost circle represents 4.5 Ga, and the outermost circle represents the most recent time (0 Ga). (D–F) Mean \pm SE for the total number of carbons, carbons less than and equal to 60 and carbons greater than 60 in all membrane lipids for all membrane lipids, common membrane lipids, and uncommon membrane lipids, respectively, of archaea, eubacteria, and eukaryotes. Inset in F shows uncommon lipids after excluding 11 more common lipid compositions between eubacteria and eukaryotes. For D–F, three independent t tests ($\alpha = 0.05$) were applied to three data sets for archaea, eubacteria, and eukaryotes; only p values > 0.05 are labeled on the respective bars. Empty circles and bars represent archaeal lipids, gray circles and bars represent eubacteria, and black circles and bars represent eukaryotes. Black arrows on the bars represent the presence of Simpson's paradox in the respective bars.

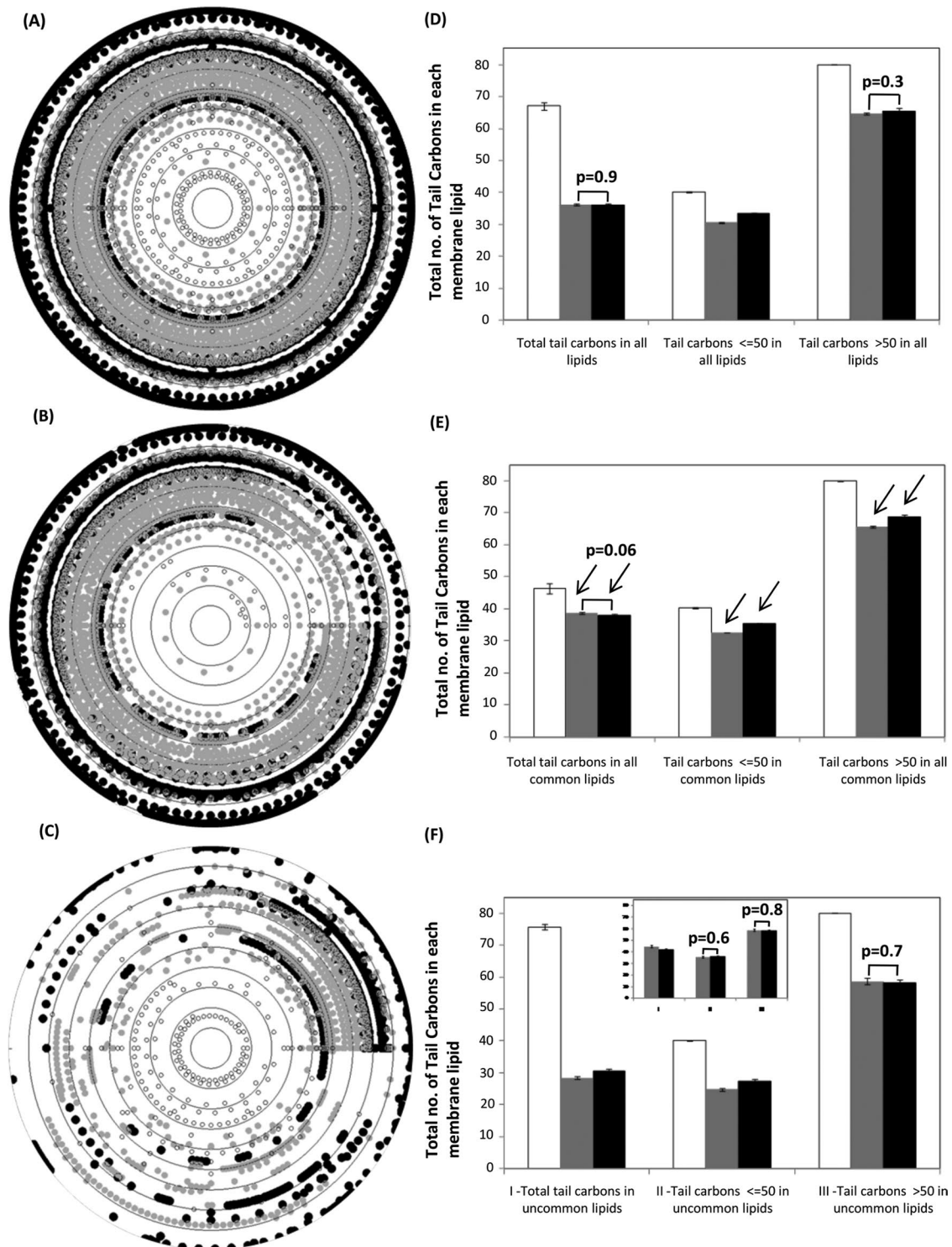


FIGURE 3: Variation in the total carbon content of tails of membrane lipids of each domain. Total number of tail carbons in (A) all membrane lipid types (5382 lipid structures), (B) 10 common membrane lipid types (3911 lipid structures), and (C) 82 uncommon membrane lipid types (1471 lipid structures) were plotted against the time of evolution of their respective species. Concentric circles represent evolutionary time, with the innermost circle representing 4.5 Ga and the outermost circle represents the most recent time (0 Ga). (D–F) Mean \pm SE for the total number of tail carbons, tail carbons less than and equal to 50, and tail carbons greater than 50 in all membrane lipids for all membrane lipids, common membrane lipids, and uncommon membrane lipids, respectively, of archaea, eubacteria, and eukaryotes. Inset in F shows uncommon lipids after excluding 11 more common lipid compositions between eubacteria and eukaryotes. For D–F, three independent t tests ($\alpha = 0.05$) were applied for three data sets of archaea, eubacteria, and eukaryotes; only *p* values > 0.05 are labeled on the respective bars. Empty circles and bars represent archaeal lipids, gray circles and bars represent eubacteria, and black circles and bars represent eukaryotes. Black arrows on the bars represent the presence of Simpson's paradox in the respective bars.

