Molecular paleontology and complexity in the last eukaryotic common ancestor

V. Lila Koumandou¹, Bill Wickstead², Michael L. Ginger³, Mark van der Giezen⁴, Joel B. Dacks⁵, and Mark C. Field⁶

¹Biomedical Research Foundation, Academy of Athens, Soranou Efesiou 4, Athens 115 27, Greece, ²School of Biology, University of Nottingham, Nottingham, NG7 2UH, UK, ³Faculty of Health and Medicine, Lancaster University, Lancaster, LA1 4YG, UK, ⁴Biocatalysis Centre, Biosciences, College of Life & Environmental Sciences, University of Exeter, Stocker Road, Exeter, EX4 4QD, UK, ⁵Department of Cell Biology, University of Alberta, Edmonton, Alberta, T6G 2H7, Canada, and ⁶Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QT, UK

Abstract

Eukaryogenesis, the origin of the eukaryotic cell, represents one of the fundamental evolutionary transitions in the history of life on earth. This event, which is estimated to have occurred over one billion years ago, remains rather poorly understood. While some well-validated examples of fossil microbial eukaryotes for this time frame have been described, these can provide only basic morphology and the molecular machinery present in these organisms has remained unknown. Complete and partial genomic information has begun to fill this gap, and is being used to trace proteins and cellular traits to their roots and to provide unprecedented levels of resolution of structures, metabolic pathways and capabilities of organisms at these earliest points within the eukaryotic lineage. This is essentially allowing a molecular paleontology. What has emerged from these studies is spectacular cellular complexity prior to expansion of the eukaryotic lineages. Multiple reconstructed cellular systems indicate a very sophisticated biology, which by implication arose following the initial eukaryogenesis event but prior to eukaryotic radiation and provides a challenge in terms of explaining how these early eukaryotes arose and in understanding how they lived. Here, we provide brief overviews of several cellular systems and the major emerging conclusions, together with predictions for subsequent directions in evolution leading to extant taxa. We also consider what these reconstructions suggest about the life styles and capabilities of these earliest eukaryotes and the period of evolution between the radiation of eukaryotes and the eukaryogenesis event itself.

Introduction

The origin of eukaryotes is considered, with justification, as one of the major evolutionary transitions for life on Earth (Maynard Smith & Szathmáry, 1995). It brought with it sophisticated intracellular compartmentalization, separation of translation and transcription (permitting increased complexity in gene expression (Martin & Koonin, 2006)), superior capabilities for genetic reassortment and, potentially, alterations to evolvability (Poole et al., 2003). Each advance individually is potential justification for the emergence of the eukaryotes, and so the coalescence of these mechanistic and cellular advances provides a compelling cohort of selective advantages along the pathway of prokaryote to eukaryote transition. However, an apparent burst of innovation is inevitably an oversimplification, due in large part to a paucity of data from the earliest periods of the eukaryotic period, suggested to have occurred up to two billion years ago (Butterfield et al., 1990; Chernikova et al., 2011; Parfrey et al., 2011; Peterson & Butterfield, 2005). Few unambiguous fossils (in the sense of having confidently assigned taxonomy) are documented from the sediments laid down in the Proterozoic era of the Precambrian (Cavalier-Smith, 2006), and the information content of many specimens is limited in terms of describing what cellular systems these organisms possessed and the molecules that facilitated construction of these systems. However, several specimens do suggest potentially complex life styles, with obvious molecular ramifications (Butterfield et al., 1988, 1990; Knoll et al., 2006). A fuller understanding of these earliest events requires a molecular paleontology, i.e. reconstructing ancient gene complements. With the improved availability of genome sequence data from diverse taxa, and the improved computational ability to analyses those data, the era of molecular paleontology is now upon us.

Eukaryotes are frequently considered a sister lineage to the archaea, on account of sharing multiple structures and features and as originally revealed from rRNA sequencing, a hypothesis known as the three primary domain model (Woese & Fox, 1977) (Figure 1A). The closer relationships
Figure 1. Unresolved questions in the early evolution of eukaryotes. (A) How many domains of life are there? The traditional view of the tree of life places all three of the major domains, i.e. bacteria, archaea and eukaryota as monophyletic (top tree). This implies that the eukaryotes branched from the Archaea as a separate and independent lineage, with a stepwise topology, i.e. bacteria emerged first, from which archaea arose and then finally the eukaryota. An alternate hypothesis, however, suggests that the eukaryotes are essentially a branch within the archaea, and that archaea and eukaryota are, therefore, monophyletic (lower tree), allowing for coevolution of an archaeal/eukaryote precursor prior to speciation. Attempts to reconstruct the topology of the archaeal/eukaryote differentiation have so far been inconclusive, with both models receiving support, although this support is far from unequivocal (Gribaldo et al., 2010). OoL; origin of life, FECA; first eukaryotic common ancestor. (B) When did the mitochondrial symbiont arrive? Many proposals for the origin of the first eukaryotic cell have been offered. Of the models that have garnered support, two simple common schema can be extracted, and a third more complex possibility. Left; A fusion event occurred between an archaea and an \( \alpha \)-proteobacterium (the source of the mitochondrial genome and functions). Central; Significant development of endogenously-derived membranous and other structures by the Archaeal ancestor arose prior to endosymbiosis of the \( \alpha \)-proteobacterium. The latter mechanism may have involved more complex fusion events, for example including methanogens or an endosymbiotic origin for the nucleus (discussed in Embley & Martin, 2006), while the metabolic capabilities of ancestral cells are essentially ignored. Right; A third possibility is that the mitochondrion arose comparatively late, after much of the complexity of the protoeukaryote had evolved, and following fusion between a bacterium (khaki lozenge) and an archaean (right scheme). While multiple endosymbiont events are considered by many as highly unlikely, the point at which the mitochondrion came on board, as well as when a true nucleus arose remain controversial and unresolved. However, most models agree that the LECA possessed mitochondria, substantial internal differentiation and a well-defined nucleus. Probing beyond LECA is critical for understanding these earliest events. Gray lozenge; Archaeal ancestor, purple lozenge; \( \alpha \)-proteobacteria (the mitochondrion is drawn in purple in the LECA), blue lozenge; protonuclear endosymbiont (the nucleus is drawn in blue in the LECA). (C) Eukaryotic tree of life with examples of sequenced organisms from currently recognized supergroups. The curved dotted-line indicates the separation of lineages included in the unikonts and the bikonts. SAR + CCTH: stramenopiles (heterokonts), alveolates, and Rhizaria plus cryptomonads, centrohelids, telonemids and haptophytes (see color version of this figure at www.informahealthcare.com/bmg).
between eukaryotic and archaean transcription and translation systems, in particular, speaks strongly to this intimate relationship. However, what is unclear is precisely how the archaea and eukaryotes are related; while the concept of sister lineages (i.e. arising as separate but related branches) is supported by some molecular phylogenies, a second topology, the two primary domain model that suggests that eukaryotes emerged within the archaea and are hence monophyletic with them, is supported by other analyses. Confidently differentiating these models remains intractable (Forterre, 2011; Gribaldo et al., 2010). If eukaryotes arose after the archaea, as suggested by the two primary domain model, this predicts that phylogenetic reconstructions would reflect independent differentiation of multiple archaean lineages, only one of which gave rise to eukaryotes. Most significantly, the two-domain model may imply that potentially more sophisticated and non-universal archaean features were present in the ancestral lineage of the last eukaryotic common ancestor (LECA), as the eukaryotes represent only a single taxon within the archaea. Clearly, the order of events has important implications for the genetic repertoire that LECA would have inherited.

Further, there remains debate concerning the precise mechanisms behind eukaryogenesis, i.e. the events leading up to the first eukaryotic common ancestor (FECA) and the subsequent evolution of the LECA (Figures 1B and 2). For convenience here we are defining FECA as the first ancestor in a lineage that lead to LECA, and which is presumed to have acquired one or a few eukaryotic-specific cellular features, while LECA is the first eukaryote, minimally defined by having a mitochondrion and a nuclear envelope. However, as the discussion below will demonstrate, it is now clear that LECA most probably possessed most of the sophistication of modern eukaryotes, i.e. multiple intracellular compartments, a cytoskeleton and also complex metabolic and gene regulatory mechanisms.

Taking a very simplistic view, and ignoring many excellent but less well-supported models, the proposed trajectories for...
eukaryotic cell origins can be grouped into two major categories; a fusion first model where an endosymbiosis event delivering the mitochondrion came extremely early, or a fusion later model where endosymbiosis occurred after development of several intracellular structures (discussed in Embley & Martin, 2006; O’Malley, 2010). Arguments for the mitochondrion first model are based primarily on energetic considerations (e.g. Lane & Martin, 2010), while the second model places emphasis on a requirement for phagocytosis-like mechanisms to be present to facilitate endosymbiont acquisition. There is also some speculation about a prior fusion event between archaea and bacteria, to produce a FECA which then took on the mitochondrial endosymbiont, reflected in a third model (Figure 1B, rightmost).

As strong as the connection with the archaea is, there is also evidence uniting eukaryotes and bacteria. This comes not only from the use of ester-linked phospholipids, but also from the eubacterial origins of many eukaryotic metabolic pathways. While this contribution could be dismissed as either derived from mitochondrial endosymbiosis or as a result of horizontal gene transfer, it may represent evidence for a fusion between archaea and bacteria to produce the FECA, which subsequently phagocytosed the primordial mitochondrion (Forterre, 2011), essentially a synthesis of the two previous models. Again, excellent arguments for all models have been made, but definitive discriminatory evidence remains lacking. There is also the very real need to appreciate that these are all singular events and hence stochastic influences are likely. Where all models agree is that, at a very early point in their evolution, eukaryotes possessed a fully functional mitochondrion, a nucleus and additional intracellular structures (Dacks & Doolittle, 2001; Embley & Martin, 2006; Roger, 1999). What remains is determining the origins of these features, which ones FECA possessed and how the rest arose post-FECA, and significantly, how long these processes took (Chernikova et al., 2011).

Much insight has been made possible recently through analysis of the genomes of extant organisms (Figure 1C). The greatly increased availability of molecular data, combined with improved phylogenetic tools and reconstructions of eukaryotic phylogeny, permits piecemeal reconstruction of likely LECA biology: the methodology has been discussed elsewhere (e.g. Koumandou & Field, 2011). A major surprise is that when molecular-level reconstructions of major cellular systems or protein families have been attempted, frequently these predict that LECA possessed a remarkably modern configuration, extending from cytoskeletal systems, through endomembrane and protein processing, metabolic capabilities and on to encompass meiosis, organization of the genome into linear chromosomes with telomeric ends and RNA processing (Figure 3). Perhaps even more remarkable is the realisation that LECA may have been, in multiple aspects, more

Figure 3. A generalized model for LECA with emphasis on the major systems proposed as present and discussed here. It is now clear that the LECA was both a flagellate and capable of movement by actin-based pseudopodia and possessed a sophisticated cytoskeleton, including large families of kinesin and dynein motors (not shown for clarity). It possessed a complex and likely very flexible, metabolism and a fully functional mitochondrion. Endomembrane compartments would have been essentially indistinguishable from modern cells, and included the endoplasmic reticulum, the Golgi complex, endosomes, autophagosomes and others (many not shown for clarity). The LECA was also capable of both conventional endocytosis and phagocytosis. The nucleus was fully differentiated with nuclear pore complexes and a sophisticated system for organization and regulation of chromatin. A high energy burden is clearly implied by this architecture and required to construct and maintain these compartments and systems plus a differentiated cytoskeleton to coordinate location and function. Heterochromatin in some form could also support life-cycle and/or environmental cue-dependent coordinate gene expression. LECA also supported meiosis. Systems are exploded with examples of the complex aspects associated with a reconstructed LECA (see color version of this figure at www.informahealthcare.com/bmg).
sophisticated than a significant number of extant eukaryotes. Here we aim, in overview, to assemble some of the complexity inferred for the LECA and to consider what type of organism or organisms LECA would have been.

