Although much of cell biology focuses on diverse proteins and the machines constructed from them, thousands of molecular forms of lipids are utilized across the Tree of Life, with dozens to hundreds frequently being deployed within individual species (Fahy et al. 2005; Oger and Cario 2013; Brügger 2014; Sohlenkamp and Geiger 2015; Buehler 2016). Lipid membranes play a central role in all cells, providing a barrier to the external environment, and in eukaryotes providing the basis for intracellular substructure. By colocalizing genomes with the products they produce, cell membranes ensure individuality, a critical requirement for heritable evolutionary processes. Consisting of millions to billions of noncovalently linked molecules, membranes are typically highly fluid, while also providing platforms for the residence of key membrane-bound proteins with diverse functions. Most notable are the trans-membrane proteins associated with channels, importers, and exporters (for ion and nutrient acquisition and balance; Chapter 17), electron transfer chains and ATP synthases (for energy production; Chapter 18), and signal transduction (for environmental sensing and communication; Chapter 21). Taken together, the proteins involved in these diverse functions typically comprise 10 to 30% of the total set of proteins encoded in the genomes of species.

Although lipids are used for multiple cell functions, including energy storage and occasionally as cofactors for protein function, we focus here specifically on membranes, most notably on their unique structural properties and considerable energetic costs. As will be discussed in detail below, lipid membranes constitute a two-dimensional liquid with intrinsic biophysical features endowing both flexibility and resistance to breakage and leakage. Specific aspects of their structural features play a central role in molding different cellular functions. Once these key structural features are understood, it becomes clear that the universal use of lipids in biology is unlikely to be simply a frozen accident. Indeed, it is difficult to see how the establishment and diversification of cellular life would be possible without them.

The matter of the energetic cost of lipids is particularly germane to understanding the evolution of eukaryotic cells, which are typically laden with internal membrane boundaries delineating various internal organelles, and this will be taken up in detail in Chapter 16. One can marvel at the many aspects of cell biology accomplished with organelles, but it is highly uncertain that eukaryotic cell structure is intrinsically superior to that of prokaryotes. Although the evolution of the eukaryotic cell introduced new ways of living, this was accompanied by a substantial increase in investment in the basic cellular machinery imposed by the deployment of internal membranes. As will be reviewed in the latter part of the chapter, establish-
ment of the intricate system of vesicle transport in eukaryotes was also associated with a significant investment in a diverse repertoire of proteins required in vesicle formation, transport, and localization.

**Molecular Structure**

Rather than presenting an encyclopedic coverage of the various classes of membrane lipids, the intention here is simply to provide an overview of the key relevant issues from an evolutionary perspective. The majority of membrane lipids in eukaryotes and bacteria reside in two families, the glycerophospholipids and the sphingolipids, which share some basic structural features. In both cases, a polar (hydrophilic) head group is attached to a negatively charged phosphate, which in turn connects to a linker, glycerol in the case of glycerophospholipids and sphingosine in the case of sphingolipids (Figure 14.1). Glycerophospholipids have two fatty-acid chains attached to the glycerol linker, whereas in the case of sphingolipids, sphingosine provides one built-in chain which joins with another fatty acid.

Such modular structure allows for enormous diversity of lipid types through the exchange of variable parts, including the nature of the head group. Structural diversity is also associated with the number of carbon atoms in the fatty-acid chains and the number and locations of double C=C bonds in the fatty-acid chains. (Double bonds are referred to as unsaturated, as the carbon atoms are bound to only single hydrogens). The most common head groups in glycerophospholipids are choline, ethanolamine, serine, glycerol, inositol, and phosphatidyl glycerol, and the resultant lipids are known as phosphatidylcholines, phosphatidylethanolamines, phosphatidylserines, phosphatidylglycerols, phosphatidylinositols, and cardiolipins. Fatty-acid chain lengths are typically in the range of 14 to 22 carbons, whereas the number of C=C bonds is usually 0 to 5, and these features have a strong influence on membrane width and flexibility.

In contrast to the water compatible head-groups of membrane lipids, the fatty-acid tails are highly hydrophobic, earning lipids the distinction of being amphipathic (or synonymously, amphiphilic) molecules. As a consequence of this structure, the roughly cylinder-shaped lipid molecules naturally self-associate into organized aggregates, with their hydrophobic tails lying parallel to each other (Figure 14.2). Moreover, the most thermodynamically stable state is one in which two sheets (leaflets) of such aggregations align with their tails contraposed. Such a configuration minimizes the contact of hydrophobic groupings with water, with the water-exposed sides consisting of flexible walls of noncovalently bound head groups. Lipid bilayers serve as the outer membranes of all cells, and also comprise the membranes surrounding all of the organelles of eukaryotic cells, including the endoplasmic reticulum, the golgi, the nuclear envelope, mitochondria and plastids, and transport vesicles. The internal hydrophobic environment of lipid bilayers makes them extremely impermeable to charged ions, which must then be imported / exported through gated channels comprised of proteins (Chapter 17).

Frequently, the proportional usage of lipid types differs between leaflets, endowing them with different properties. Moreover, in various phylogenetic groups, there are added layers of complexity to the combinatorial scheme noted (Buehler 2016).
The head groups of some lipids are modified by additions of various small molecules. Some fatty acids contain methyl side branches, and some contain ring structures at the ends. Although the precise functions of most such variants are unknown, many appear to play roles in thermal stability, permeability, and protection from various damaging agents.

Most classes of phospholipids are shared among bacteria and eukaryotes, although their relative usages can vary dramatically (Table 14.1). The maintenance of such lipid diversity over billions of years of evolution is likely related to the variation in structural flexibility endowed by alternative head groups and fatty-acid chains. Indeed, microbial species are commonly capable of shifting the lipid profiles of their membranes in response to environmental change, e.g., using glyceroglycolipids instead of phospholipids when phosphorus is limiting. With increasing temperature, many cells physiologically remodel their membranes to contain lipids with longer and more saturated fatty acids or to incorporate different head groups. By this means, membrane fluidity and permeability is kept relatively constant, a process known as homeoviscous adaptation (Sinensky 1974; van de Vossenberg et al. 1995). Without a shift in lipid composition, increased temperature magnifies membrane permeability and fluidity, eventually leading to the loss of homeostasis, the disorganization of membrane proteins, and cell death. Homeoviscous adaptation has been observed in all domains of life (Hazel 1995; Toyoda et al. 2009; Oger and Cario 2013; Ernst et al. 2016), and can be especially refined in organisms such as mammalian pathogens that regularly experience large shifts in temperature (environment vs. host) (Li et al. 2012). The mechanisms for membrane monitoring, essential for an adaptive physiological response, involve proteins that regularly probe membranes for their fluidity (Harayama and Riezman 2018).