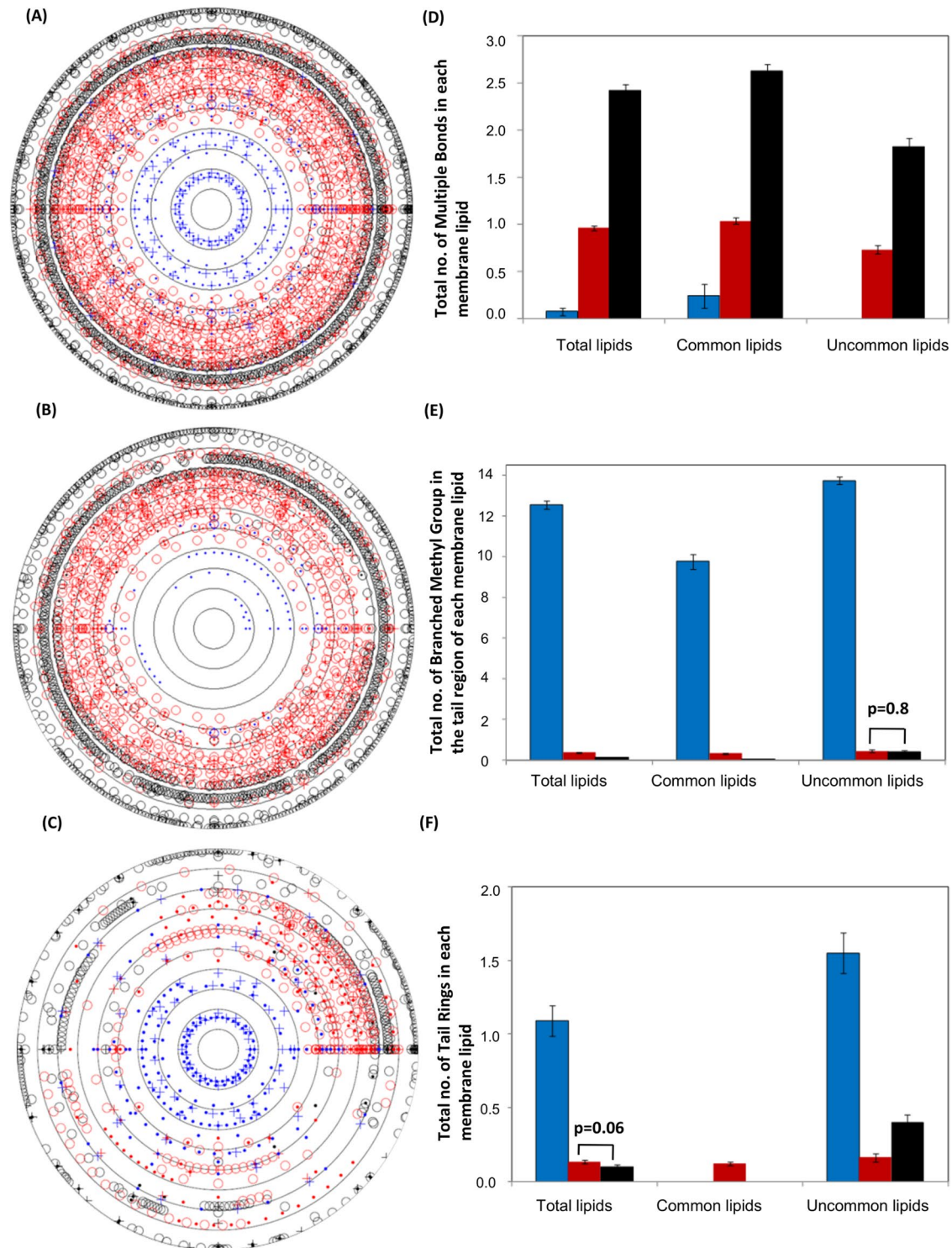


FIGURE 4: Structural variation in membrane lipids of each domain. Total number of multiple bonds, branched methyl groups, and rings in the tail region of (A) all membrane lipid types (5382 lipid structures), (B) 10 common membrane lipid types (3911 lipid structures), and (C) 82 uncommon membrane lipid types (1471 lipid structures) were plotted against the time of evolution of their respective species. Concentric circles represent evolutionary time, with the innermost circle representing 4.5 Ga and the outermost circle representing the most recent time (0 Ga). (D–F) Mean \pm SE for the total number of multiple bonds, branched methyl groups, and rings, respectively, in the tail regions of all membrane lipids, all common membrane lipids, and all uncommon membrane lipids of archaea, eubacteria, and eukaryotes. For D–F, three independent *t* tests ($\alpha = 0.05$) were applied for three data sets of archaea, eubacteria, and eukaryotes; only *p* values > 0.05 are labeled on the respective bars. Empty circles represent multiple bonds, dots represent branching, and a plus sign (+) was used for tail branching in archaeal, eubacterial, and eukaryotic lipids. Blue symbols and bars represent archaeal lipids, red symbols and bars represent eubacterial lipids, and black symbols and bars represent eukaryotes.

nonelectrolytes (Balleza *et al.*, 2014), and promotion of tubulation/curvature formation in ether lipids (Kozlov *et al.*, 2014), we found tail branching in abundance in archaea but decreased in eubacteria and completely vanished from the common membrane lipid types of eukaryotes (Figure 4, A–C and E). In contrast with the above, tail rings completely disappeared not only from common membrane lipid types of eukaryotes but also from archaeal common lipids (Figure 4, A–C and F). Tail rings were found only in the archaeal uncommon lipids and were completely replaced by a new class of lipids having fused ring-like structures (known as *sterols*) in eubacteria and eukaryotes. All these structural changes in membrane lipids clearly show increased disorder and mobility of lipids, which in turn enhances the fluidity of the plasma membrane of these domains in favor of changing environmental conditions.