Systems

The nucleus and the nuclear envelope

The defining feature of eukaryotic cells, the nucleus, is responsible for packaging the genetic material and coordinating gene expression, amongst other roles (Martin & Koonin, 2006). An expansion in the physical size of eukaryotic genomes, with a greater frequency of noncoding DNA also likely necessitates more sophisticated organization and structural support. This is provided in large part by histones, which are universal amongst eukaryotes, and most of which are also present in archaea and bacteria, and therefore, likely arose pre-FECA (Kasinsky et al., 2001; Sandman & Reeve, 2005). Importantly, separation of transcription and translation also facilitated evolution of complex gene expression regulatory mechanisms, which may in part result from evolutionary exploitation of histone packaging systems, and allowed splicing and other RNA processing events to emerge (Martin & Koonin, 2006). We have little de novo information on mechanisms controlling histone assembly and function beyond basic chromatin regulatory processes and most are shared between animals, fungi and other lineages; where data exist these aspects appear near universal (e.g. Figueiredo et al., 2009). At this level the system is probably very highly conserved. Furthermore, while RNA and DNA polyadenylases, despite eukaryotic elaborations, are clear prokaryotic descendants, transcriptional mechanisms are remarkably variable, with polycistronic transcription versus one-gene one-promoter systems providing an example of a process where even a well-conserved gene cohort operates in a distinct manner between lineages, albeit still retaining significant mechanistic similarities (Daniels et al., 2010; Moore & Russell, 2012; Morton & Blumenthal, 2011).

Much understanding of the structural organization of the nucleus derives from metazoan lamin proteins. These 60 kDa coiled-coil proteins have multiple functions, interacting with the nuclear pore complex (NPC), organizing heterochromatin and positioning of chromosomes and also subtending interactions with the cytoskeleton via interactions with the LINC complexes that span the nuclear envelope (Starr & Fridolfsson, 2010). Further, lamins, coiled coil intermediate filament proteins, are essential for the structural integrity of the nucleus as well as other higher order organizational functions (Simon & Wilson, 2011). Until recently, lamins appeared specific to metazoas, with no evidence for a presence in any other lineage, including the related fungi. The recent discovery of a lamin-like protein in the amoeba Dictyostelium discoideum (Krüger et al., 2012) suggests lamins originated as early as the unikont root (Cavalier-Smith, 2003; Figure 1C). Lamins remain unikont specific at this time, despite evidence for heterochromatin and other lamina-requiring functions in the bikonts. Putative lamin analogs have been identified in Arabidopsis thaliana while a small family of coiled-coil LINC proteins are associated with the A. thaliana nucleus and nuclear periphery and appear to control nuclear size and chromosomal segregation (Dittmer et al., 2007), but these proteins also seem to be land plant specific. Recently NUP-1, a large coiled-coil protein was described in trypanosomes, which performs many of the roles ascribed to lamins, including maintaining nuclear structure and defining heterochromatin (DuBois et al., 2012). While similarities between NUP-1 and lamin functions are striking, in silico analysis has not identified NUP-1 orthologs outside of the kinetoplastida, or common ancestry with unikont lamins or plant LINC proteins. This may reflect both evolutionary distance and the low complexity amino acid composition of coiled-coil domains, but at present data are nondiscriminatory concerning the ultimate origins of lamins, LINC proteins and NUP-1 and their presence in the LECA. However, what is clear is that many taxa possess lamin functional equivalents, so it is rather likely that such functions were part of the LECA cellular physiology.

Integration of nuclear and other cellular functions requires bidirectional transport across the nuclear envelope. As translation is cytoplasmic this requires that all tRNA, rRNA and mRNAs be exported, while proteins required for DNA replication, transcription, transcriptional regulation, RNA processing and overall nuclear organization are imported. Import and export across the nuclear envelope is the function of the NPC, a huge structure that in Saccharomyces cerevisiae comprises ~30 different proteins at multiple copy numbers, with a total subunit tally exceeding 430 (Alber et al., 2007). Transport is mediated by several mechanisms, mainly via recognition of a cis-acting signal within the transported protein by a karyopherin (KAP). Transport is powered by a gradient of GDP- versus GTP-bound Ran, a small Ras-like GTPase (Grossman et al., 2012). Export of mRNA, in some systems at least, utilises non-KAP factors, including Mex67, and is Ran-independent (Oeffinger & Zenklusen, 2012). There are differences in molecular mechanisms of nucleocytoplasmic transport between yeast, plants and mammals, and specifically in how the Ran gradient is controlled; in metazoa, RanGAP activity (GAPs stimulate GTPase activity converting the GTPase from a GTP to GDP-bound form) is associated with the nuclear pore complex, but in A. thaliana the RanGAP is targeted to the nuclear envelope by nuclear envelope-embedded trans-membrane domain proteins (Meier et al., 2007; Xu et al., 2007), while in S. cerevisiae RanGAP is not targeted to the nuclear envelope at all. There are fourteen β-KAP subclasses, together with a single α-KAP that has undergone paralogous expansions in metazoa (Mason et al., 2009; O’Reilly et al., 2011). Some KAPs are highly specific, while others remain less well understood. Moreover, there is functional promiscuity, with distinct β-KAP families able to assume the roles of others when the primary β-KAP is deleted; the evolutionary basis for retention of such secondary functions is unclear, but may contribute to KAP cohort evolvability (O’Reilly et al., 2011). The fourteen basal β-KAP clades are represented in all eukaryotic supergroups indicating that the basic repertoire was present early in eukaryotic evolution, i.e. pre-LECA. KAP paralogs expanded in some lineages, while secondary loss is common. Most remarkably, with the exception of a plant-specific KAP clade, there has been little evolution of lineage-specific KAPs, and little evidence for evolution of
Endomembrane compartments and trafficking

Organelles of the endomembrane system include the entirety of the secretory/exocytic and endocytic pathways and the nuclear envelope, which is contiguous with the endoplasmic reticulum (Figure 3). Proteins destined for the surface or to be secreted, enter the system by co- or post-translational translocation across the ER, are folded by numerous chaperones and monitored by quality control mechanisms (Alberts et al., 2002). Most protein export proceeds via the Golgi complex, where molecules enter the cis-face, traverse several cisternae and are exported from the trans-most cisternae. Plasma membrane delivery is achieved via secretory vesicles budding from the trans-Golgi compartment. Endocytic pathways originate by invagination of the plasma membrane, fusion or maturation of the resulting vesicles into endosomes and subsequent sorting to one of several destinations. The major destinations are recycling to the cell surface, a common pathway for nutrient receptors for example, and which intersects post-Golgi exocytic routes, or delivery to terminal degradative endosomes, variously termed lysosomes, vacuoles or reservosomes. This latter pathway, mediated by ubiquitination, is important for turnover of surface proteins, signaling receptors and destruction of immune factors for example, as well as degradation of material for nutritional purposes.

In a general sense all of these pathways are well understood and were present in the LECA. There is excellent evidence for universal conservation of the ER translocation machinery, with origins in the prokaryotic SecY system (Jungnickel et al., 1994). The major chaperone subclasses are also likely ancient, encompassing categories involved in folding, disulphide bond rearrangement and quality control, including sensing mechanisms based on N-glycan glycosylation and retro-translocation of terminally mal-folded proteins (Field et al., 2010). It is unclear if any of the proteins mediating these latter functions have direct prokaryotic ancestors, although several are part of the huge and universal HSP family. The Golgi complex was almost definitely present in the LECA (Klute et al., 2011), but interestingly has taken multiple evolutionary trajectories. At its most extreme the canonical Golgi stacked cisternae morphology has been lost, and in several cases there is no microscopic evidence for the organelle. This occurred in multiple lineages, indicating convergent evolution (Mowbrey & Dacks, 2009). However, the basic functions of protein targeting and N-glycan processing are retained in such lineages, which raises the issue of a potentially cryptic Golgi complex or repurposing of other compartments that assume these functions (Dacks et al., 2003).

Further, the Golgi complex demonstrates quite extreme morphological variability. While many microbial eukaryotes possess a single Golgi complex, as seen in trypanosomes, others including prominent organisms such as apicomplexans (T. gondii aside) and ciliates have a reduced Golgi complex, i.e. a single cisterna (reviewed in Mowbrey & Dacks, 2009). Golgi organelles have also expanded in some lineages, such as the ribbon like, interconnected Golgi stacks in mammalian tissue culture cells, extensive and mobile stacks in A. italiana (Staehelin & Kang, 2008) and beautifully expanded Golgi bodies of parabasalid taxa with large numbers of both stacks...
and cisternae per stack (e.g., Brugerolle, 2004). Even closely related yeast species display variance in Golgi complex morphology (Suda & Nakano, 2011). While it is near certain that LECA possessed a stacked Golgi apparatus, it is unclear what the precise LECA configuration would have been, and the molecular steps that facilitate extreme morphological plasticity within a central organelle of the eukaryotic cell are unknown.

Endocytosis, at least as far as we presently understand it, presents a more complex evolutionary story. The basic system is once more a feature of LECA, with the major endosomal coat protein, clathrin, being (so far) universal. In several lineages, including Trypanosomatids, Apicomplexa and plants, clathrin-mediated endocytosis may represent the sole mechanism for endocytosis (Field et al., 2007). By contrast, in metazoan cells multiple modes of endocytosis co-exist, with caveolin-, Cdc42-, RhoA- and flotillin-mediated pathways being frequently viewed as restricted to the Opisthokonta. Some of these pathways, for example Cdc42 and RhoA-dependent pathways, likely await more detailed functional dissection, as the presence of these multifunctional GTPases is, of itself, insufficient to define an endocytic route as mechanistically distinct from others (Sandvig et al., 2011). The evolutionary history of flotillin-mediated endocytosis, associated with both clathrin-dependent and -independent endocytic pathways, is less clear (Otto & Nichols, 2011). Flotillin orthologs are present in bacteria and archaea, and in Bacillus associate with detergent-resistant membranes (DRMs); the precise function is unknown or even if this represents true endocytosis (Lopez & Kolter, 2010). These findings are particularly significant as they suggest that bacteria, as well as eukaryotes, segregate their plasma membranes based on protein-lipid physicochemical properties, and these similar biochemical underpinnings, i.e. isoprenoid-derived metabolites and lipid-binding flotillins, operate (Lopez & Kolter, 2010). Additionally, signaling complex proteins are enriched in DRMs in both prokaryotes and eukaryotes, suggesting conserved function. The presence of GPI-anchored proteins in DRMs is apparently a result of membrane physicochemical properties, and these similar biochemical underpinnings, i.e. isoprenoid-derived metabolites and lipid-binding flotillins, operate (Lopez & Kolter, 2010). Additionally, signaling complex proteins are enriched in DRMs in both prokaryotes and eukaryotes, suggesting conserved function. The presence of GPI-anchored proteins in DRMs is apparently a result of association with pre-existing aspects of membrane physiology, and potentially involvement of flotillin could likewise represent recruitment to endocytic functions. Finally, while the distribution of flotillin in eukaryotes is broad, there is frequent secondary loss. The functions of these additional endocytic pathways are unclear, and while there may be other undiscovered taxon-specific pathways, it is firmly established that the LECA possessed a clathrin-based endocytic system, and possibly also a flotillin-mediated mode.