Finally, as noted in Chapter 1, the structures of lipid molecules in archaea differ significantly from those of eukaryotes and bacteria (Koga and Morii 2007; Chong 2010; Oger and Cario 2013; Buehler 2016). Most notably, archaea generally utilize isoprenoid hydrocarbon chains (which contain methyl side groups off the tails, rather than simple hydrogen atoms). Despite these differences, however, most of the head groups utilized in phospholipids in eukaryotes and bacteria are also deployed in their archaeal counterparts. A particularly unique aspect of archaeal membranes is the partial use of bipolar lipids with head groups at both ends, which span the entire width of the membrane. Some archaeal tails also contain cyclopentane carbon rings.

<table>
<thead>
<tr>
<th>Organism</th>
<th>PC</th>
<th>PE</th>
<th>PG</th>
<th>PI</th>
<th>PS</th>
<th>C</th>
<th>LPG</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacillus subtilis</td>
<td>0.00</td>
<td>0.24</td>
<td>0.35</td>
<td>0.00</td>
<td>0.00</td>
<td>0.18</td>
<td>0.23</td>
<td>0.00</td>
</tr>
<tr>
<td>Caulobacter crescentus</td>
<td>0.00</td>
<td>0.00</td>
<td>0.88</td>
<td>0.00</td>
<td>0.00</td>
<td>0.12</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Escherichia coli</td>
<td>0.00</td>
<td>0.75</td>
<td>0.19</td>
<td>0.00</td>
<td>0.00</td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Staphylococcus aureus</td>
<td>0.00</td>
<td>0.00</td>
<td>0.53</td>
<td>0.00</td>
<td>0.00</td>
<td>0.07</td>
<td>0.40</td>
<td>0.00</td>
</tr>
<tr>
<td>Zymomonas mobilis</td>
<td>0.13</td>
<td>0.62</td>
<td>0.20</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Table 14.1. Fractional contributions of lipid molecules to plasma membranes in select species. The surveys exclude contributions from sterols and proteins, and are generally given for optimal growth conditions. The central point is that distantly related species often utilize the same types of lipids, although at different frequencies.
Eukaryotes:

<table>
<thead>
<tr>
<th>Organism</th>
<th>PC</th>
<th>PE</th>
<th>PG</th>
<th>PI</th>
<th>PS</th>
<th>C</th>
<th>LPG</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mus musculus, thymocytes</td>
<td>0.57</td>
<td>0.21</td>
<td>0.00</td>
<td>0.07</td>
<td>0.10</td>
<td>0.00</td>
<td>0.00</td>
<td>0.06</td>
</tr>
<tr>
<td>Vigna radiata, seedlings</td>
<td>0.47</td>
<td>0.35</td>
<td>0.05</td>
<td>0.05</td>
<td>0.08</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Dictyostelium discoideum</td>
<td>0.29</td>
<td>0.55</td>
<td>0.01</td>
<td>0.08</td>
<td>0.02</td>
<td>0.02</td>
<td>0.00</td>
<td>0.03</td>
</tr>
<tr>
<td>Dunaliella salina</td>
<td>0.15</td>
<td>0.41</td>
<td>0.15</td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.22</td>
</tr>
<tr>
<td>Saccharomyces cerevisiae</td>
<td>0.17</td>
<td>0.18</td>
<td>0.00</td>
<td>0.23</td>
<td>0.21</td>
<td>0.03</td>
<td>0.00</td>
<td>0.19</td>
</tr>
<tr>
<td>Schizosaccharomyces pombe</td>
<td>0.42</td>
<td>0.23</td>
<td>0.00</td>
<td>0.25</td>
<td>0.03</td>
<td>0.06</td>
<td>0.00</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Abbreviations: PC, phosphatidylcholine; PE, phosphatidylethanolamine; PG, phosphatidylglycerol; PI, phosphatidylinositol; PS, phosphatidylserine; C, cardiolipin; LPG, lysophosphatidylglycerol; O, other.

References: Mm: van Blitterswijk et al. (1982); Vr: Yoshida and Uemura (1986); Dd: Weeks and Herring (1980); Ds: Peeler et al. (1989); Sc: Zinsen et al. (1991); Tuller et al. (1999); Blagovic et al. (2005); Sp: Koukou et al. (1990); Bs: Op den Kamp (1969); Lopez et al. (1998); Cc: Contreras et al. (1978); Ec: Raetz et al. (1979); Rietveld et al. (1993); Sa: Haest et al. (1972); Mishra and Bayer (2013); Zm: Carey and Ingram (1983).

### Membrane Structure

Owing to the difficulties of moving a polar headgroup through the hydrophobic interior of a bilayer, flip-flops of molecules between layers are negligible unless promoted by specialized transport proteins. However, as the individual molecules within lipid bilayers are held together by noncovalent forces, movement of molecules within a leaflet is essentially unavoidable. Thus, membranes have an enormous amount of flexibility and fluidity, while also maintaining relatively stable sheet-like structures.

This being said, the multitude of lipid types are not homogeneously distributed within membranes. Rather, like molecules tend to aggregate as they encounter each other by diffusion, leading to a sort of self-organizing phase separation, and generating patchy variation in membrane properties. These island-like domains, often referred to as lipid rafts, are themselves capable of diffusive movement across the membrane. Such variation is relevant to the distribution of membrane proteins, as a stable platform for a membrane-spanning protein requires a good match between the membrane thickness and a protein’s hydrophobic trans-membrane domains. Hence, particular types of proteins are associated with particular lipid rafts, further adding to the membrane chemical heterogeneity (Mitra et al. 2004).

The individual lipid molecules within a membrane rotate around their long axes, with the lateral diffusion coefficients of individual glycerophospholipid molecules in a bilayer being $D \approx 2$ to $4 \mu m^2/sec$ at $25^\circ$ C (Devaux and McConnell 1972; Wu et al. 1977; Jin et al. 1999). Thus, letting $D = 3$, and assuming unbiased directional movement such that the mean squared distance traveled over a two-dimensional surface is $4Dt$, where $t$ is measured in seconds (Chapter 2), the mean absolute distance traveled is $\sim 2\sqrt{Dt}$, or $\sim 3.5 \mu m$ in 1 sec. This implies that individual molecules can diffuse the full lengths of moderately sized cells in a matter of seconds. Rates of diffusion of lipid rafts are one to two orders of magnitude lower, declining with the size of the raft (Schitte et al. 2017; Zeno et al. 2018).