The key observations and inferences from the above are as follows: 1) the presence of multiple bonds only in archaeal common membrane lipids, which was completely absent in their uncommon lipids; 2) a lower amount of branching in common membrane lipids than the uncommon membrane lipids of archaea; 3) complete disappearance of tail rings only in common membrane lipid types of archaea and eukaryotes; 4) complete disappearance of branching only in common membrane lipids of eukaryotes; and 5) these favorable changes, based on comparative analysis of common and uncommon lipids, were more pronounced in the chemical structures of common lipids as compared with uncommon lipids of all three domains (Figure 4, D–F). These observations clearly indicate that only common membrane lipids of archaea were altered in their hydrophobic tail regions in response to the changing environment and thus were preferably selected and exchanged and survived as integral structural features of plasma membranes over time in all domains. Thus structural adaptability of common membrane lipids allowed their successful exchange and/or incorporation into different domains. On the other hand, each domain also evolved with a few more membrane lipid types specific to their own niche (uncommon membrane lipids). Alternatively, transfer of genetic components of common membrane lipids' biosynthetic machineries might have allowed their codevelopment into plasma membrane assemblies. Literature studies revealed that the homologues of key enzymes for fatty acid synthesis and degradation were found to be universally present in all three domains (Lombard *et al.*, 2012a). Apart from this, enzymes that synthesize phospholipids (which is one out of the three common membrane lipid classes) and facilitate attachment of glycerol phosphate to their tail and polar head regions were also found to be commonly conserved in all three domains (Lombard *et al.*, 2012b). Common membrane lipids of the three domains of life strongly indicated the possible exchanges of the membrane lipids or transfer of genetic components of common lipids biosynthetic machineries among domains and/or their common origins from an ancestral population that gave rise to the statistical paradox (Simpson's paradox) that is observed for these lipids only. To our knowledge, this is the first evidence, based on the "third front" (i.e., lipids) and achieved through compositional analysis of lipidomics data, on the symbiotic origins of eukaryotic cells and demonstrating a larger role for eubacteria in formation of plasma membranes.