Later endocytic processes include the sorting of proteins by the ESCRT system, intimately involved in the generation of multi-vesicular bodies and pre-lysosomal compartments. ESCRTs were originally identified as class E vacuolar sorting mutants (vps) in S. cerevisiae, but their importance has become more extensive with subsequent analysis (Field & Dacks, 2009); their principal function in the endocytic system is the recognition of ubiquitylated endocytic cargo as well as the invagination of membrane in late endosomes to create multi-vesicular bodies (MVBs), a function which appears to be intrinsic to the snf7 and vps4 ATPase subunits (Hanson et al., 2008). The entire system is comprised of five subcomplexes, which together contain ~25 distinct proteins. The core of this system is near universal, and indicates that the LECA possessed ESCRT machinery and hence likely the ability to sort ubiquitylated endocytic cargo and form MVBs, although the absence of one subcomplex (ESCRT 0) outside of animals and fungi, and responsible for cargo recognition suggests that some lineage-specific mechanisms must be present (Leung et al., 2008). Broad representation is unsurprising as several ESCRT components are present in the archaea, and significantly the snf7/vps4 orthologs play a role in the curving of membrane during cytokinesis, where they are recruited by CdvA to form helical fibers (Dobro et al., 2013; Samson et al., 2011). This more ancient role for a subset of ESCRT proteins seems to have been maintained in eukaryotes, where ESCRT factors are recruited late during membrane scission and appear to participate in the final steps of cytokinesis, together with a number of eukaryotic-specific proteins of the endosomal system (reviewed in Chen et al., 2012).

Control of vesicular traffic and definition of compartments is highly dynamic and the result of collaborations between large cohorts of proteins. Much of this complexity is the result of expansions of several gene families (Dacks & Field, 2007). Major players include SM proteins, tethers, Rab and ARF family GTPases, SNAREs, adaptins and coat proteins, the evolution of which have now been investigated in considerable detail. The presence of paralogs at the core of these systems helps to explain two important features; how new compartments arise and why there is plasticity within the endomembrane system when considering the diverse configurations present in divergent taxa. We suggested a model for the evolution of new endomembrane compartments, which we term organellar paralogy, and which suggests a simple mechanism for the integration of new paralogs into pre-existing complexes and subsequent neo-functionalisation (Dacks & Field, 2007; Elias et al., 2012; Figure 4).

GTPases play an important role in vesicle trafficking, in both the formation and fusion of transport intermediates. The Rab subfamily is the prime mediator of compartmental identity and controls the fusion events between transport intermediates and organelles. This family is perhaps the premier example of a large paralogous family with well-detailed evolutionary histories. Recent data indicate a substantial family of over twenty Rab proteins in the LECA (Diekmann et al., 2011; Elias et al., 2012). Importantly, reconstruction of Rab evolution indicates emergence of the major organelle-specific subfamily members in the LECA, and also pre-LECA emergence of primordial endocytic and exocytic Rabs (Elias et al., 2012), suggesting a stepwise, ongoing increase in complexity. This view is further supported by the continual emergence and elimination of Rab5 isoforms, which contribute to early endocytosis (Dacks et al., 2008; Pereira-Leal, 2008), indicating that evolution of intracellular compartments is continuing in modern lineages. However, there appears to be a limit to the level of specialization possible, and which may be due to energetic constraints or more simply no need for more than about three distinct routes. Further, there is increased Rab GTPase family complexity within animals and fungi, but also evidence for the emergence of lineage-specific Rab proteins in all supergroups (Diekmann et al., 2011; Elias et al., 2012).
Finally, the LECA Rab complement is greater than some well-studied extant organisms, many of which, however, are highly derived taxa, suggesting secondary loss as a major driver for sculpting the endomembrane system. Recent analysis also suggests a large complement of Rab GTP-activating proteins and GTP-effector proteins in LECA (Gabernet-Castello et al., 2013).

The ARF GTPases provide a contrast to Rab evolution and offer one of a restricted number of examples of primordial simplicity in the LECA (Li et al., 2004). While there is good evidence for ARF participation in membrane transport, the contribution to the LECA was likely modest. ARF expansions in multiple taxa suggest a single ARF LECA ancestor, and that elaboration of the family is supergroup specific (Berriman et al., 2005). The basis for this curious evolutionary history is unclear at present, and the drivers that propelled expansion of the ARF families are less obvious than for Rabs, despite the clear widespread influence and requirement for a substantial ARF family. Interestingly, a recent analysis of Arf-GAPs shows much greater complexity in the LECA, with at least six ancient subfamilies (Schlacht et al., 2013), suggesting that perhaps GAPs provided an early source for functional Arf diversity.

The other prominent example of simplicity in the LECA is that the multiple Qa-SNAREs involved in anterograde transport in the early and late endosomes respectively in animals, fungi and plants appear to have evolved from a single primordial endosomal SNARE (Dacks et al., 2008).

Nonetheless, while analysis is less extensive than for Rabs, a large SNARE cohort was also present in the LECA. Both comparative genomics (Yoshizawa et al., 2006) and phylogenetics (Dacks & Doolittle, 2002, 2004; Vedovato et al., 2009) have reconstructed not only the Qa, Qb, Qc and R families of SNAREs present in the LECA, but several major organelle-specific subfamilies as well.

Heterotetrameric adaptin (AP) complexes act as cargo receptors, and only four AP complexes were known until recently (Dacks et al., 2008; Robinson, 2004). Recently a fifth, highly divergent AP-5, operating in late endosomal transport, was identified and was probably present in the LECA (Hirst et al., 2011). While the AP family is less extensive than the Rab/SNARE families, it appears more stable over time. Most taxa possess at least AP-1, 2, 3 and 4, with a significant level of secondary loss of AP-5 from many lineages, for example, the trypanosomatids. However, examples of losses of other AP complexes have also been described, including AP-2, 3 and 4 (Berriman et al., 2005; Field et al., 2007; Manna et al., 2013; Nevin & Dacks, 2009), suggesting sculpting of this feature of the endomembrane system as well. Furthermore, detection of a divergent AP complex provides a salient lesson on the ability of search methods to call the limits of protein families: it is essential to maintain an open mind as to where such limits may lie.

SM proteins and tether complexes are involved in control of vesicle fusion and interact intimately with the Rab and SNARE proteins. While the total number of gene products...
involved is quite large, their evolution appears to be comparatively stable. As far as we are aware, there are four broadly distributed SM proteins, likely present in the LECA. The tether system is more complex, as these factors comprise several complexes of varying size, and in many taxa several subunits are lost, but most lineages have at least a representative of all examined tethering complexes, consistent with a complement of at least seven complexes in the LECA (Koumandou et al., 2007). Limited sequence similarity is apparent between the tether complex subunits, but structural biology has identified a similar “CATCHR” fold in subunits from multiple tether complexes, suggesting the presence of paralogous subunits (Bröcker et al., 2010; Spang, 2012). Extreme examples of specific expansions of tether subunits are known and include Exo70, an exocyst subunit, in A. thaliana, where there are over 20 paralogs (Chong et al., 2010). At least some of the expansion appears to be due to tissue-specific expression (Li et al., 2010), but the absence of such numbers of Exo70 paralogs from other complex multicellular organism lineages, e.g. metazoa, suggests that there is a more sophisticated driver at work.

A unification between the large cohorts of proteins involved in cytoplasmic membrane transport and nucleocytoplasmic transport hinges on the presence of a highly conserved architecture within these proteins (Devos et al., 2004). Specifically, clathrin, β-COP, adaptins, Sec13/31 of COPII and several subunits of the intraflagellar transport system share the same β-propeller and α-solenoid secondary structure present in many NUPs (deGrasse et al., 2009; Devos et al., 2004; van Dam et al., 2013). Further there are suggestions that similar architectures are also present in the HOPS/Corvet and SEA complexes which have roles in endocytosis, although formal solution of the structures themselves remains to be achieved (Dokudovskaya et al., 2011). As all of these protein families were fully established before the LECA, this also provides a model by which a primitive membrane deforming complex could have given rise to the numerous systems present in the LECA cell, essentially through simple paralogous expansion during the transition period between the FECA and LECA.

**Prokaryotic origins of eukaryotic trafficking systems**

Although intracellular vesicle trafficking is a hallmark of eukaryotic cells, many of the components have ancestral prokaryotic orthologs. Most were identified by comparing 3D structures as sequence identity between prokaryotic and eukaryotic orthologs is frequently insignificant, and many examples have only been identified recently due to increased genomic data from a variety of bacteria and archaea. For example, prokaryotic V4R proteins, which currently have no clear function, have low sequence, but significant structural similarity to the Bet3 subunit of TRAPPI (Podar et al., 2008), a tethering complex component involved in attachment of TRAPPI to Golgi membranes (Kim et al., 2005). Similarly, prokaryotic members of a family of protein cargo receptors possibly involved in vesicle formation and protein trafficking in eukaryotes have recently been identified through PSI-BLAST, HHMer, and secondary structure predictions (Saudek, 2012), and the secreted MPT63 protein of *Mycobacterium tuberculosis* has structural similarity to adaptins (Goulding et al., 2002). However, at present these links are tentative. Several prokaryotic trafficking-related factors have been studied in considerable detail, and here the evidence for common descent is compelling. For example, bacterial dynamin-like proteins have been studied in the cyanobacterium *Nostoc punctiforme* (BDLP) (Low & Lowe, 2006; Low et al., 2009) and the gram-positive bacterium *Bacillus subtilis* (DynA) (Burmann et al., 2011), and homologs are found in many bacterial lineages as well as in certain archaea (Methanomicrobia) (Bramkamp, 2012). Structural and functional studies suggest a role in cytokinesis and/or membrane fission, similar to eukaryotic dynamin. Rab and Arf GTPases are central to vesicle transport, and Ras homologs are present in several bacteria and archaea, suggesting a possible prokaryotic origin for the Rab and Arf GTPases central to vesicle transport (Dong et al., 2007). The bacterial Ras-like GTPase MglA, along with its cognate GAP, MglB, are required for sporulation and motility in the gram-negative soil bacterium *Mycobacterium xanthus* (Hartzell, 1997), and also in the regulation of cell polarity, i.e. the localization of proteins to the leading or lagging cell pole of motile cells (Bulyha et al., 2011; Leonardy et al., 2010; Zhang et al., 2010). The polar localization of motility proteins by MglA may also involve the actin-like protein MreB (Maurolo et al., 2010). MglB does have homologs in eukaryotes but not to known eukaryotic GAPs; rather, it contains a dynein light chain domain and may define a novel GAP family (Wanschers et al., 2009). MglA and MglB orthologs are present in many phylogenetically distant bacteria and archaea (Koonin & Aravind, 2000), suggesting that regulation of polarity by a Ras-like G-protein and a GAP is possibly a general prokaryotic feature.