To put this into perspective, recall from Chapter 3 that diffusion rates of proteins are on the order of 10 to 40 $\mu m^2/sec$ within a cytoplasmic environment. Thus,
despite the fluidity of membranes, the diffusion inhibition from molecular crowding within membranes is substantially greater than in the cytoplasm. The lateral diffusion coefficients of membrane proteins is lower than that for lipids, e.g., $\sim 0.04$ to $0.3 \, \mu m^2/sec$ for mitochondrial proteins (Gupte et al. 1984), and Kumar et al. (2010) found that such coefficients decline from $\sim 0.2$ to $0.02 \, \mu m^2/sec$ in E. coli as the number of membrane-spanning helices in proteins increases from 3 to 14. Thus, although membrane proteins are mobile in an absolute sense, they are effectively stationary from the perspective of cytoplasmic proteins.

Lipid molecules are not strictly cylindrical in shape. Rather, depending on the size of the head group relative to the tail width, the overall shape can be closer to a cone or a reverse cone, with unsaturated fatty-acid tails tending to fan out. As a consequence, curvature is induced when molecules of particular geometric shapes associate with each other (Figure 14.2). This serves as a simple structural mechanism to reduce the bending energy necessary to mold membranes into particular shapes, and in part explains the differential distribution of lipid types on the inner vs. outer leaflets of membranes.

Generation of stronger curvature (as in vesicle formation) typically requires energy input beyond that associated with background thermally induced molecular motion (Helfrich 1973), implying the involvement of ATP- or GTP-hydrolyzing processes as sources of bending energy. A wide variety of membrane proteins have functions specifically associated with the bending and sculpting of membranes into specific forms (Shibata et al. 2009; Jarsch et al. 2016). The insertion of hydrophobic wedges will naturally cause a membrane to bend towards the narrower end of the inserted protein, as occurs when ATP synthase molecules inhabit the tips of cristae on the internal membranes of mitochondria. Motor proteins moving along microtubules or actin filaments (Chapter 15) can pull membranes into tubular forms. Assemblies of scaffolding proteins with natural curvature and an affinity for lipid head groups can force lipid bilayers to conform to the same curvature, and are widely used in the formation of vesicles (as described in more detail below). Transmembrane proteins, which transverse the space between two membranes, help maintain specific distances between layered sheets.

**Eukaryotes and the Organellar Explosion**

One of the eukaryotic hallmark distinctions from prokaryotes is the proliferation of internal membrane-bound organelles. Prominent in almost all eukaryotic cells are the endoplasmic reticulum (the site of production of many proteins and lipids), the golgi (the site of secondary processing and transport), and lysosomes and peroxisomes (devoted to degradation), all of which must date to LECA. Special attention will be given to mitochondria and chloroplasts in Chapter 22, as these are exogenous in origin, having a well-known endosymbiotic ancestry dating back to a specific member of the alpha-proteobacterial lineage (as recorded in mitochondrial genomes). All other eukaryotic organelles are believed to be endogenous in origin, having developed by descent with modification, although a possible indirect association with the colonizing mitochondrion has been suggested (Gould et al. 2016).

A key problem is how individual organelle types achieve their distinctive shapes.
This may in part be a consequence of the physical features of the component molecules described above, combined with the tendency for self-aggregation, but other directly evolved mechanisms may be involved. For example, major portions of the core endoplasmic reticulum (ER, the source of much of the eukaryotic cell’s lipid and protein production) have a layered but spiraling architecture, resembling a parking structure (Terasaki et al. 2013), and the ER is also continuous with the nuclear envelope (Foundations 14.1). The peripheral ER often exhibits a matrix-like structure involving narrow tubules (Nixon-Abell et al. 2016). It appears that these and many alternative morphologies can be generated by the relative concentrations of just two types of membrane-shaping proteins, one encouraging flat sheets and the other curvature (Shemesh et al. 2014).

Many eukaryotes lack classical stacked golgi (the central clearing house for vesicle traffic and post-translational modification), yet contain the genes associated with golgi trafficking, suggesting independent loss of this morphology at least eight times and hence dramatic variation in the architectural features of this organelle as well (Dacks et al. 2003; Mironov et al. 2007; Mowbrey and Dacks 2009). Many organelles have unfused membrane-contact sites (e.g., covering between 2 and 5% of the the surface area of the ER and contacted mitochondria; Phillips and Voeltz 2016), with functional relevance as means for inter-organellar communication, further contributing to the complex interior of eukaryotic cells.

Often, it is assumed that the invention of organelles led to fundamentally superior organisms, with increased cellular complexity being viewed as a positive endowment (Lane and Martin 2010; Gould 2018). However, although organelles allow eukaryotes to accomplish cellular tasks in novel ways relative to prokaryotes, there is no evidence that these are more efficient in any absolute sense, and as will be discussed below, the investment in internal membranes comprises a substantial energetic burden on cells. There is also no evidence that the emergence of eukaryotes led to the extinction of any prokaryotic lineages. In the absence of a cellular fossil record, obtaining such evidence is admittedly challenging, but as an explicit counterexample, consider the frequent dominance of cyanobacteria over various eukaryotic phytoplankters in the open waters of freshwater lakes.

Moreover, prokaryotes are not constrained from evolving internal cellular structures. For example, the planctomycetes, a group of aquatic bacteria, have a substantial tubular network of internal membranes (Fuerst and Sagulenko 2011; Acehan et al. 2014; Boedecker et al. 2017), reminiscent of but not clearly of the same origin as the endomembrane system of eukaryotes. Although the functions of such membranes are not fully resolved, one structure (the anammoxosome) sequesters a reaction that converts nitrite and ammonium ions to nitrogen gas (van Niftrik and Jetten 2012). The planctomycetes are also capable of endocytosis and reproduce by budding, both of which may involve the use of eukaryote-like mechanisms. Other related groups of bacteria (e.g., verrucomicobia and chlamydiae) also appear to have endomembranes, and given the recent discoveries noted in Chapter 2, it would not be surprising if lineages of archaea are found to have them.

Many other examples of compartmentalized organelles can be given for bacteria. The photosynthetic machinery in cyanobacteria is sequestered within a carboxysome (Savage et al. 2010). A microcompartment for ethanolamine metabolism consisting of hexameric protein subunits is present in E. coli (Tanaka et al. 2010), and
Salmonella harbors another such structure for propanediol utilization (Chowdhury et al. 2015). In all of these cases, the intracellular compartment consists of an assembly of several protein multimers, much like the capsids of viruses. However, other lipid-bound organelles are known as well. For example, magnetotactic bacteria contain magnetosomes, crystals of magnetite or other iron-phosphate granules enclosed by phospholipid membranes (Byrne et al. 2010; Jogler et al. 2011). The giant cells (up to 1 mm in length) of Epulopiscium, a symbiotic bacterium inhabiting triggerfish guts, contain stacked “vesicles” of unknown function near the cell membrane (Robinow and Angert 1998). Nor can the existence of intracellular endosymbionts be viewed as purely eukaryotic attributes. For example, the β-proteobacterial endosymbionts of mealybugs themselves harbor γ-proteobacterial symbionts (von Dohlen et al. 2001; Thao et al. 2002), and bacteria are known to inhabit the cytoplasm of the cyanobacterium Pleurocapsa (Wujek 1979).