Proposed model to explain the origin of eukaryotes based on the membrane lipidomics of three domains

Remarkably, we were able to capture significant statistical signatures (from Simpson's paradox and *t* test analysis) from the plasma membranes of the three domains of life, despite their being highly dynamic and exposed to diverse environmental conditions. The data

clearly support the idea that the eukaryotes inherit the majority of their plasma membrane lipids from eubacteria. Further, to investigate the origin of mitochondria/chloroplasts (i.e., energy-producing organelles) in eukaryotes, we collected available ~400 lipids from the membranes of mitochondria and chloroplasts of 13 eukaryotes. We observed that membranes of these organelles consist of 24 different chemical compositions that include 10 compositions common to all three domains, eight compositions uniquely common to eubacterial plasma membranes, and six unique to the mitochondria (but also present in the plasma membranes of eukaryotes; see Figure 1). Here it is important to note that these eight commonly shared lipid compositions between mitochondrial membranes and plasma membranes of eubacteria were found to be different from 11 uniquely shared lipids between plasma membranes of eubacteria and eukaryotes. However, 10 membrane lipid types common to all three domains were found to be exactly similar to the common lipids of their plasma membranes. Further, quantitative comparisons revealed that the mean value of the total number of carbon atoms and tail carbon atoms was found to be higher in the membrane lipids of eukaryotic organellar membranes than in their plasma membranes (Figures 2, 3, and 5). Student's *t* tests show that the variation found in the total carbon and tail carbon content of membrane lipids of mitochondria/chloroplasts of eukaryotes and plasma membranes of eubacteria is not statistically significant and thus they might have derived these membranes from the same ancestral population (see Figure 5, A–F). Clearly, since the extrinsic environment for organellar membranes is different from the plasma membranes of eukaryotes and eubacteria, similarities found in the membrane lipids of mitochondria/chloroplasts of eukaryotes and plasma membranes of eubacteria are not as prominent or evident compared with their plasma membranes.

In view of above observations, we propose a hypothetical model to illustrate the evolution of mitochondriate eukaryotes based on membrane lipidomics (Figure 6). It is important to note that the composition of the plasma membrane of the cenancestor remains a topic of debate in evolutionary studies. There are recent reports supporting the idea that the plasma membrane of the cenancestor was composed of two stereotypes; glycerol-1-phosphate (G-1-P) and glycerol-3-phosphate (G-3-P), that is, ether- and ester-bond membrane lipids (Lombard *et al.*, 2012b). Lombard *et al.* (2012b) proposed that the diversification of these two stereotypic lipids might have led to the emergence of two different lineages, that is, archaea and eubacteria. On the other hand, our results indicate that the diversification of these two stereotypic lipids (included in common membrane lipid types) from the cenancestral plasma membrane led to the emergence of prearchaeal and prebacterial cell lineages. Then these common membrane lipids of prearchaeal cells might have adapted several structural changes and acquired unique membrane lipid types to survive in their niche environments, thus evolving into archaeal cells. Similarly, pre-eubacterial cells might have acquired several changes in the structure of their common lipids and synthesized new lipids to adapt to the changing environment. During this process of evolution of pre-eubacterial cells to eubacteria, some pre-eubacterial cells might have engulfed eubacteria, giving rise to primitive mitochondriate eukaryotes. Hence, while these endosymbiont eubacteria gave rise to the formation of mitochondria, host pre-eubacterial cells shared the nucleic acids and protein pools present in their cytoplasm with the eukaryotes. As these pre-eubacterial cells evolved from the cenancestral cells, eukaryotes shared significant similarities in their information system (i.e., components of replication, transcription, and translation) with archaea (Gribaldo *et al.*, 2010). Later these primitive eukaryotes