Ubiquitination is a key sorting signal for protein sorting and degradation in eukaryotes. A functionally similar system, PUPylation, targets PUPylated proteins to the proteasome for degradation in *Mycobacterium*. Similar to ubiquitin, Pup is post-translationally transferred to proteins on lysine residues, but the enzymes involved are fewer than in eukaryotes and do not exhibit significant sequence or structural similarity to the eukaryotic ubiquitin-ligase system (Burns & Darwin, 2010; Pearce et al., 2008). A ubiquitin-like proteasome system has also been described in the halophilic archaeon *Haloferax volcanii*, and is shared with various archaeal species (Humbard et al., 2010). In addition, the genome of the archaea *Candidatus Caldiarchaeum subterraneum* harbors an operon-like gene cluster encoding homologs of eukaryote-type E1 and E2 ubiquitin ligases, as well as a small Zn-finger protein containing a RING finger motif that, in eukaryotes, mediates the ubiquitin ligase activity of RING-type E3s. This suggests the presence of an unprecedented eukaryote-type ubiquitin ligase system in archaea. HGT from eukaryotes is considered unlikely, given that the individual components acquired by HGT would need to have reorganized to form a gene cluster; however, such an arrangement may facilitate coordinate expression with significant selective advantage (Nunoura et al., 2011).

In the Crenarchaeae, where FtsZ and MreB are absent, cell division is mediated by the Cdv complex, with several components orthologous with eukaryotic genes, including the
Vps2/Snf7 subunits of the ESCRT-III system and Vps4 (Bernander & Ettema, 2010; Lindas et al., 2008; Makarova et al., 2010; Samson et al., 2008). Interestingly, an archaeal-specific factor recruits Cdv to the membrane (Samson et al., 2011). Cdv forms ring structures between segregating nucleoids, which constrict during cell division. In eukaryotes, ESCRT-III-derived curved filaments are involved in vesicle formation during endosomal protein sorting; Vps2/Snf7 mediates membrane bending itself while Vps4, an ATPase, is responsible for disassembly. The Cdv system may be a preferential cytokinesis mechanism in some thaumarchaeal species that possess both FtsZ and Cdv, and which may be an archaeal lineage most closely related to eukaryotes (Busiek & Margolin, 2011; Peive et al., 2011). The bacterial proteins PspA and Vipp1 also have homology to Vps2, and Vipp1 is thought to function in membrane stabilisation or vesicle traffic for thylakoid biogenesis within chloroplasts of cyanobacteria (Vothknecht et al., 2012).

The Cdv system is also involved in outer membrane vesicle (OMV) formation in archaea (Ellen et al., 2010). Various archaeal and bacterial species lacking known ESCRT-like factors can release proteins packaged into small (10–300 nm in diameter) membrane vesicles that emerge from the cell surface. OMVs are implicated in a variety of processes, including release of bacterial toxins and quorum-sensing factors (Ellen et al., 2010; Ellis & Kuehn, 2010; Lee et al., 2009). In fact, the quorum-sensing hydrophobic molecules packaged into OMVs directly affect OMV biogenesis (Masshburn & Whiteley, 2005; Masshburn-Warren et al., 2008). Furthermore, OMVs have a distinct lipopolysaccharide composition from the outer membrane, which could influence sorting of specific proteins into these structures (Haurat et al., 2011). While no prokaryotic vesicle coats have been formally demonstrated yet, some evidence suggests that vesicle budding may simply be a physicochemical process in cells synthesising excess membrane for their surface and/or that lack the factors to support a certain cell shape (Bendezu & de Boer, 2008; Erickson & Osawa, 2010; Leaver et al., 2009). Imbalances in protein associations between outer/inner membrane and the peptidoglycan wall (Deatherage et al., 2009; Moon et al., 2012), as well as hydrophobic molecules preferentially intercalating into the outer leaflet of the membrane bilayer (Schertzer & Whiteley, 2012) also induce membrane curvature and OMV budding. Furthermore, two bacterial proteins, SpoVM and DivLVA, preferentially associate with positively or negatively curved membranes, respectively, and apparently without the need for adaptors or other sorting signals (Shapiro et al., 2009). Lipid microdomains also affect protein localization in bacteria. For example, cardiolipin preferentially associates with negatively curved membranes and mediates polar positioning of the proteins MinD and ProP (Renner & Weibel, 2011; Romantsov et al., 2007), while sterol rich flotillin-containing microdomains in B. subtilis and other bacteria have already been mentioned (Lopez & Kolter, 2010).

Internal membranes are far from unique to eukaryotes and present in multiple bacterial and archaea lineages (reviewed in Fuerst & Sagulenko, 2012). Considerable interest in the Planctomycete bacteria and Gemmata obscuriglobus, in particular, has led to detailed examination of these bacteria, and how their internal membrane systems relate to eukaryotes (Fuerst, 2005; Santarella-Mellwig et al., 2013). The observation that G. obscuriglobus DNA is contained within a membrane-bounded structure has been used as evidence for a nucleus-like precursor in these organisms (Fuerst, 2005; Fuerst & Sagulenko, 2011), and evidence for an energy-dependent, endocytosis-like mechanism by a Planctomycete, has also been presented (Lonhienne et al., 2010). However, most recently it has been shown that the putative nucleoid is open and that the endomembrane system within G. obscuriglobus, while being highly complex, is similar to those found more widely in bacterial cells (Santarella-Mellwig et al., 2013). Structure prediction suggests the presence in Planctomycetes of proteins with the β/α architecture, a hallmark of the proteocamer superfamily (Devos et al., 2004; Santarella-Mellwig et al., 2010). These findings have caused considerable controversy, with suggestions for HGT being put forward as an explanation for the presence of β/α proteins in Planctomycetes (Devos, 2012; McInerney et al., 2011). However, the β and α architectures are incredibly common, and while the type of α-solenoid that features in proteocamers is a specific subfamily (Field et al., 2011), it is difficult to imagine that the β/α topology can be used alone as evidence for common ancestry between bacterial and eukaryotic proteins, rather than simple convergence. Forterre has argued for a model whereby the proteoeukaryote arose through fusion of a Planctomycete, or close relative, with a thaumarchaeon, which is potentially consistent with direct descent (Forterre, 2011), although definitive functions for the Planctomycete β/α proteins remain to be reported. Regardless of their status as eukaryotic precursors or not, these β/α proteins are of considerable interest to prokaryotic biology and the general understanding of membrane trafficking in a broader context.

In summary, while no unequivocal prokaryotic vesicle coat or SNARE-like proteins have been reported, there are candidate prokaryotic homologs for Rab-like GTPases, cytoskeletal components (see below), putative adaptors, tethering complex precursors, a subset of the ESCRT system and a ubiquitin-like-sorting system, as well as membrane microdomains characteristic of lipid rafts, as well as possible β/α architecture proteins. Together, this suggests a probable prokaryotic origin for at least some portions of the eukaryotic trafficking machinery, and it is tempting to hypothesize that the eukaryotes did not innovate all trafficking factors de novo, but instead repurposed pre-existing prokaryotic systems as vesicle budding and trafficking machinery within a cell with multiple membrane-bounded compartments. The prokaryotic ESCRT system and the actin/tubulin/dynamin homologs are likely involved in cell division (Field & Dacks, 2009), whereas prokaryotic lipid-rafts and GTPases mediate cell polarity. However, many of these connections must be treated with caution as the ability to detect common ancestry in silico falls at the edge of statistical significance, and we must be aware that precise mechanistic details are lacking for many of these examples, and which are an important step in confirming putative relationships. It may also be that various combinations of prokaryotic trafficking factors were exploited in lineages pre-LECA, with HGT and novel innovations such as coatamer proteins also involved, but that the LECA was so
successful that the vast majority of these other configurations failed to survive into the extant eukaryotic lineage.

Cytoskeletal filaments

The cytoskeletons of all extant eukaryotes are dominated by two filament-forming protein families; tubulin and actin. Both protein families were already present as multiple paralogs in LECA. Tubulin had already diversified into the two major microtubule paralogs (α and β), a nucleating paralog (γ), and two paralogs associated with the axoneme/basal body (δ and ε) (Vaughan et al., 2000), while actin had expanded to produce several actin-related protein (ARP) families, including paralogs associated with dynactin (ARP1), nucleation (ARP2 and ARP3), and four nuclear families (ARP4, 5, 6 and 8) (Sehbring et al., 2007; Schafer & Schroer, 1999). Three tubulin paralogs, α, β and γ, and actin, appear to be ubiquitous to all eukaryotes studied to date, while the other paralogs have been lost from at least some lineages. As expected from their primary functions, δ and ε-tubulin are absent from species lacking cilia/flagella, but curiously are also absent from Thalassiosira pseudonama, Caenorhabditis elegans, dipterans and lepidopterans (Hodges et al., 2010). Both ARP2 and ARP3 are lost from several protistan/algae lineages including some diatoms, the red alga Cyanodioschyzon merolae and the Apicomplexa (Wickstead & Gull, 2011a). This is rather surprising, since together ARP2 and ARP3 are part of an otherwise highly-conserved actin nucleator, which is essential in yeast and nematodes (Lees-Miller et al., 1992; Sawa et al., 2003; Schwob & Martin, 1992). ARP1 has also been lost several times: in ciliates, Theileria annulata, metamonads (e.g. Giardia lamblia) and trypanosomes (Wickstead & Gull, 2011a). In Trypanosoma and Leishmania, the absence of ARP1 is part of a general loss of all dynactin complex components, except cytoplasmic dynein 1 itself (Berriman et al., 2005). These findings indicate plasticity in the eukaryotic cytoskeleton that is not apparent from a consideration of any one lineage, and similarly to some Rab GTPases, genes that are essential in some taxa can be lost from others. Further, organisms in particular groups have reduced their dependence on specific aspects of the cytoskeleton while elaborating others; the excavae Giardia lamblia reduced its actin cytoskeleton to the extent that the actin gene is highly divergent and no families of ARPs or actin-based motors have been identified in the genome (Morrison et al., 2007).

Both tubulin and actin have prokaryotic ancestors with low sequence similarity but clear tertiary structural conservation (Bork et al., 1992; de Boer et al., 1992; Mukherjee et al., 1993; RayChaudhuri & Park, 1992). Tubulin is homologous to the prokaryotic proteins FtsZ, TubZ and RepX, the latter two encoded by bacterial plasmids. Due to divergence it is difficult to robustly determine which prokaryotic gene is the true eukaryotic tubulin ortholog. However, given widespread occurrence of FtsZ in both bacteria and archaea (although, notably not the Crenarchaeota), it is reasonable to assume that this is the nearest extant relative. Heterodimeric BtubA/B, found in some Prosthecobacter species, is much more similar to eukaryotic tubulins than other bacterial homologs (Jenkins et al., 2002; Vaughan et al., 2004). Lack of strong phylogenetic affinity for any extant tubulin families, coupled with an ability to fold in the absence of chaperones, has been argued to point to these proteins being representatives of an ancient tubulin ancestor (Pilhofer et al., 2011). However, BtubA/B are extremely limited in their distribution and show strong evidence of horizontal gene transfer (Pilhofer et al., 2007), making divergence of sequence following transfer of tubulin from a eukaryote the more likely scenario.