Vesicle Trafficking

Although many cellular functions in prokaryotes are governed by diffusion-like processes, the large cells of eukaryotes rely on active transport of macromolecules for numerous cellular functions. Such processes range from the internalization of extracellularly derived cargoes, the movement of molecules from one organelle to another, and the transport of proteins and RNA molecules across the nuclear envelope. Intracellular transport raises many of the same issues encountered in the consideration of metabolic (Chapter 18) and transcription regulatory (Chapter 20) networks, most notably the specificity of intermolecular interactions. Essentially every transport pathway involves one or more modes of intermolecular communication – correct substrates must be identified to the exclusion of erroneous and sometimes harmful cargoes, and these must be delivered to their appropriate destinations.

Eukaryotes achieve such ends by deploying lipid-bound vesicles in a wide range of activities, including endocytosis and exocytosis, digestion, and transport between the endoplasmic reticulum and the golgi. The life cycle of a vesicle involves assembly at sites of initiation, delivery through the intracellular domain, and docking and fusion to another lipid-bound compartment at the site of delivery (Figure 14.3). As these processes are all occurring simultaneously, and in bidirectional fashions, the quantitative partitioning of lipid membranes throughout the cell can remain in a roughly steady-state condition, such that the lipids lost by donors are balanced by those gained by recipients, despite substantial traffic between compartments. The rate of membrane flux can be quite high. For example, amoeboid cells can internalize the equivalent of the entire surface membrane in the form of endocytic vesicles in just an hour (Ryter and de Chastellier 1977; Bowers et al. 1981; Steinman et al. 1983).

The origins of vesicle transport remain obscure, with few obvious orthologs of any components known in prokaryotes. However, the planctomycete situation lends credence to the idea that some aspects of an endomembrane system may have been present in the primordial eukaryote, i.e., the first eukaryotic common ancestor (FECA) (Lonhienne et al. 2010). Such a hypothesis is consistent with phylogenetic analyses suggesting the origin of various aspects of endocytosis as well
as the secretory system prior to LECA (Jékely 2003; Dacks et al. 2008; Wideman et al. 2014; Klinger et al. 2016).

Pointing out that many bacteria release outer-membrane vesicles into the extracellular environment, Gould et al. (2016) suggested that the eukaryotic endomembrane system, along with the replacement of archaeal membrane components (and many other features) by those from bacteria (Chapter 2), originated via the mitochondrial endosymbiont. One concern with this argument is the absence of any mechanistic evolutionary argument for how the simple production of intracellular vesicles by the primordial mitochondrion could have become whole-sale transformed into a highly organized and nuclear-encoded vesicle transport system (involving hundreds of genes) by the host cell. Given the energetic costs of internal membranes (Chapter 16), and the myriad of microbes that do fine without them, it is far from clear that the evolution of the eukaryotic endomembrane system was promoted by adaptive processes. Indeed, once in place, an endocytic pathway provides a direct route for exploitative cellular entry and exit by pathogens (e.g., Heuer et al. 2009; Szumowski et al. 2014; Shen et al. 2015; Renard et al. 2015; Shi et al. 2016).

As will be outlined below, although the basic mechanisms are conserved across all eukaryotes, the vesicle-transport system is quite complex, involving numerous specialized sorting pathways with respect to donor and recipient membranes, cargoes, etc. Typically, hundreds of genes are devoted to such processes. How such complexity and diversity arose is far from clear. One idea, the organelle-paralogy hypothesis (Figure 14.4), invokes repeated rounds of gene duplication and joint co-evolution of clusters of components toward more specialized functions (Dacks and Field 2007; Mast et al. 2014). Although such descent with modification provides a logical argument for diversification (Ramadas and Thattai 2013), the population-genetic conditions that would allow for such specialization have not been worked out and seem rather formidable, as the subcomponents of each descendent pathway must not only evolve pathway-specific features but also relinquish preduplication features that would promote crosstalk.

The following provides a brief overview of what little is known about the various steps from cargo uptake to delivery. The focus here is on general principles, and not on the enormously detailed molecular mechanisms that can be explored further in many specialized publications.

Fundamentals 14.1. Probability of preservation and subdivision of labor by duplicated interactions. In Chapter 10, the concept of subfunctionalization of the two members of a duplicated gene pair was introduced. The central point is that genes often have multiple, independently mutable subfunctions; after gene duplication, these can become reciprocally silenced, leading to more specialized daughter genes. The question of interest here is how frequently pairs of interacting genes (e.g., members of a transport pathway) can partition up their functions after both members of the pair are simultaneously duplicated (as would occur following a whole-genome duplication event). In the extreme, this can lead to two independently operating pathways. Before addressing this point, it will be useful to understand the quantitative expectations for the single-gene case.

Consider the case illustrated in Figure 14.5, where initially a single protein-coding
gene has a coding region and two regulatory elements for different subfunctions. It will be assumed here that all mutations with significant effects on gene activity are degenerative in nature, with loss of single subfunctions occurring at rate $\mu_s$ for each regulatory element and mutations that eliminate whole-gene function arising at rate $\mu_n$. Following gene duplication under this model, there are two possible fates: one of the genes will become completely silenced (nonfunctionalization), returning the system to the initial state of a single active gene, or the two genes will become mutually preserved by subfunctionalization, as in this case joint retention is necessary to retain the full complement of gene activity. It will be assumed that the loss of single gene features is a neutral process owing to the redundancy of the two-gene system, so that each step of permissible mutations proceeds at a rate equivalent to the mutation rate.