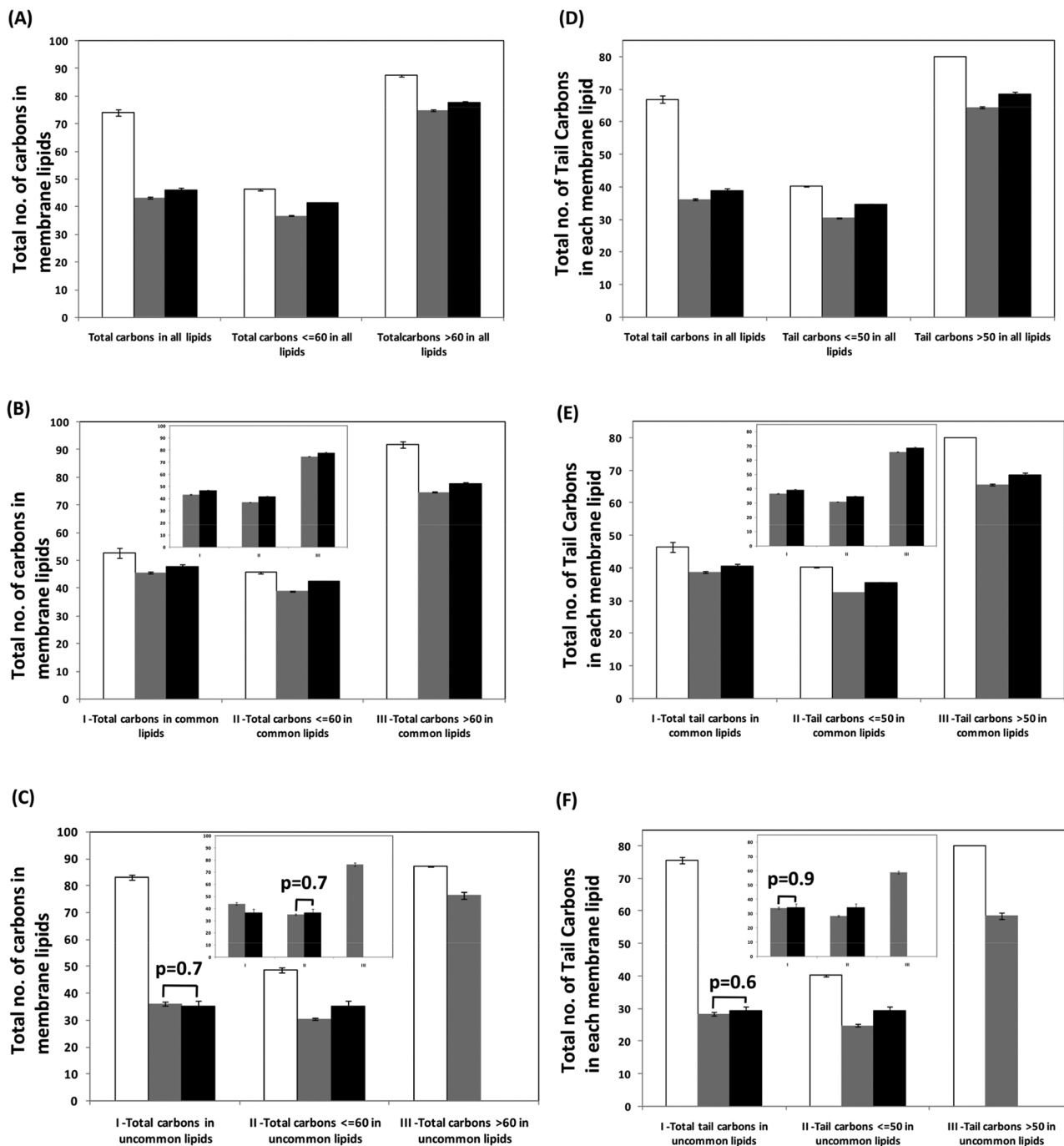


FIGURE 5: Variation in the hydrophobic content of membrane lipids of organellar membranes of eukaryotes and plasma membranes of archaea and eubacteria. (A–C) Mean \pm SE for the total number of carbons, carbons less than and equal to 60, and carbons greater than 60 for all membrane lipids (2343 lipid structures), common membrane lipids (1775 lipid structures) and uncommon membrane lipids (568 lipid structures), respectively, of archaea, eubacteria, and eukaryotes. (D–F) Mean \pm SE for the total number of tail carbons, tail carbons less than and equal to 50, and tail carbons greater than 50 in all membrane lipids for all membrane lipids, common membrane lipids, and uncommon membrane lipids, respectively, of archaea, eubacteria, and eukaryotes. Insets in B and E show common lipids after including eight more common lipid compositions shared by eubacteria and eukaryotes. Insets in C and F show uncommon lipids after excluding eight more common lipid compositions shared by eubacteria and eukaryotes. For A–F, three independent *t* tests ($\alpha = 0.05$) were applied for three data sets of archaea, eubacteria, and eukaryotes; only *p* values > 0.05 are labeled on the respective bars. Empty bars represent archaeal lipids, gray bars represent eubacteria, and black bars represent eukaryotes.

might have undergone several structural changes in their shared membrane lipids and also might have synthesized numerous unique membrane lipid types. Further, generation of an endomembrane

system from a host plasma membrane might have led to the formation of membrane-bound compartments and a nucleus around the host genetic material.