FtsZ is also found in some eukaryotes alongside tubulin. This eukaryotic FtsZ is plastid-derived and serves a similar role in the division of the chloroplast and/or mitochondrion that it once did in their free-living ancestors. FtsZ mediates prokaryotic cell division, and mitochondrial and plastid division in eukaryotes, by forming a dynamic ring between prospective daughter cells (or daughter organelles) before cytokinesis (see Wickstead & Gull, 2011a).

Actin is a member of a large superfamily of ATPases that includes prokaryotic MreB, FtsA, AlfA and ParM, but also Hsp70 chaperones and several classes of sugar/sugar alcohol kinases (Derman et al., 2009; Flaherty et al., 1991; Jockusch & Graumann, 2012). Actin and MreB have similar fold structures (Kabsch & Holmes, 1995), but until recently it was unclear which of the many prokaryotic ATPases was most closely related to eukaryotic actin/ARPs. This was resolved by the discovery of “crenactin” – an archaeal actin-like protein with a localization similar to bacterial MreB in bacteria, but which is monophyletic with eukaryotic actin (Ettema et al., 2011; Yutin et al., 2009). Interestingly, this actin ortholog is only present in Crenarchaeota and some basal archaeal lines. This, together with the distribution of FtsZ, suggests the prokaryotic ancestor of the FECA probably arose near the base of the archaea (see Wickstead & Gull, 2011a).

MreB filaments are involved in maintenance of cell shape, forming a helix below the cell membrane and influencing cell wall synthesis. Another prokaryotic actin homolog, ParM, along with the tubulin homolog TubZ, have roles in plasmid segregation. In contrast, eukaryotic DNA segregation is dependent on the tubulin-based cytoskeleton in all systems studied, whereas cytokinesis involves actin–myosin (Wickstead & Gull, 2011a). Surprisingly, recent data show that in yeast a form of rudimentary nuclear division can proceed in the absence of a microtubule-based spindle and that this auxiliary system is potentially actin-based (Castagnetti et al., 2010). This possibly reflects a persistence of an ancestral mechanism, although given the ubiquity of tubulin-based mitosis in eukaryotes and the absolute requirement of other systems on the presence of the spindle, this interpretation is rather unlikely.

Intermediate filaments (IF) are a distinct class of cytoskeletal elements and, unlike tubulin and actin, lack directionality and cytomotivity in their own right, and have no known dedicated motor proteins (see Fuchs & Weber, 1994). Vertebrates possess many IF protein classes, including keratins, vimentin and desmin, α-internexin and laminin. It is likely that lamins were the first IF proteins to evolve within the unikonts and from these all other IF proteins evolved (Erber et al., 1998; Weber et al., 1989). Interestingly, in the cnidarian Hydra spp., a lamin-related protein, lacking both a
nuclear-localization signal and farnesylation site, is present in the mechanosensory cilia of nematocysts (Hwang et al., 2008). This gene arose through duplication of the nuclear lamin gene and provides an example of the evolution of cytoplasmic IFs from laminas in a manner independent of that which occurred in animals more generally. No good candidates for prokaryotic ancestors of eukaryotic IF proteins have been identified. The ‘‘IF-like’’ protein crescentin in the bacterium Caulobacter crescentus is more likely to be an example of convergence than true homology (see Wickstead & Gull, 2011a), and the absence of detectable lamin homologs outside of the unikonts may indicate the absence of the IF class of cytoskeletal proteins from other arms of the eukaryotic lineage (Figure 1C).

Cytoskeletal motors

Eukaryotic cytoskeletal function is hugely augmented by the recruitment of motors, kinesins and dyneins, to the tubulin-based cytoskeleton, and myosins, to F-actin. None of the eukaryotic cytoskeletal motors have characterized prokaryotic homologs with a similar motor function. However, systems such as the AglQRS system for gliding in the bacterium Myxococcus xanthus, suggest that trafficking on the bacterial cytoskeleton analogous to that seen extensively in eukaryotes has evolved (Sun et al., 2011). Kinesin and myosin motors share a common structure and are distant relatives (Kull et al., 1998, 1996). It is likely that they do not share a single common motor ancestor that walked on both actin- and tubulin-based filaments, but evolved independently from the same superfam-ily of proteins (Leipe et al., 2002). The ancestral superfamily of P-loop NTPases also gave rise to many other eukaryotic families, including the Ras-superfamily GTPases and additional G protein families (Leipe et al., 2002).

In contrast to kinesin/myosin, dyneins belong to the large, diverse AAA+ superfamily. Each dynein heavy chain contains six AAA+ domains of ~220 residues which form a hexameric ring (Carter et al., 2011; Samsó et al., 1998). Most prokaryotic AAA+ proteins contain a single AAA+ domain, but many assemble into homomeric rings (Lupas & Martin, 2002). Dynein most likely evolved by duplication and subsequent divergence of a single AAA+ domain. Much of this divergence occurred before the LECA and prior to emergence of the major dynein classes. Due to their small size and degree of divergence, the evolutionary history of AAA+ domains is also extremely challenging, but there is some evidence that the closest prokaryotic family to dynein may be MoxR and its relatives (Iyer et al., 2004; Snider & Houry, 2006). MoxR family members have diverse roles in prokary-otes and no known motor function, but have instead chaperone-like properties, suggestive of a similar role for the original ancestor of dynein in the pre-eukaryotic cell.

The cytoskeletal motors comprise superfamilies containing several paralogous classes, many of which are ancient. At least eleven ancient kinesin paralogs were present in the LECA under any likely model of early eukaryotic branching – namely, Kinesin-1, 2, 3, 4/10, 5, 8, 9A, 9B, 13, 14 and 17 (Wickstead et al., 2010). This provides molecular evidence for several key aspects of LECA biology. It can be inferred that the LECA built a bidirectional spindle with antagonistic plus-end directed (Kinesin-5) and minus-end directed (Kinesin-14) motors, which was modulated by microtubule depolymerizing motors (Kinesin-8 and -13). Also, LECA trafficked membrane-bounded organelles within the cytoplasm using Kinesin-1 and -3 and built a cilium/flagellum containing a 9 + 2 axoneme (Kinesin-9 A) by intraflagellar transport (Kinesin-2). However, post LECA all of these families have experienced multiple losses, so that no family is now ubiquitous. In spite of this, no eukaryote entirely lacking kinesins has been discovered, although the apicomplexan T. annulata completes its life-cycle with only two kinesins, both depolymerizing motors (Kinesin-8 and -13) (Wickstead et al., 2010).

Myosin diversity may be even higher than that of kinesins, with descriptions of up to 35 classes (Foth et al., 2006; Odonitz & Kollmar, 2007). However, much of this diversity may result from difficulties in phylogenetic reconstruction, as seen in the large number of apparently lineage-specific myosin families, and more conservative estimates are similar to kinesins (Richards & Cavalier-Smith, 2005). Even with these considerations, at least three myosin families can be traced back to the LECA. Myosins are entirely absent from excavates (G. lamblia and Trichomonas vaginalis) and the red alga Cyanidioschyzon merolae (Richards & Cavalier-Smith, 2005).

Recent phylogenetic analyses of dynein heavy chains suggests nine major classes (Morris et al., 2006; Wickstead & Gull, 2007; Wilkes et al., 2008), which encompass seven classes built into the axoneme of motile cilia/flagella and two cytoplasmic classes. All nine dynein classes were present in the LECA, but multiple losses during eukaryote diversification are clear. Cytoplasmic dynein 2 is the retrograde motor for IFT and required for construction of the axoneme in almost all lineages (Briggs et al., 2004; Rosenbaum & Witman, 2002).

Unsurprisingly, loss of cilia/flagella is associated with loss of flagellar dyneins or cytoplasmic dynein 2 (Wickstead & Gull, 2007). Cytoplasmic dynein 1 has also been lost independently at least three times, and the amoeba E. histolytica, red alga C. merolae and all angiosperms lack dyneins entirely (Lawrence et al., 2001; Wickstead & Gull, 2011b).

These analyses provide both molecular evidence for the existence of key motor-related functions in LECA and show that a large proportion of motor family diversity had already arisen in this ancient lineage. The advent of motor proteins was likely critical for eukaryotic cellular compartmentalization and facilitated the increased cellular complexity between the FECA, with a prokaryote-like cytoskeleton, and the more sophisticated LECA.

The axoneme

In many lineages, the cytoskeleton is used to form flagella and/or cilia, constructed from a microtubule axoneme extending from the basal body. In spite of notable absences, in angiosperms and most fungi, flagella/cilia are widely distributed (Carvalho-Santos et al., 2011; Hodges et al., 2010). In nearly all organisms the axoneme and basal body retain their iconic nine-fold symmetry. The basal body of flagella/cilia is an identical structure to the centriole, which is embedded in the primary microtubule organising centre of metazoan cells, the centrosome (see Azimzadeh & Marshall, 2010). Given their distribution in extant eukaryotes, cilia/
flagella arose pre-LECA (Cavalier-Smith, 1978) and molecular data suggest that the LECA flagellum possessed both sensory and motility functions (Mitchell, 2007).

There are obvious cytoplasmic analogs for much of the axonemal machinery, including core microtubules, dynein motors and several IFT components, presenting a plausible route for evolution of the axoneme from cytoskeletal factors; an autogenous origin for the axoneme is now accepted (Mitchell, 2007; Pickett-Heaps, 1974). However, the precise pathway by which the flagellum formed is still unclear and three alternative hypotheses have been proposed. The “sensation-first” model suggests axoneme-like structures evolved from microtubule-based protrusions with an exclusively sensory function (Cavalier-Smith, 1978), “beat-first” models place motility as the most ancestral function, while “gliding-first” models propose that the original function was motility, but driven by gliding resulting from an IFT-like motor, rather than microtubule sliding (Mitchell, 2004, 2007). Since dynein and axonemal evolution are intimately linked, dynein phylogenies can distinguish between these alternate hypotheses if the tree can be accurately rooted, such that the order of dynein family emergence can be inferred. If the ancestral dynein is assumed as a cytoplasmic dynein 1 (Hartman & Smith, 2009; Wilkes et al., 2008), then analyses support the sequential appearance of IFT and then axonemal beating, consistent with sensation-first and gliding-first hypotheses. Gibbons suggested that the homodimeric nature of cytoplasmic dynein 1 represents a more “primitive” arrangement than some of the axonemal dyneins (Gibbons, 1995), but the subsequent discovery of simple axonemal single-headed dyneins refutes this. Moreover, proposing that the axoneme evolved from cytoplasmic components does not necessitate cytoplasmic dynein 1 as the progenitor dynein. In contrast, rooting of dynein phylogenies using the closest eukaryotic relative of dynein, midasin (Garbarino & Gibbons, 2002; Iyer et al., 2004), suggests that microtubule sliding was much closer to the origin of proto-cilia and evolved before the specialized IFT machinery (Wickstead & Gull, 2011b). This implies that the cilium may have evolved not from an immotile protrusion, but from a motile cytoplasmic microtubule bundle analogous to the axostyles of oxymonads (McIntosh, 1973; McIntosh et al., 1973). This bundle would have been assembled in an IFT-independent manner, as are the axonemes of several extant organisms (Briggs et al., 2004; Witman, 2003).