If subfunctionalization is to occur, the first mutation to fix must be of the subfunctionalizing type, the probability of which is $2\mu_s/(\mu_n + 2\mu_s)$. This expression follows from the fact that there are three ways to mutate each fully endowed gene, two of which eliminate single subfunctions. Conditional upon arriving at this initial state, the remaining fully intact gene cannot be nonfunctionalized, as this would fully eliminate one subfunction entirely, although it can lose the remaining redundant subfunction, and the partially partially incapacitated copy can be completely silenced by either a nonfunctionalizing mutation or by a mutation to the remaining subfunction. The total permissible mutation rate during the second step is then $\mu_n + 2\mu_s$, with the probability that the second mutation leads to joint subfunctionalization being $\mu_s/(\mu_n + 2\mu_s)$. The total probability of subfunctionalization is equal to the product of the two stepwise probabilities,

$$P_{\text{sub,1}} = \frac{2\mu_s^2}{(\mu_n + 2\mu_s)^2},$$

with the probability of nonfunctionalization being

$$P_{\text{non,1}} = 1 - P_{\text{sub,1}}.$$  \hspace{1cm} (14.1.1a)

Now consider the situation in which a pair of interacting genes (e.g., a donor and its recipient) is duplicated simultaneously, with each pair having two independently mutable interactions (as indicated by the different colors and complementary shapes in Figure 14.5). Following the same mutation scheme noted above, there are four possible final fates of this system: 1) complete subfunctionalization and the preservation of two specialized single-subfunction interactions; 2) one fully endowed donor gene, and two specialized recipients; 3) two specialized donors, and one fully endowed recipient (not shown); and 4) nonfunctionalization of one donor and one recipient and return to the single-pair (ancestral) situation.

Multiple paths involving multiple steps lead to each of these final outcomes, rendering the book-keeping difficult, so only a few of the results will be sketched out. It is relatively straight-forward to obtain the probability of complete subfunctionalization, as this requires that a series of four subfunctionalizing mutations occur before any gene is completely nonfunctionalized. Moreover, specific subfunctions must be retained in each gene – the two donor genes must preserve alternative subfunctions, as must the two recipient genes. The probability of each specific subfunctionalizing mutation is $\mu_s/(\mu_n + 2\mu_s)$, and because there are two ways by which the donor copies can be resolved (blue in one, and green in the other, in either order), and likewise for the recipient genes, the probability of preservation of the four-gene set by subfunctionalization is

$$P_{\text{sub,2}} = \frac{4\mu_s^4}{(\mu_n + 2\mu_s)^4},$$

which is equivalent to $P_{\text{sub,1}}^2$.

We next consider the probability of return to a single-pair system, which requires the complete loss of function of one donor and one recipient gene. There are three
ways by which this endpoint can come about. First, if the initial mutation is non-
functionalizing, which occurs with probability \( \frac{\mu_n}{\mu_n + 2\mu_s} \), the system effectively
returns to a one-gene system, as only the remaining pair of duplicates is now capable
of further evolution. The net probability of return to the ancestral state by this path
is then simply
\[
P_{\text{non},a} = \frac{\mu_n P_{\text{non},1}}{\mu_n + 2\mu_s}.
\]

Second, there are two paths to a one-pair system if the first mutation is of the
subfunctionalizing type (probability \( \frac{2\mu_s}{\mu_n + 2\mu_s} \)) and the second is nonfunctional-
izing. When there are three fully functional and one subfunctionalized genes, the total
rate of permissible mutations in the next step is \( d = 3(\mu_n + 2\mu_s) \). The probability that
the subfunctionalized copy is silenced in the next step is \( \frac{\mu_n + \mu_s}{\mu_n + 2\mu_s} \), and this
returns the system to the identical situation noted in the previous paragraph – one
fully endowed gene of one type and two of the other, with a probability of nonfunc-
tionalization of \( P_{\text{non},1} \) in the final step. Alternatively, a member of the pair of fully
endowed genes will be nonfunctionalized with probability \( \frac{2\mu_n}{d} \), in which case the
remaining single-subfunction gene will be lost with probability \( \frac{\mu_n + \mu_s}{\mu_n + 2\mu_s} \),
leaving one fully endowed donor and recipient gene. Collecting terms, the probability
of complete nonfunctionalization by these two path types is
\[
P_{\text{non},b} = \frac{2\mu_s(\mu_n + \mu_s)}{3(\mu_n + 2\mu_s)^2} \left( P_{\text{non},1} + \frac{2\mu_n}{\mu_n + 2\mu_s} \right).
\]

The third potential path to complete nonfunctionalization follows when the first
two mutations are of the subfunctionalizing type. This can only occur if one of each
such mutations is allocated to a donor and the other to a recipient gene (as otherwise,
both members of donor and/or recipient would be permanently preserved by subfunc-
tionalization). The probability of this starting point is \( \frac{\mu_s}{\mu_n + 2\mu_s} \cdot \frac{4\mu_s}{d} \).
Completion of the path to complete nonfunctionalization then requires that one of the
single-subfunction genes is silenced by the next mutation, the probability of which is
\( \frac{\mu_n + \mu_s}{\mu_n + 2\mu_s} \), and that the final single-subfunction gene is also silenced in the
remaining step, which also occurs with probability \( \frac{\mu_n + \mu_s}{\mu_n + 2\mu_s} \). Collecting
terms,
\[
P_{\text{non},c} = \frac{8\mu_s^2(\mu_n + \mu_s)^2}{3(\mu_n + 2\mu_s)^4}.
\]

Summing up terms, the total probability of return to a single-pair system by
random silencing of one donor and one recipient gene is
\[
P_{\text{non},2} = P_{\text{non},a} + P_{\text{non},b} + P_{\text{non},c}.
\]

The probability of partial preservation is
\[
P_{\text{par}} = 1 - P_{\text{sub},2} - P_{\text{non},2},
\]
with half of these cases involving two specialized donors and one two-subfunction
recipient, and the other half the reciprocal situation.

The solutions of the above formulae, given in Figure 14.5, are simple functions of
the ratio \( \mu_s/\mu_n \). As noted already, the probability of complete subfunctionalization of
a two-component pathway is substantially smaller than that of subfunctionalization of
a single two-function gene, being equivalent to the square of the latter. On the other
hand, the probability of partial preservation of a pathway (involving just one member
of the pair) is \( \approx 2P_{\text{sub},1} \), provided \( \mu_s < \mu_n \) (which is likely to be the usual case).
Thus, given that there are two genes involved, only one of which will be preserved by
subfunctionalization, the probability of preservation per gene is very nearly the same with the duplication of a two-gene system as in the case of single-gene duplication.

**Vesicle production.** Rather than forming *de novo*, vesicles are typically derived via the invagination (endocytosis) or budding (inter-organelle transport) of a pre-existing membrane, with the lipid molecules then being recycled via cycles of fission and fusion. Vesicle birth generally involves the recruitment of specific proteins dedicated to inducing membrane curvature. Several classes of membrane-deforming proteins are known (Field et al. 2011), although we will focus on the most well-studied types, all of which assemble into cage-like lattices that support developing vesicles before they are eventually pinched off from parental membranes by a rope of dynamin molecules.