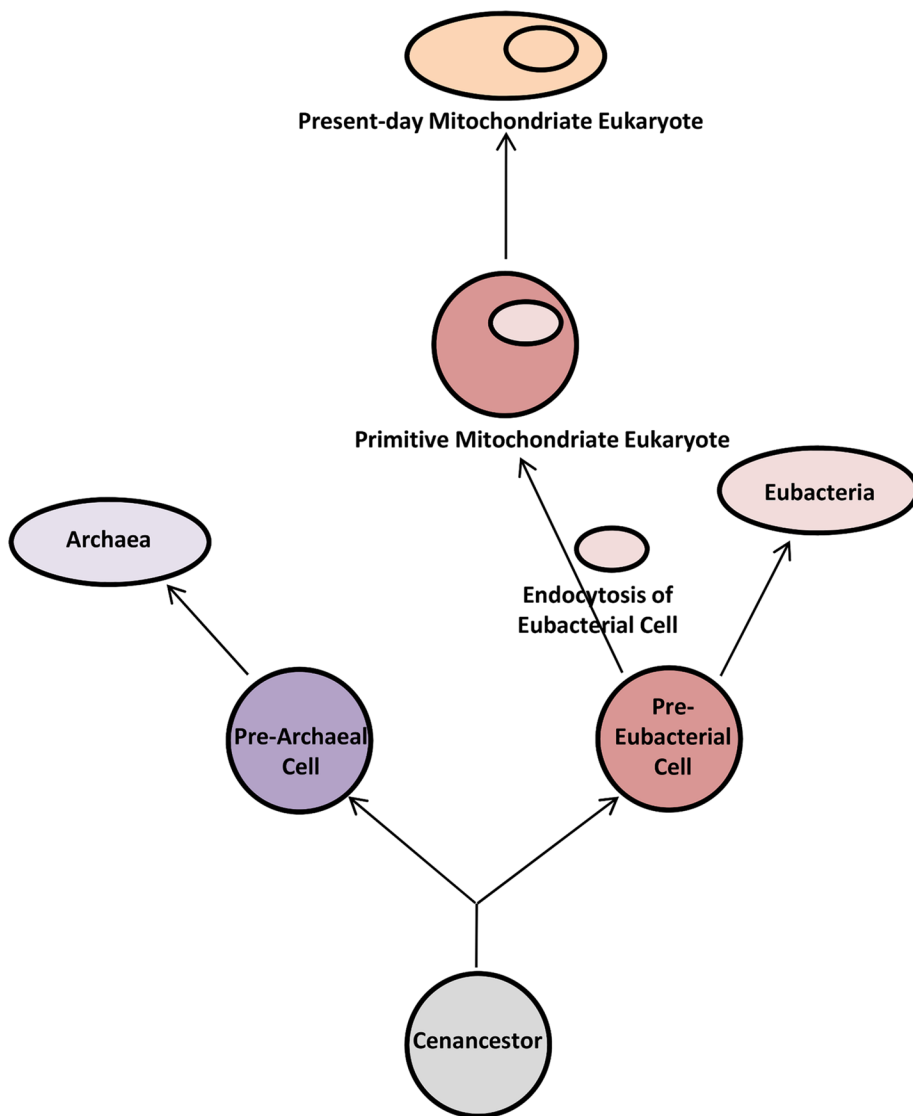


FIGURE 6: Proposed model for the origin of eukaryotes based on the membrane lipidomics of three domains. The plasma membrane of the cenancestor has been proposed to be made up of two stereotypes of G-1-P and G-3-P, ether- and ester-bond membrane lipids (Lombard *et al.*, 2012b). Divergence of prearchaeal and pre-eubacterial lineages is thought to have emerged after encapsulation of one of these two specific stereotypic lipids (including common membrane lipid types with different stereochemistry) in their respective membranes. Later the plasma membrane of prearchaeal cells might have incorporated several structural changes in their common membrane types and acquired unique membrane lipids favorable to their environmental conditions and specific to their archaeal lineage. Pre-eubacterial cell membranes may also have evolved to form eubacterial membrane-like lipids. During this process, some pre-eubacterial cells might have endocytosed eubacterial cells to give rise to the primitive mitochondriate eukaryotes.

Conclusions

Comparative compositional studies of two key components of central dogma (i.e., nucleic acids and proteins) have provided major insights into the origins of eukaryotes. However, to date, few reports have been available on the evolution of eukaryotic membranes based on the study of components of plasma membranes. Biological membranes are essential primary building blocks for the formation of stable and dynamic living systems. Polymerization of lipid amphipaths not only possibly provided the very first compartment for the origin of life on earth but also changed with the evolution of various species in response to their changing environmental conditions. Thus we

also, for the first time, carried out a comprehensive chemical composition-based analysis on the “third front” (i.e., lipids) of plasma membranes of 273 species collected from the three domains of life. Qualitative comparative analysis of classified lipids resulted in three classes of lipids common to all three domains. Further, extensive quantitative analysis of the chemical structures of these common lipids revealed the remarkable Simpson’s paradox. Common membrane lipids of eubacteria and eukaryotes showed reverse average trends in their carbon content as compared with their two mutually exclusive subsets. Interestingly, this paradoxical condition was not observed in the case of archaea and for uncommon membrane lipids. This revealed the significant presence of (hidden) interaction(s) and/or association(s) in membrane lipids of eubacteria and eukaryotes. This indicates the symbiotic origin of eukaryotic cells, which utilized eubacterial lipids as a major source of their plasma membranes. We provide further compositional evidence that the common membrane lipids having chemical structures contributing to key membrane functionalities (e.g., permeability, lipid mobility, membrane fluidity) were selected in response to changing environments (passed to the next generation or exchanged) in the three domains individually. Hence the possible exchange of common membrane lipids or the transfer of genetic components of their biosynthetic machineries and/or their emergence from a common ancestral population were found to be the source of this hidden interaction(s) and gave rise to the statistical anomaly discussed earlier. We also observed significant similarities between the mitochondrial membrane lipids of eukaryotes and plasma membrane lipids of eubacteria. Thus, through compositional analysis of “third front” (i.e., lipids) data, we provide the first evidence on the symbiotic origins of eukaryotic cells majorly involving eubacterial lipids in the formation of their plasma membranes and mitochondrial membranes. Further, as the assembly of lipid amphipaths resulting in membrane structures is largely govern by their molecular shapes (Tanford, 1973; Baumgart *et al.*, 2003; Bansal and Mittal 2013), future work on shape/structural classification of lipids in light of the compositional analysis here may shed more light on development of cell structures.