Cell division

There are two major modes of cell division in eukaryotes, mitosis and meiosis. The former is the mechanism that underpins somatic or non-reductive division, resulting in two daughter cells with similar DNA content, while meiosis is a sexual process resulting in production of gametes. Depending on the configuration of the genome, this proceeds either by reductive cell division from a diploid state prior to gamete production and fusion or mating between two haploid cells prior to the reductive divisions, although other modes have been described. The primary mechanisms of eukaryotic cell division require participation of the cytoskeleton, for construction of the spindle, nuclear division and cytokinesis itself. Prokaryotic origins for the cytokinesis machinery and the role of the cytoskeleton in nuclear events have been described above. Meiosis, which for most organisms is a facultative mode of division, requires the participation of a specialised set of gene products that are frequently only expressed during the meiotic process, and which include Spo11, Hop1, Dmc1 and others (Peacock et al., 2011; Ramesh et al., 2005, Schurko & Logsdon 2008). Meiosis is clearly very distinct from bacterial conjugation. The ability to reassort genomes during meiosis is a major step in evolution and can facilitate the sweep of new traits more rapidly through a population.

Early phylogenetic analysis suggested that sex was likely widespread and that the LECA was a facultative sexual organism, despite evidence for loss of such activity in some lineages (Dacks & Roger, 1999), and the inherent difficulty in observing such behaviour in most taxa (e.g. Peacock et al., 2011). Importantly, a facultative sex mode ensures that the major cost to meiotic division, i.e. disruption of potentially successful genotypes, is only paid when the environmental conditions change. Comparative genomics further demonstrated the presence of meiotic genes in most lineages (Malik et al., 2007a; Ramesh et al., 2005; Schurko & Logsdon 2008), although significantly, even in organisms where meiosis clearly occurs, the entire meiotic gene cohort may not be present, or in taxa where a substantial cohort are present, the precise mechanism may be distinct. For example, Giardia, considered by some to be descended from an early branching eukaryotic lineage, is able to exchange chromosomes during specific stages in the life cycle without a true meiosis (Carpenter et al., 2012; Ramesh et al., 2005).

There is some evidence for the evolution of meiosis genes prior to the LECA. Spo11, a critical topoisomerase essential for meiosis is derived from archael topoisomerase VI (Bin3), and which is also retained by eukaryotes (Malik et al., 2007b). Based on the conservation of Spo11 paralogs in extant taxa, there were likely three Spo11 genes present in the LECA. Only Spo11-1 and Spo11-2 paralogs are meiosis specific, and significantly at least one of these is retained by all major lineages, but with evidence for frequent secondary losses. Only higher plants retain all three Spo11 paralogs, while Spo11-3 (Top6A) and Bin3 (Top6B) were lost from the unikont common ancestor. Hence the conclusion that the LECA possessed meiotic capability is well supported, with clear implications for its life style.

Mitochondrial origins and LECA

Unlike the majority of intracellular compartments, the mitochondrion has a non-endogenous origin, arising from once free-living bacteria, specifically an α-proteobacterium-like organism. The non-endogenous origin of mitochondria was first postulated by Portier & Wallin nearly a century ago (reviewed in Martin, 2007). The presence of DNA (Nass & Nass, 1963), semi-autonomous replication (Mitchell & Mitchell, 1952) and a supposed role in oxygen removal allowed others to postulate an endosymbiotic mitochondrial origin (Sagan, 1967). Subsequent accumulation of mainly molecular data provided clear evidence for a prokaryotic
origin but the timing remains contentious with, in principle, two conflicting hypotheses: the phagotrophy and syntrophy models (O’Malley, 2010; Figures 1 and 2). The phagotrophy model argues that the protoeukaryote required a cytoskeleton to facilitate endocytosis, after which an α-proteobacterium was engulfed (Sagan, 1967; Whatley et al., 1979). Based on metabolic and energetics arguments, the syntrophy model postulates that the origin of eukaryotes and establishment of the mitochondrion was one event (Lane & Martin, 2010; Martin & Müller, 1998).

Initial ultrastructural studies of various eukaryotes believed to be primitive, including Entamoeba, Giardia and Trichomonas, suggested the absence of typical cristate mitochondria, leading to proposal of the Archezoa as a primitively amitochondriate eukaryote group, whose extant representatives are these amitochondrial taxa (Cavalier-Smith, 1987). This implied that LECA could have lacked a mitochondrion. However, subsequent unambiguous identification of diagnostic mitochondrial features in each of these organisms plus the discovery of highly derived mitochondrial organelles in all other studied Archezoan taxa, resulted in the hypothesis being rejected. Mitochondrial variants go by various names as mitosomes, hydrogenosomes, mitochondrion relicts or mitochondrial-like organelles (Müller et al., 2012; van der Giezen, 2009). Discovery and assignment of these organelles is strong evidence against any eukaryotic mitochondrial lineage, with the consequence that, if not one and the same event, eukaryogenesis and the origin of mitochondria were chronologically closely linked, so that while the status of FECA remains uncertain, LECA definitely possessed mitochondria (Figure 2). More recently, thermodynamic and bioenergetic arguments have been used to support a syntrophic origin of eukaryotes and, simultaneously, mitochondria (Lane et al., 2010; Lane & Martin, 2010; Lane, 2011).

One early event during mitochondrial establishment was loss of much of the genome of the α-proteobacterial progenitor, and transfer of many of these genes to the nucleus. In comparison with the >1000 proteins imported into aerobic mitochondria, very few mitochondrial proteins are encoded by extant mitochondrial genomes (Barbrook et al., 2010; Bullerwell & Gray 2004; Gray et al., 2004). Based on modern bacteria, the endosymbiont likely possessed a ∼1–2 Mb genome encoding at least a thousand proteins and RNAs. Extant mitochondrial genomes can be over 2 Mb in size, but in terms of gene content currently range from three protein-coding genes on a genome of ∼6 kb to 97 protein coding genes on a ∼69 kb genome (Bullerwell & Gray 2004). Therefore, the vast majority of proto-mitochondrial genes that were originally endosymbiont-encoded, have been lost with substantial portions transferred to the genome of the host (Rivera et al., 1998). Analyses of extant nuclear genomes indicate that most operational genes, i.e. encoding proteins involved in DNA and RNA functions, have an archaeal origin, while genes encoding metabolic proteins processes are eubacterial (Pisani et al., 2007; Rivera et al., 1998). As the host must have possessed genes encoding metabolic proteins, irrespective of the mechanism of mitochondrial origin (O’Malley, 2010), this suggests that FECA had archaeabacterial metabolism and LECA replaced this system using endosymbiont genes (Doolittle, 1998; Ginger et al., 2010; Müller et al., 2012). An immediate consequence of transferring any essential genes from endosymbiont to host is to make the endosymbiont dependent, and potentially requiring the host to return essential proteins whose genes have been transferred to the nuclear genome. This has been viewed as enslavement, but as the process of endosymbiotic gene transfer seems inevitable, it seems the endosymbiont and host had no choice as to the outcome of their relationship (Doolittle, 1998). This is clearly a further massive revolution in function in the transitional period between FECA and LECA.

Extant eukaryotes use complex protein import mechanisms to target nuclear encoded proteins to various mitochondrial compartments (Neupert & Herrmann, 2007), although more recent studies indicate that many microbial eukaryotes seem able to survive with simpler import mechanisms (Busu et al., 2013; Burri et al., 2006; Dagley et al., 2009; Dolezal et al., 2010; Eckers et al., 2013). Significantly, evolution of mitochondrial targeting signals is not that difficult in terms of sequence evolution and synthetic evolution experiments suggest that such signals can arise readily, while selective pressure is likely greater for prevention of import of inappropriate polypeptides than for failure to translocate a bona fide mitochondrial protein (Allison & Schatz, 1986; Lemire et al., 1989). Further, the mitochondrial import machinery did not arise de novo in the early eukaryotes but via existing bacterial membrane proteins, such as OmpA (Clements et al., 2009; Hewitt, et al., 2011; Selkirk et al., 2012), further lowering the selective barrier.

A major mitochondrial function is to support the electron transport chain (ETC). The ETC transfers electrons from donors (NADH) to an acceptor (O$_2$ in aerobes), via several enzyme complexes residing in the mitochondrial inner membrane, while also pumping protons across that membrane. The proton gradient drives ATP synthesis using a specialised mitochondrial ATPase (F$_{0}$F$_{1}$ ATPase). The ETC is an ATP generation mechanism common to prokaryotes and eukaryotes, suggesting that LECA possessed a mitochondrial ETC. Studies focusing on aerobes suggested an ETC containing five complexes (Complex I; NADH:quinone oxidoreductase, Complex II; succinate dehydrogenase, Complex III; the cytochrome bc1 complex, Complex IV; cytochrome c oxidase Complex V; F$_{0}$F$_{1}$ ATPase). On the one hand, taxonomically broader sampling suggests that the five complex ETC is lineage-restricted and that a variety of shorter ETCs exist across the eukaryotes, although these are likely derived states (Müller et al., 2012). On the other hand, many eukaryotes (some animals, plants and unicellular eukaryotes) possess an alternative oxidase, which sits in the inner mitochondrial membrane and passes electrons directly from a reduced quinone to oxygen. Here electron transfer to/from a quinol is not coupled to proton-pumping. The distribution of alternative oxidase is broad, suggesting that alternative oxidase was present in LECA (Müller et al., 2012), but was subsequently lost from many lineages, rather than being an enzyme introduced by repeated lateral gene transfer. However, alternative oxidase is seldom present in extant α-proteobacteria or archaea for which complete genome sequences are available. This leaves uncertainty as to whether
alternative oxidase was present in the endosymbiont that became the proto-mitochondrion; rather, the alternative oxidase likely provides an example of a feature of the core metabolism introduced early into the eukaryotic lineage, but coming neither from the \(\alpha\)-proteobacterial endosymbiont nor necessarily from the archaean host. This is a theme that is repeated in other core aspects of eukaryotic metabolism.

Recently, it has become evident that mitochondria play a crucial role in the production of iron-sulphur proteins, cofactors with essential roles for many different enzymes (Lill & Kispal, 2000). In fact, FeS clusters are so widespread that it is probable that iron and sulphur played an important role in the origin of life (Hall et al., 1971). There are three principal distinct FeS cluster synthesizing mechanisms: a nitrogen fixation system (NIF), the iron-sulphur cluster system (ISC) and the mobilization of sulphur system (SUF) (Bandypadhyay et al., 2008; Xu & Moller, 2011). Until recently, the ISC system was viewed as the sole essential mitochondrial pathway, as all FeS proteins in the cell depended on this biosynthetic mechanism (Lill & Kispal, 2000), but more recently evidence for a eukaryotic SUF system has appeared (Tsaousis et al., 2012). As archaeabacteria generally use SUF and eubacteria use ISC, it is likely that the mitochondrial endosymbiont brought ISC to the eukaryotes. An important component of SUF is the cysteine desulphurase, which contains the essential Isd11 protein in eukaryotes (Richards & van der Giezen, 2006). Isd11 is absent from eubacteria; instead bacterial cysteine desulphurase requires a distinct protein, YfhJ to function (Pastore et al., 2006; Shimomura et al., 2005). It seems YfhJ was lost early in eukaryotic evolution and was rapidly replaced by the eukaryotic Isd11.