Three types of vesicle coating are known. Clathrin-coated vesicles import cargoes across the cell membrane in the form of endosomes, and are also deployed in the trans-golgi network. COPI (coat protein I)-coated vesicles carry cargoes between different golgi compartments and from the golgi to the endoplasmic reticulum. COPII-coated vesicles export cargoes from the endoplasmic reticulum. In all cases, protein lattices are produced via the assembly of radiating trimers (clathrin and COPI) or dimers (COPII), which then coassemble into higher-order structures with distinct geometric shapes and sizes (Figure 14.6).

Although many of the details remain to be worked out, clathrin vesicle formation initiates when specialized adaptor proteins bind to the source membrane and then recruit the coat proteins (McMahon and Boucrot 2011; Boettner et al. 2012; Kirchhausen et al. 2014). Adaptor proteins (often called adaptins) are thought to recognize specific cargo-recruitment molecules, which in turn have affinities to specific cargo types. In this sense, adaptins serve as an intermediary link between specific cargoes and coat recruitment. This is a simplified view in that many other ancillary proteins are thought to be involved in clathrin recruitment, some of which appear to be lineage specific (Adung’a et al. 2013). The details on how clathrin-coated pits come to contain their cargoes or even whether cargoes are essential to trigger vesicle formation remain unclear (Kaksonen and Roux 2018), although it appears that pits may stochastically develop and abort, with cargoes and their receptors diffusing laterally through the lipid membrane of the cell until they blunder upon a developing pit. Once within an incipient pit, stochastic dissociation becomes progressively more difficult as vesicle curvature becomes more pronounced (Weigel et al. 2013). Cargoes with higher affinities for such settings will naturally accumulate to a greater extent.

Given the widespread presence of clathrin throughout the eukaryotic domain, the logical conclusion is that LECA deployed clathrin-coated vesicles (Field et al. 2007). Although no organized comparative study has been performed, it is clear that substantial diversification of clathrin-coated vesicles has occurred, with diameters ranging from about 30 to 200 nm among observed species (McMahon and Boucrot 2011; Kaksonen and Roux 2018). It remains to be seen whether a simple physical basis to such size differences resides in the architecture of the clathrin molecule itself, e.g., the numbers of α helices constituting the long connecting arms, or in
turgor-pressure differences among cell types (which would influence the bendability of membranes), or both. Structural flexibility also results in variation in the sizes of COPI- and COPII-coated vesicles (Faini et al. 2012).

Several different adaptor proteins are known to be associated with different clathrin-coated vesicles, and although little is known about their modes of specificity, all are heterotetramers comprised of two large, one medium, and one small subunit. Each subunit has orthologs across all adaptors and is also related to a particular protein involved in the COPI coat protein (which itself is a heptameric complex) (Schledzewski et al. 1999). Moreover, the two large subunits appear to have arisen by a gene duplication that preceded the origin of the different adaptor complexes, and the same is true of the medium and small subunits. These observations suggest that the ancestral adaptor may have been heterodimeric (or perhaps a dimer of heterodimers) consisting of just single small and large subunits (Schledzewski et al. 1999). Under this hypothesis, subsequent duplication of both subunits followed by divergence led to the heterotetrameric state, with further duplications and divergence of all subunits leading to the various classes of adaptors and to the COPI-coat protein.

Because all five known adaptor proteins as well as COPI-coated vesicles are found throughout the eukaryotic domain, their diversification must also have preceded LECA. Using the form of the genealogical relationships among the various complexes then provides a potential means for ordering events in the diversification of vesicle-trafficking pathways on the lineage connecting FECA to LECA (Figure 14.7). Such a perspective leads to a scenario in which the COPI coat diverged from an early adaptor, which then underwent a series of four duplications leading to a total of five different adaptors (Hirst et al. 2011). This implies that one of the first events in the evolution of vesicle trafficking was a split between COPI coating of vesicles and the ancestral adaptor, which underwent further rounds of diversification, possibly prior to the emergence of clathrin. The form of relationship between the gene-family members further suggests that the deployment of adaptors in endosomes emerged prior to the expansion of their use in the trans-golgi network, and that the adaptor involved in internalization was one of the latest arrivers.

It is likely that other adaptor-like complexes remain to be discovered, given that a distantly related ortholog has recently been found based on structural information (rather than sequence divergence) (Hirst et al. 2014). Phylogenetically, this complex appears to be nestled between the adaptor proteins and the COPI subunits, and although the full complex is present in green plants and slime molds, it has been partially or fully lost from several eukaryotic lineages (including metazoans). Unlike the complexes described previously, this new complex is a heterohexamer.

Central to the completion of vesicle formation, at least in metazoans, is the protein dynamin (Praefcke and McMahon 2004). After the development of clathrin-coated invaginations begin, dynamin assembles into collar-like helical structures and uses mechanical energy derived from GTP hydrolysis to pinch off the neck. Oligomerization stimulates GTPase activity, resulting in a chain-like reaction that generates the overall mechanical force once a critical length of the collar has been achieved. Dynamin appears to be absent from most eukaryotic lineages, which nonetheless often harbor a separate clade of dynamin-like proteins that likely serve a similar function (Liu et al. 2012; Briguglio and Turkewitz 2014).
**Vesicle delivery.** Once formed, vesicles must find their way to an appropriate donor, and in the process, avoid fusing with inappropriate membranes. Central to these processes are members of the RAB GTPase family of proteins, which specify the locations to which vesicles are delivered. RABs act as switches by undergoing conformational changes when bound by GDP (inactive state) or GTP (active state). Specific GEFs (guanine exchange factors, which promote GDP release) catalyze conversion from the GDP- to GTP-bound forms, leading to activation, whereas GAPs (GTPase-activating proteins) do the reverse, leading to GTP hydrolysis and inactivation. Still other proteins are involved in RAB activation/deactivation cycles; e.g., RAB escort proteins deliver their cognate RABs to specific cellular locations, whereas after inactivation, RABs are recycled back to their membranes of origin via specific GDP-dissociation inhibitors (GDIs). The N-terminal residues of RABs contain specificity information, whereas the C-terminals are involved in targeting and adhesion to destination lipid membranes; still other enzymes endow these regions with post-translational modifications that confer specificity (Pylypenko and Goud 2012).

The main point here is that the transport of specific kinds of vesicles to specific locations involves an elaborate choreography of several layers of specialized protein-protein interactions – RABs and their various interactors. Eukaryotic species typically harbor 10 to 100 distinct RABs, and phylogenetic analysis suggests that LECA may have contained up to 23 RAB genes (Elias et al. 2012; Klöpper et al. 2012), with some lineages then experiencing losses of distinct family members. Fungi commonly encode no more than a dozen (Brighouse et al. 2010).