MATERIALS AND METHODS

Membrane lipid compositions of a total of 273 species (74 archaea, 144 eubacteria, and 55 eukaryotes) were collected from the literature along with the species’ estimated time of evolution (Hedges *et al.*, 2006; see Supplemental Table S1). A total of 5714 membrane lipids were collected from all of these species. Of these membrane lipids, complete chemical structures of 5382 membrane lipids were

available and were therefore recorded. Of these 5382 membrane lipids, molar fractions of 2343 lipids were available and were therefore collected from the literature. On the basis of their chemical composition, these membrane lipids were first classified into 23 different classes with a total of 92 different chemical compositions (see Supplemental Figure S2). Classification of lipids was done by analyzing the chemical composition of each membrane lipid for 1) the type and 2) the number of sugars, sulfates, and/or amines of the core, tail, and other additional molecules present in the head and/or tail region of these lipids (we ignored the difference of ether and ester linkages in archaea and other domains). Archaea consisted of a total of 11 different classes with 34 different membrane lipid compositions, eubacteria contained a total of 14 classes with 45 compositions, and eukaryotes consisted of a total of 16 classes with 47 membrane lipid compositions. We also collected 408 chemical structures of membrane lipids of mitochondria and chloroplasts from 13 eukaryotes (out of the 55 eukaryotes). Organellar membranes consist of eight classes with 24 membrane lipid compositions.

ACKNOWLEDGMENTS

S.B. is grateful to the Council of Scientific and Industrial Research, Government of India, for research fellowship support.