**General cellular metabolism**

As mentioned above, a eubacterial origin for many eukaryotic central metabolic enzymes was evident from early genomic comparisons (e.g. Esser et al., 2004), and is supported by recent analyses (e.g. Thiergart et al., 2012). However, several major pathways, such as glycolysis, have probably experienced significant HGT and gene displacement post-LECA, most notably during the evolution of anaerobic/microaerophilic lineages (Liapounova et al., 2006; Stechmann et al., 2006). Thus catabolism of glucose, the carbon source for ATP production preferred by the majority of eukaryotes, is catalyzed by orthologs of classic Embden-Meyerhof glycolytic enzymes, and not the variants present in some archaea (Sato & Atomi, 2011; Siebers & Schönheit, 2005). If endosymbiosis with the \(\alpha\)-proteobacterial mitochondrial pro-genitor was the key event in eukaryogenesis, facilitating enhanced energy production via oxidative phosphorylation (Lane & Martin, 2010), the preponderance of eubacterial homologs within eukaryotic metabolism suggests that balancing metabolic regulation from the eubacterial endosymbiont might have been easier than integrating and co-regulating metabolism from both the endosymbiont and archaenal host during the FECA to LECA transition.

The presence of mitochondrial sirtuins in animals and trypanosomes (Alsford et al., 2007; Katada et al., 2012) may indicate the possible extent of integration between mitochondrial and nuclear-encoded metabolism in LECA. Sirtuins catalyse NAD\(^+\)-dependent deacetylation of a variety of substrates, including histones. Together with other chro-matin re-modelling enzymes, sirtuins regulate nuclear gene expression in response to metabolite changes in mammalian cells and yeast (discussed in Katada et al., 2012). Thus far in mitochondria, the identified targets of sirtuins are metabolic enzymes. However, the compact packaging of mitochondrial DNA from taxonomically diverse sources, including histone-like proteins in trypanosomatids (Avliyakulov et al., 2004), possibly points towards the presence of a sophisticated mechanism for dual regulation of both mitochondrial and nuclear genomes, and that epigenetic regulation of metabolism was established in the LECA.

Potential similarities between metabolic regulation in LECA and extant eukaryotes are further illustrated by conservation of AMP-activated kinase (AMPK) in all eukaryotes except the microsporidian *Encephalitozoon cuniculi* (Hardie, 2011). In animals, some plants and yeasts, AMPK is responsible for maintaining cellular energy homeostasis, acting on key catabolic and anabolic targets and regulating mitochondrial biogenesis and turnover, suggesting that in LECA the ancestral AMPK functioned similarly (Hardie et al., 2012; Thelander et al., 2004). In *E. cuniculi* the absence of AMPK is an example of reductive evolution, and likely a result of reliance on the host for ATP (Tsaousis et al., 2008). Even in *Giardia*, an excellent example of genomic minimalism, orthologs of AMPK subunits are retained despite a simplified metabolism where mitochondria do not contribute directly to ATP production (Hardie et al., 2003; Morrison et al., 2007). However, determining if an ancestral role for AMPK was to trigger a switch to oxidative metabolism in response to nutrient deprivation requires experimental analysis of taxonomically diverse protists (Hardie, 2011).

For anabolic biochemistry, the biosynthesis of heme, nucleotides and multiple coenzymes is undertaken by many unicellular and multicellular eukaryotes. In contrast, biosynthetic pathways leading to the ‘essential’ amino acids have been lost from animals and many taxonomically diverse protists (Guedes et al., 2011; Payne & Loomis, 2006). The latter organisms tend to be either phagotrophs or parasites. However, retention of highly conserved pathways for essential amino acid biosynthesis, as well as for NO\(_3^-\)/NO\(_2^-\) and sulphate assimilation in fungi, strongly suggests that LECA was capable of extensive macromolecular biosynthesis and similar to many bacteria, for example *Escherichia coli*, which can grow in minimal media. Although LECA was almost certainly a heterotroph, capable of utilising glucose, fatty acids or amino acids as carbon sources, retention of a broad repertoire of amino acid biosynthetic pathways may have been required to support the evolution of the cellular complexity underpinning the FECA-LECA transition. Since the transition was unlikely to have occurred in ecological isolation, FECA and its immediate descendants would have likely been competitively disadvantaged in comparison to prokaryotes if their anabolic metabolism was more limited.

Beyond a conserved core there are clear eukaryote-specific metabolic innovations. A hallmark of eukaryotic biology is sterol biosynthesis, and the few bacteria capable of sterol
biosynthesis are often presumed to do so via HGT from eukaryotes (e.g. Desmond & Gribaldo, 2009). In contrast, sterols are either synthesized or acquired by all eukaryotes, testifying to an ancient origin and function of a biosynthetic pathway, with over 25 enzyme-catalyzed reactions required for de novo synthesis. This also explains the potential of persistent sterol breakdown products as eukaryotic biomarkers within the geological record (Love et al., 2009; Summons et al., 2006). Sterols function most obviously as membrane constituents, regulating membrane fluidity and microdomain partitioning, but also encompass other activities, including key roles in many aspects of development and survival in mammals (Entchev & Kurzchalia, 2005; Kurzchalia & Ward, 2003). A so-called “sparking” role for trace amounts of sterol or sterol-derived metabolites in cell cycle regulation of yeasts and some protists has also been proposed (Nes et al., 2012; Parks et al., 1995), and this raises the possibility of a role for sterols in cell cycle “sparking” in either LECA or an older eukaryotic ancestor.

Further, peroxisomes provide an excellent example of metabolic organelles that are the products of eukaryotic innovation. The presence of peroxisome-bearing organisms in all major eukaryotic groups, albeit with considerable diversity in likely composition between distinct taxa, indicates that peroxisomes were present in the LECA and it is probable that these organelles functioned in diverse lipid metabolism (Gabaldón et al., 2006, 2010). The possibility that peroxisomes had endosymbiotic origins has been overturned by recent phylogenetic analysis of the peroxisomal protein import apparatus and proteomes (Bolte et al., 2011; Gabaldón et al., 2006; Gabaldón & Capella-Gutiérrez, 2010) together with cytological analysis of trafficking pathways of essential peroxisomal membrane proteins (reviewed in Tabak et al., 2008). Crucially, de novo peroxisome formation in S. cerevisiae is dependent upon routing of Pex3 and Pex19, also required for peroxisome formation in mammals, to newly forming peroxisomes via the endoplasmic reticulum, suggesting that peroxisomal membranes arise from the ER (Hoepfner et al., 2005). Lipid-associated pathways reconstructed for LECA peroxisomes include long-chain fatty acid β-oxidation, α-oxidation of branched-chain fatty acids, the glyoxylate cycle, ether lipid biosynthesis and several enzymes of the mevalonate pathway. In many instances the peroxisomal location of these pathways also dictates the presence within the organelles of enzymes associated with NADPH formation (e.g. isocitrate dehydrogenase, glucose-6-phosphate dehydrogenase) and the detoxification of reactive oxygen species (most notably catalase and superoxide dismutase) (Gabaldón et al., 2006; reviewed in Gabaldón, 2010), suggesting that similar metabolic functions plausibly featured in the peroxisomes of LECA.

Of the more extreme examples of metabolic compartmentalization is the essential and exclusive compartmentalisation to peroxisomes of either the first six or seven glycolytic enzymes in trypanosomatids (depending upon the species or life cycle stage examined) (see Gualdrón-López et al., 2012). A recent report of peroxisomal targeting of glycolytic enzyme isoforms in diverse fungi and possible peroxisomal targeting of glycolytic enzymes in mammals (Freitag et al., 2012), suggests either a fascinating example of convergent evolution or implies that the metabolic capabilities of peroxisomes in extant eukaryotes, LECA, or both are significantly underappreciated.

As metabolic compartmentalisation clearly arose early during eukaryotic evolution, so would a need to regulate and co-ordinate Fe-S cluster assembly for incorporation into proteins across multiple cellular compartments, including mitochondria, cytosol, nucleus and ER (Balk & Pilon, 2011). As discussed earlier, mitochondria seemingly retain a universal and conserved role in assembly of Fe-S clusters (see also Lill et al., 2012), and mitochondrial involvement is at several points: in provision of the Fe-S clusters themselves, regulating overall cellular iron homeostasis and providing an activated sulphur compound for export to the cytosol and thence cytosolic Fe-S cluster assembly. The cytosolic Fe-S cluster machinery appears to be another eukaryotic-specific innovation (Lill, 2009; Stehling et al., 2012) that is conserved in all eukaryotes and required for the maturation of essential cytosolic and nuclear proteins, albeit with individual components whose origins can be readily traced to prokaryotes (Allen et al., 2008; Basu et al., 2013; Boyd et al., 2009; Horner et al., 2002). The eukaryotic-specific cytosolic Fe-S cluster assembly pathway is particularly relevant if one considers the enzymes and other proteins present in LECA and potentially FECA, but generally absent from extant α-proteobacteria and archaea. Cytosolic Fe-S cluster assembly requires a protein closely related to Fe-hydrogenase, the defining enzyme of hydrogenosomes and responsible for anaerobic H₂ production in eukaryotes (Balk et al., 2004; Luo et al., 2012; Müller et al., 2012; Song & Lee, 2011). The Fe-hydrogenase-like protein (NAR1 in yeast or iron-only hydrogenase-like protein 1 (IOP1) in animals) probably evolved from a bona fide Fe-hydrogenase (Horner et al., 2002). However, Fe-hydrogenase is present widely in prokaryotes, but poorly represented in extant α-proteobacteria or archaea, suggesting that the eukaryotic Fe-hydrogenase was unlikely derived from the mitochondrial progenitor or archaeal host. A presumed early origin of eukaryotic Fe-hydrogenase may suggest either a more complex syntrophic model for eukaryotic origins (e.g. López-García & Moreira, 1999) or HGT during early eukaryogenesis, with subsequent neofunctionalisation. As most Fe-hydrogenases are oxygen-labile, the finding that Fe-hydrogenase was present early in eukaryotic evolution is also significant in considering whether FECA evolved in an aerobic or microoxic/anoxic environment (e.g. Hug et al., 2010; Müller et al., 2012).

A final example of a metabolic regulatory mechanism is autophagy, which encompasses various pathways for selective and non-selective remodelling of cellular architecture through lysosomal degradation. Macroautophagy in S. cerevisiae utilises over 30 ATG gene-products (Yang & Klionsky, 2010), and the majority are conserved across the breadth of eukaryotic evolution, including TOR kinases that act as master autophagy regulators. This indicates a major presence in the LECA (Brennand et al., 2011; Rigden et al., 2009). Macroautophagy also has several roles in cellular differentiation (Duszenko et al., 2011; Yang & Klionsky, 2010), while a function in turnover of whole organelles in diverse eukaryotes suggests the presence of similar pathways in
versatile early eukaryotes (Herman et al., 2008; Manjithaya et al., 2010). Curiously, while understanding the mechanistic basis and regulation of autophagy in model systems has advanced during the last 15–20 years, the origin of much of the canonical, conserved autophagy machinery remains enigmatic, including the origin of the autophagosome membrane (Mari et al., 2011).

Despite the more limited metabolic repertoire in extant eukaryotes generally when compared with prokaryotic taxa, an essential Fe-hydrogenase-related protein in cytosolic Fe-S cluster assembly and glycolytic enzymes imported to peroxisomes from diverse taxa demonstrate that the full appreciation of the metabolic repertoires of FECA and LECA remains a way off. Regulation of metabolic fluxes in LECA was also likely complex; AMPK was present and data suggest potential epigenetic regulation of metabolism, while a clear role for autophagy in LECA is strongly indicated based on conservation of the ATG genes across the eukaryotes.