Also involved in vesicle delivery to specific sites are a large set of SNARE (soluble N-ethylmaleimide-sensitive factor-attachment protein receptor) proteins, which act in a zipper-like fashion, with coordination between specific sets attached to vesicles and others attached to recipient membranes. The appropriate recognition groups attach to each other in four-helix bundles known as SNARE pins. The four main types of SNAREs are classified into four subsets reflecting their positions in the four-helix bundles (Klöpper et al. 2007). Like the coatamer proteins, each of the classes of these attachment proteins diversified into subfamilies prior to LECA. Many of the subfamilies have expanded in lineage-specific ways, but with no obvious relationship to organismal complexity (Sanderfoot 2007; Kienle et al. 2009).

**Evolutionary issues.** Although many details remain to be elucidated, it is clear that the vesicle trafficking system of Eukaryotes diversified through gene duplication prior to LECA. This is true for essentially all of the components that have been subject to comparative sequence analysis – the adaptor proteins, the RABs, the SNARES, etc. However, gene duplication does not immediately ensure diversification in function, especially in a multilayered system that requires coordinated behavior of the component parts. One possibility (raised in previous chapters) is that at least two whole-genome duplication events occurred on the path from FECA to LECA, providing opportunities for coevolutionary coordination among duplicated parts. However, how such coordinated evolution proceeds is unclear, as a mechanism (such as sub- or neofunctionalization) is required to ensure the joint and coordinated preservation of mutually interacting pairs of components.

Equally challenging is understanding how the multiple layers of communication
necessary for specialized trafficking pathways evolve. Adaptor proteins provide the interface between various cargoes and the specific coat proteins of vesicles; specific types of RAB proteins must specify specific types of vesicles as well as specialized effector molecules associated with subcellular localization; and specific pairs of vesicle and target SNARE proteins must recognize each other to ensure vesicle delivery to proper destinations. Although such a system might be viewed as exquisitely intricate, it comes at a substantial energetic cost, and the target size of a pathway for degenerative mutations is proportional to the number of components.

The Nuclear Envelope

If there is an iconic feature of the eukaryotic cell, it is the housing of the genome inside the nucleus. Rather than floating free in the cytoplasm, the nucleus is surrounded by a double membrane (involving two bilayers), with the outer layer being continuous with the ribosome-associated ER and periodically bending around at nuclear pores to form the inner membrane (Figure 14.8). There is also a proteinaceous support layer interior to the nuclear envelope, consisting of lamins in metazoa and amoebozoa (Simon and Wilson 2011; Burke and Stewart 2013) and apparently unrelated proteins in plants and other organisms (Cavalier-Smith 2005).

Among other things, genomic sequestration separates transcription from translation, paving the way for the emergence of introns that must be spliced out of pre-messenger RNAs (Lynch 2007). It is through the nuclear pores that mRNAs are actively exported to the cytoplasm and nuclear proteins (e.g., transcription factors and histones) are imported. There is evidence in flies that clusters of proteins are sometimes exported as particles to the cytoplasm by budding of the inner nuclear membrane and vesicle transport to the outer membrane (Speese et al. 2012). However, it remains unclear whether this is a common phenomenon, and it is virtually certain that the bulk of transport proceeds through pores.

**Nuclear-pore architecture.** Nuclear pores are lined with a giant nuclear-pore complex (NPC), consisting of ∼500 to 1000 individual Nup (nucleoporin) proteins encoded by ∼30 separate genes. The NPC is the largest protein complex in most eukaryotic cells, exceeding the mass of a ribosome by more than ten-fold (Field et al. 2014; Devos et al. 2014; Beck and Hurt 2017). A brief excursion is now taken to describe the NPC, as this will reinforce the contention that large complexes within eukaryotic cells are typically grown out of a series of gene duplication events, while also illustrating that despite its conserved functions, the NPC has experienced considerable diversification at the architectural level. There are interesting lessons in coevolution to be learned as well, as pathogens that require entry into the nucleus (e.g., for replication and/or transcription) must successfully navigate the NPC.

The core of the NPC is both vertically and radially symmetrical, consisting of four stacked rings (two on the nuclear side and two on the cytoplasmic side), each comprised of eight spokes, which in turn consist of two parallel columns of several proteins (Figure 14.8). The proteins in adjacent columns are related as pairs, each of which is derived by gene duplication (Alber et al. 2007). This one-to-one correspondence of multiple pairs of duplicates is again consistent with a massive
amount of duplication activity, if not a whole-genome duplication, in the ancestor leading to LECA.

Consistent with the theme of descent with modification, the evolutionary origin of the NPC appears to be intimately related with the proteins associated with the vesicle-production machinery. Most notably, the observation that the core proteins of the inner rings appear to be related to the membrane-bending proteins involved in vesicle formation (COPI, COPII, and clathrin) motivates the hypothesis that all of these molecules are derived from a common-ancestral molecule, deemed the protocoatamer (Devos 2004; Mans et al. 2004; Alber et al. 2007a,b; Brohawn et al. 2008).

Comparison of the parts list in diverse species suggests that LECA had an NPC structure very much like that in today’s species. Further evidence that the details of nuclear-pore construction are derived from predecessors in the vesicle-transport system is the use of a membrane-sculpting protein, known as ESCRT, for fusing the nuclear membranes at the pore junctions and in the development of internal vesicles of cells (Vietri et al. 2015).

There are, however, some component differences among lineages (Mans et al. 2004; Bapteste et al. 2005; DeGrasse et al. 2009; Neumann et al. 2010; Devos et al. 2014). As one example, the multimeric coatamer complex used in the two yeasts S. cerevisiae and S. pombe differ in terms of the number of subunits (Liu et al. 2012; Stuwe et al. 2013). The overall mass of the S. cerevisiae NPC is only ~50% of that of the human NPC, owing to a reduction in the number of subunits banding together to comprise the ring in S. cerevisiae. Experiments have shown that simple changes in the expression of subunit genes can lead to an alteration in the overall structure, suggesting a simple path to variation in pore composition/size within and among species (Rajoo et al. 2018). Larger compositional changes are known as well. For example, the deployment of proteins on the nuclear and cytoplasmic sides of the pore is asymmetrical in the case of yeast, animals, and land plants, but relatively symmetrical in the case of trypanosomes (Obado et al. 2016).

Finally, it is worth noting that the NPC has evolved a number of secondary functions including involvement in chromosome organization and positioning and the mediating of transcription of tRNAs and mRNAs (Fahrenkrog et al. 2004; Xu and Meier 2007; Strambio-De-Castillia et al. 2010; Ikegami and Lieb 2013; Vaquerizas et al. 2010). In yeast, and likely many other species, a number of genes have short motifs that target them spatially to the nuclear periphery via interactions with the NPC (Ahmed et al. 2010). Thus, the NPC evolved to become the hub of many activities beyond cargo transport.