REFERENCES

- Balleza D, Garcia-Arribas AB, Sot J, Ruiz-Mirazo K, Goni FM (2014). Ether- versus ester-linked phospholipid bilayers containing either linear or branched apolar chains. *Biophys J* 107, 1364–1374.
- Bansal S, Mittal A (2013). Extracting curvature preferences of lipids assembled in flat bilayers shows possible kinetic windows for genesis of bilayer asymmetry and domain formation in biological membranes. *J Membr Biol* 246, 557–570.
- Baum DA, Baum B (2014). An inside-out origin for the eukaryotic cell. *BMC Biology* 12, 76.
- Baumgart T, Hess ST, Webb WW (2003). Imaging coexisting fluid domains in biomembrane models coupling curvature and line tension. *Nature* 425, 821–824.
- Chen IA, Walde P (2010). From self-assembled vesicles to protocells. *Cold Spring Harb Perspect Biol* 2, a002170.
- Daiyasu H, Kuma K, Yokoi T, Morii H, Koga Y, Toh H (2005). A study of archaeal enzymes involved in polar lipid synthesis linking amino acid sequence information, genomic contexts and lipid composition. *Archaea* 1, 399–410.
- Forterre P (2011). A new fusion hypothesis for the origin of Eukarya: better than previous ones, but probably also wrong. *Res Microbiol* 162, 77–91.
- Fuerst JA, Webb RI (1991). Membrane-bounded nucleoid in the eubacterium *Gemmata obscuriglobus*. *Proc Natl Acad Sci USA* 88, 8184–8188.
- Goldfine H (2010). The appearance, disappearance and reappearance of plasmalogens in evolution. *Prog Lipid Res* 49, 493–498.
- Gribaldo S, Poole AM, Daubin V, Forterre P, Brochier-Armanet C (2010). The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? *Nature Rev Microbiol* 8, 743–752.
- Hartman H, Fedorov A (2002). The origin of the eukaryotic cell: a genomic investigation. *Proc Natl Acad Sci USA* 99, 1420–1425.
- Hedges SB, Dudley J, Kumar S (2006). TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* 22, 2971–2972.
- Kanhere A, Bansal M (2005). Structural properties of promoters: similarities and differences between prokaryotes and eukaryotes. *Nucleic Acids Res* 33, 3165–3175.
- Knight RD, Freeland SJ, Landweber LF (2001). A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. *Genome Biol* 2, RESEARCH0010.
- Koga Y (2012). Thermal adaptation of the archaeal and bacterial lipid membranes. *Archaea* 2012, 789652.
- Kozlov MM, Campelo F, Liska N, Chernomordik LV, Marrink SJ, McMahon HT (2014). Mechanisms shaping cell membranes. *Curr Opin Chem Biol* 29, 53–60.
- Lindley DV, Novick MR (1981). The role of exchangeability in inference. *Ann Stat* 9, 45–58.
- Lindmark DG, Muller M (1973). Hydrogenosome, a cytoplasmic organelle of the anaerobic flagellate *Trichomonas foetus*, and its role in pyruvate metabolism. *J Biol Chem* 248, 7724–7728.
- Lombard J, Lopez-Garcia P, Moreira D (2012a). An ACP-independent fatty acid synthesis pathway in archaea: implications for the origin of phospholipids. *Mol Biol Evol* 29, 3261–3265.
- Lombard J, Lopez-Garcia P, Moreira D (2012b). The early evolution of lipid membranes and the three domains of life. *Nature Rev Microbiol* 10, 507–515.
- Lombard J, Moreira D (2011). Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. *Mol Biol Evol* 28, 87–99.
- Margulis L, Chapman M, Guerrero R, Hall J (2006). The last eukaryotic common ancestor (LECA): acquisition of cytoskeletal motility from aerotolerant spirochetes in the Proterozoic Eon. *Proc Natl Acad Sci USA* 103, 13080–13085.
- Martin W (2005). Archaeobacteria (Archaea) and the origin of the eukaryotic nucleus. *Curr Opin Microbiol* 8, 630–637.
- Martin W, Koonin EV (2006). Introns and the origin of nucleus-cytosol compartmentalization. *Nature* 440, 41–45.
- Martin W, Muller M (1998). The hydrogen hypothesis for the first eukaryote. *Nature* 392, 37–41.
- Mittal A, Acharya C (2013). Protein folding: is it simply surface to volume minimization? *J Biomol Struct Dyn* 31, 953–955.
- Mittal A, Jayaram B (2011). Backbones of folded proteins reveal novel invariant amino acid neighborhoods. *J Biomol Struct Dyn* 28, 443–454.
- Mittal A, Jayaram B (2012). A possible molecular metric for biological evolvability. *J Biosci* 37, 573–577.
- Mittal A, Jayaram B, Shenoy S, Bawa TS (2010). A stoichiometry driven universal spatial organization of backbones of folded proteins: are there Chagaff's rules for protein folding? *J Biomol Struct Dyn* 28, 133–142.
- Monnard PA, Deamer DW (2002). Membrane self-assembly processes: steps toward the first cellular life. *Anat Rec* 268, 196–207.
- Moreira D, Lopez-Garcia P (1998). Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J Mol Evol* 47, 517–530.
- Mulkidjanian AY, Galperin MY, Makarova KS, Wolf YI, Koonin EV (2008). Evolutionary primacy of sodium bioenergetics. *Biol Direct* 3, 13.
- Nussinov R (1984). Doublet frequencies in evolutionary distinct groups. *Nucleic Acids Res* 12, 1749–1763.
- Oger PM, Cario A (2013). Adaptation of the membrane in Archaea. *Biophys Chem* 183, 42–56.
- Pagano RE, Huang L (1975). Interaction of phospholipid vesicles with cultured mammalian cells. II. Studies of mechanism. *J Cell Biol* 67, 49–60.
- Pereto J, Lopez-Garcia P, Moreira D (2004). Ancestral lipid biosynthesis and early membrane evolution. *Trends Biochem Sci* 29, 469–477.
- Rasmussen S, Bedau MA, Chen L, Deamer D, Krakauer DC, Packard NH, Stadler PF (2009). *Protocells: Bridging Nonliving and Living Matter*, Cambridge, MA: MIT Press.
- Reed CF (1968). Phospholipid exchange between plasma and erythrocytes in man and the dog. *J Clin Invest* 47, 749–760.
- Rivera MC, Jain R, Moore JE, Lake JA (1998). Genomic evidence for two functionally distinct gene classes. *Proc Natl Acad Sci USA* 95, 6239–6244.
- Rucker G, Schumacher M (2008). Simpson's paradox visualized: the example of the rosiglitazone meta-analysis. *BMC Med Res Methodol* 8, 34.
- Simpson EH (1951). Interpretation of interaction in contingency tables. *J R Stat Soc Ser B Stat Methodol* 13, 238–241.
- Tanford C (1973). *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*, New York: Wiley-Interscience.
- Tu YK, West R, Ellison GT, Gilthorpe MS (2005). Why evidence for the fetal origins of adult disease might be a statistical artifact: the "reversal paradox" for the relation between birth weight and blood pressure in later life. *Am J Epidemiol* 161, 27–32.
- Vesteg M, Krajcovic J (2008). Origin of eukaryotic cells as a symbiosis of parasitic alpha-proteobacteria in the periplasm of two-membrane-bounded sexual pre-karyotes. *Commun Integr Biol* 1, 104–113.
- Wagner CH (1982). Simpson's paradox in real life. *J Amer Stats* 36, 46–48.
- Woese CR (2000). Interpreting the universal phylogenetic tree. *Proc Natl Acad Sci USA* 97, 8392–8396.