**FECA to LECA: A severe bottleneck?**

The events leading to eukaryogenesis, and the transitional period between FECA and LECA, may have become clearer, but the precise sequence of events, what drove them, how much diversity arose and how much was lost during this period remain less well defined. Moreover, it is unknown what the duration of the transition period was in terms of cell/organismal generations. We previously suggested that colonization of the eukaryotic endomembrane system by proto-coatomer-based membrane deforming complexes may reflect a form of intracellular competition and natural selection (Field et al., 2011). While the presence of bona fide prokaryotic proto-coatomer is in doubt (McInerney et al., 2011), the protein architectural elements are clearly present. With expansions of many paralogous families critical for eukaryotic cells, one can speculate that similar selections for protein families able to respond to functionalization during the period prior to LECA occurred, so that specific gene families dominated. With these elements in place, massive and rapid expansion of eukaryotic diversity was perhaps inevitable. The organellar paralogy hypothesis (Dacks & Field, 2007) is an example of this mechanism, and posits evolution of new compartments based on ratchet-like replacement of paralogs within complexes, generating new functions, and could be applied to any modular system (Figure 4). This also implies that alternate evolutionary strategies by post-FECA/pre-LECA organisms were unsuccessful and lost, and the complexity of transition period organisms remains unknown.

What was required to progress from FECA to LECA? Due to prokaryotic relatives of many gene families that mediate eukaryote-specific features (however distant), only a moderate level of protein structural invention may have been required, regardless of how critical such inventions were. What permitted elaboration and expansion of gene families and the consequent rise in cellular complexity? If most of the pieces were already in place in many prokaryotes, why did this transition not happen repeatedly? One attractive explanation is that the acquisition of the mitochondrion, generally agreed to have occurred only once, massively increased energy production, and may have facilitated elaboration of sophisticated membrane structures, the cytoskeletal systems to subtend them and the eukaryotic flagellum (Lane & Martin, 2010). Membrane transport and flagellum-mediated motility are extremely expensive activities in both biosynthetic demand and direct ATP requirements. An alternate model, that phagocytosis was required for acquisition of the α-proteobacterial mitochondrial endosymbiont, is also conceivable. Specifically, an endomembrane system may have evolved early, and could even have been sufficient for dominance within the transitional period. The ability to eat, or at least out-eat, the competition would also have been a powerful selective advantage.

Is this then simply an example of contingency; the first organism to acquire the mitochondrion rapidly dominated the local environment, eliminating all but a restricted lineage of eukaryotes and their descendants? Conceivably, mitochondrial acquisition facilitated even greater exploitation of primitive eukaryotic systems, delivering the coup de grâce to all amitochondrial eukaryotes. For example, it is now clear that the IFT system is related to protocoatomer, and recent data suggest evolution from COP-I (Satir et al., 2007; van Dam et al., 2013); an interesting hypothesis would be that evolution of the flagellum had to await an enabling set of conditions, but increased the selective advantage of transitional eukaryotes once this occurred. This may have been comparatively late as all IFT subunits were present in LECA. One other possible answer to the question of why the transition did not happen frequently is that perhaps it did. Comparative genomics suggests the presence of the meiotic system in the LECA (Ramesh et al., 2005). The ability to recombine and resort genes has the obvious ability to allow rapid innovation. Furthermore, comparative analysis of sexual cycles across eukaryotes implies that facultative sexual stages are common and likely ancestral, providing benefits of meiosis without the disadvantages and dangers of obligately linking meiosis with cell replication (Dacks & Roger, 1999). The other facet to the process is syngamy, or fusion of gametes, which facilitates the sweep of advantageous alleles through a population and allows for both genetically encoded traits (e.g. complex Golgi) and cytoplasmic traits such as mitochondria to arise in independent lineages and then spread through a reticulating population. This even raises the question of whether the observed complexity arose in a single FECA lineage or through multiple transitional lineages that, via interbreeding, “multiplexed” their innovations. This idea, conceptually similar to a fusion origin of eukaryotes, emphasizes that what we reconstruct through comparative genomics is a LECA. All points prior to this are still very much *Terra incognita*, and determining the order of emergence of the various cellular innovations remains an important goal for resolution by future work.

**Reconstructing the LECA**

What has emerged from comparative analysis of cellular systems is the great complexity in the reconstructed LECA, implying an ancestral which organism possessed capabilities exceeding those of many extant eukaryotes. LECA had a mitochondrion, meiotic machinery, a sophisticated and
flexible metabolic capability, a fully differentiated endomembrane system, phagocytosis, actinomyosin and tubulin-based cytoskeletal systems together with a large complement of motor proteins, and finally a nucleus subcompartmentalised into hetero- and euchromatin, together with sophisticated nucleocytoplasmic transport. Subsequently, many lineages simplified various cellular aspects, losing genes that played important roles in their forebears. Due to the limitations of phylogenetic reconstruction, the LECA may have been even more complex than these studies suggest; for example, the presence of multiple subfamily paralogs cannot be reconstructed for LECA, with the consequence that estimates of a LECA Rab complement of twenty could easily be an underestimate (Elias et al., 2012). This caveat may be relevant for many of the gene families considered here, especially those where paralogous families are drivers of new function, including G proteins, SNAREs, protocoatomers, kinesins, dyneins and karyopherins, as well as expanded metabolic enzyme isoforms. Finally, as our understanding of eukaryotic biology is biased towards functions described in animals and fungi, it remains unknown how many gene families present in the LECA may emerge from broader sampling, but which have been specifically lost from animals and fungi. Several such examples have emerged within the membrane-trafficking system recently (Elias et al., 2012; Gabernet-Castello et al., 2013; Schlacht et al., 2013).

Why was the LECA so complex? The answer may simply be that a high level of complexity was required to dominate the early eukaryotic landscape and to occupy a successful position within the ecosystem. Having on board phagocytic capabilities, possibly amoeboid locomotion and a flagellum permits multiple modes of motility, together with the ability to feed on other organisms. As the LECA was a heterotroph and lacked a plastid, it was probably dependent on such abilities. The presence of the mitochondrion would at least partly offset energetic costs of increased complexity, while a highly sophisticated metabolic network with little dependence on a specific nutrient source allowed exploitation of a wide range of carbon sources. The LECA was an aerobe and possessed a TCA cycle, as evidence suggests that anaerobic eukaryotes have arisen from secondary losses of mitochondrial metabolic capacity. The presence of meiosis further implies that the LECA was capable of the generation of gametes, while the probable differentiation of LECA chromatin into hetero- and euchromatin, indicates potential developmental progression via repression of selected gene cohorts and the presence of distinct life stages. This latter aspect could, for example, have facilitated the emergence of quiescent forms allowing survival during unfavorable conditions, as present in testate amoebae, and which may have been present quite early on in post-LECA evolution (Butterfield, 2007). This feature both protects against transient environmental changes and aids in dispersal via traversing hostile environments between favorable locales. In essence, what has emerged is a sophisticated and potentially predatory organism, with great flexibility allowing survival in varied ecological niches.

Two modes of post-LECA evolution have now emerged. Firstly, many taxa acquired significant complexity, for example, vascular plants, metazoa and multiple protist lineages. Here, examples of paralogous expansions and the evolution of novel gene families abound. By contrast, there are many lineages where complexity decreased, encompassing many fungi, some algae, most kinetoplastids, apicomplexans and diplomonads. Many reductions are almost certainly a result of parasitism, frequently associated with reduced metabolic potential, as well as more limited trafficking and cytoskeletal arrangements. In other cases, streamlining may be a result of adaptation to specific environments, where reductive pressure includes energetic reasons to facilitate shorter cell cycle times, and increase competitive advantage (yeasts), or for protection (C. merolae has a reduced endocytic system, likely to protect against a low pH environment). Critically, a LECA with a broad metabolic and cellular functional repertoire would have been best placed for subsequent exploitation of novel niches, a well-equipped explorer, with ample capacity from which to build greater complexity, plus access to a smårgåsbord of functionality from which more limited activities could be selected, providing opportunity for adaptation to a broad range of conditions. Hence a complex LECA may explain the enormous range of life styles and cellular forms that are exhibited by living taxa.

Conclusions, perspectives and many unanswered questions

Progress in the last decade on describing the earliest events in eukaryote evolution has been spectacular, and has advanced from a rather skewed view of ever increasing complexity based predominantly on assumptions, to the appreciation and description of a LECA cellular sophistication that is based on substantial molecular data. The overriding conclusion from all of these studies is of great functional differentiation, and that the LECA was, in many ways, a surprisingly modern organism. While we may never fully understand the life cycle and life style of the LECA, we now have a far more sophisticated view of its likely capabilities, functions and even modes of gene regulation. The importance of secondary reduction to subsequent evolution is also highlighted by the evidence of simpler extant eukaryotes.

The transitional period remains poorly reconstructed, and while there are now several clear biological principles running between FECA and LECA, the ordering of events such as acquisition of the mitochondrion and evolution of the endomembrane system and flagellum remain to be resolved (Figures 2 and 3); not least concerning is the understanding of the state of the FECA/LECA transitional form that acquired the mitochondrion, and how the nuclear membrane itself arose. Resolution of some of these steps from molecular data may become possible in the future, but both this issue and the more accurate analysis of extant eukaryotes require several technical advances.

First, there remains a sensitivity issue: many homologs fail to be detected, or detected with sufficient confidence in diverse genomes for robust conclusions to be drawn. Detailed analysis is time consuming, and even with improved algorithms, models of sequence evolution or phylogenetic approaches there is a clear need for new and more robust...
methods to eliminate both much of this burden as well as, if possible, erroneous calls. Coupled to this is the continual demand for more sequence data, and specifically data from taxa residing at critical positions within the eukaryotic phylogeny. Second, it is critical that such taxa, or at least a well-chosen subset, can be analyzed functionally: such approaches are frequently the only way in which sequence-based studies can be thoroughly validated, or even understood at the cellular level. The impact of neo-functionalisation, for example, amongst paralog families may be a major evolutionary driver, and more detailed insights into these processes can only be gleaned through experiment. Third is the issue of asymmetry, the biasing of analysis due to the significantly greater understanding that we have for Opisthokonta taxa, and relatively poor details in most other supergroups. Taken together, tackling each of these issues will provide a greater appreciation of eukaryotic diversity and the role this plays in the context of ecology and disease mechanisms, potentially opening up the transitional period and finally break the ‘asymmetry’ problem (Dacks & Field, 2007). An era of molecular paleontology may have arrived.

Acknowledgements

MCF thanks Michael Rout for many insightful discussions concerning evolution of gene expression and trafficking systems and suggestions with figures.

Declaration of interest

The authors acknowledge various funding agencies that have supported many aspects of our work in this area: the Wellcome Trust (program grant 082813 to MCF), the Medical Research Council (project grants to MCF), the Marie Curie fund (FP7-PEOPLE-2011-IEF to VLK), the Johannes Veldkamp Foundation (MvDG), the Biotechnology & Biological Sciences Research Council (BW and MLG), The Royal Society (MLG), the Leverhulme Trust (MLG) and the Natural Sciences and Engineering Research Council (of Canada) (JBD). BW is a recipient of University of Nottingham New Lecturer support. JBD is the Canada Research Chair in Evolutionary Cell Biology. None of the funding agencies were involved in the decision to publish or in the selection of content. The authors have no conflict of interest to declare.