**Nuclear transport.** The nuclear pore is lined with a large number of FG Nups, each containing up to 50 phenylalanine (F)-glycine (G) repeats. These highly unstructured molecules can be viewed as a spaghetti-like sieve through which cargoes bound by appropriate nuclear transporter proteins are actively delivered and inappropriate molecules are excluded (Sorokin et al. 2007; Grünwald et al. 2011; Hülsmann et al. 2012; Vovk et al. 2016). Such selective filtering demands a means of molecular communication between the FG Nups and the transport mechanism, as well as between the cargoes and their transporters (Figure 14.8). The latter typically involves a nuclear localization signal on cargo protein molecules, which attracts a
nuclear transporter protein, often called a karyopherin or importin/exportin. Such signals are generally quite simple, typically involving three or four consecutive basic amino acids (arginine or lysine), although the consensus sequence appears to vary among species (Kosugi et al. 2009). A separate set of transporter proteins is assigned to mRNA export.

The exact mechanisms of facilitated cargo transport are not fully resolved, but the process is fast, allowing the delivery of up to 1000 molecules per second per pore (Yang et al. 2004). Transport is governed by gradients of Ran-GTP and Ran-GDP, associated with the transporter-cargo complexes – Ran-GTP binds to the import complex on the nuclear side of the membrane, releasing the cargo, and then recycles to the cytoplasm during cargo export, where upon release Ran-GDP is returned with its own carrier to the nucleus and converted to Ran-GTP, maintaining the Ran gradient necessary for efficient transport. Specific enzymes devoted to the Ran-GDP/GTP interconversion cycle reflect the need for still another means of molecular communication in the nuclear-transport pathway.

Although the basic mechanism of communication between karyopherins and FG Nups remains conserved across taxa, there is drift in the language of communication across lineages. For example, human transport substrates are not imported into the nucleus of *Amoeba proteus* unless they are coinjected with human importins (Feldherr et al. 2002). Among yeast species, the FG Nups have diverged at the sequence level at much higher rates than other genes, with the greatest elevation arising in sequences interspersed between the Nup repeats (Denning and Rexach 2007). The ciliate *Tetrahymena thermophila* harbors two nuclei (the transcriptionally silent micronucleus and the transcriptionally active macronucleus), one of which has pores lined with FG Nups, while the other has Nups with novel NIFN repeats, implying distinct permeability of the two nuclear membranes (Iwamoto et al. 2017).

**Evolutionary considerations.** The universal presence of a nuclear envelope in eukaryotes presents two major evolutionary questions. First, what were the driving forces underlying the emergence of the nucleus, if any? Second, once established, what secondary evolutionary challenges / opportunities did the nuclear envelope impose on other aspects of cellular evolution?

It is not clear that genome sequestration would have any intrinsic advantage in a prokaryote-like ancestor, and the failure of any prokaryote to make such a transition over billions of years suggests that there is none. Nonetheless, two hypotheses have been proposed for the origin of a nuclear barrier based on a negative-selection scenario surmised to have been unique to eukaryotes. Martin and Koonin (2006) proposed that the origin of introns (intervening sequences of messenger RNAs that must be spliced out to yield a productive mRNA) would have forced the evolution of the nuclear envelope to prevent the early translation of inappropriate (not yet spliced) messages. Jékely (2008) suggested an alternative scenario, with the origin of the mitochondrion forcing the sequestration of nuclear-encoded genes. Here the idea is that once the ribosomal protein-coding genes of the primordial mitochondrion were transferred to the nucleus of the host cell (Chapter 21), there would have been a risk of constructing chimeric (and likely malfunctional) ribosomes consisting of mixtures of proteins from both host and endosymbiont genomes. In principle, this problem might have been avoided by assembling the cytosolic ribosomes prior to nuclear
export, and addressing the mRNAs for the mitochondrial ribosomal protein-coding genes to the mitochondrion.

One difficulty with both of these arguments is the assumption of an initial harmful condition that the host cell is unable to escape from. If a problem was deleterious enough to encourage a massive repatterning of cellular architecture, why wasn’t the original mutational variant that created such a dire situation simply removed from the population by purifying selection? A plausible alternative is that no such threat existed, and the nuclear envelope evolved prior to introns and mitochondrion-to-host genome transfer. Although this leaves the explanation for the initial origin of the nucleus unresolved, alternatives that involve direct promotion might involve exogenous selective forces including invasive genomic parasites. Most bacterial genomes are largely devoid of mobile-genetic elements, in principle because of the typically large effective population sizes of such species (Lynch 2007). In contrast, few eukaryotes are able to cleanse themselves entirely of such elements, and a large fraction of many eukaryotic genomes is a result of the activities of parasitic DNAs (Chapter 16).

Many intracellular parasites depend on access to the host genome for survival. Among the most prominent of these are the mobile genetic-elements that literally reside within the nuclear genome – transposons and retrotransposons. To produce their encoded mobilization factors necessary for proliferation, such elements must be transcribed and translated, and the resultant products must be able to return to susceptible genomic territories. The FG-Nup-gated nuclear pores serve as a primary guardian against uncontrolled element spread, although any sequence changes used to thwart mobile-element entry must be compatible with normal cellular activities.

Dozens of examples exist for the coevolution of Nups and genomic parasites. For example, yeast retrotransposons have a requirement for the host-cell FxFG repeats in Nup124 (Dang and Levin 2000; Kim et al. 2005), although nuclear-pore associated factors have also been found to have inhibitory effects on retrotransposition (Irwin et al. 2005). Notably, the same Nup protein is exploited by HIV-1 (Varadarajan et al. 2000; Woodward et al. 2009; Lee et al. 2010). On the other hand, Nup124 prevents entry of hepatitis B virus, specifically via the FXFG repeats (Schmitz et al. 2010). Many other exogenous viruses have been found to engage in genetic conflicts with Nups of their host species (e.g., Gallay et al. 1995, 1997; Strunze et al. 2005; Satterly et al. 2007; Bardina et al. 2009; Porter and Palmenberg 2009).

Given the potentially high evolutionary rates of the nuclear-pore components driven by infectious agents, and the NPC’s involvement in chromosome organization and interactions with the spindle during meiosis, it would not be surprising if the divergence of the NPC at the sequence level played a central role in the emergence of species isolating barriers. This could happen if coevolutionary changes of interacting NPC components within species led to situations in which cross-species interactions in hybrids led to negative functional and/or assembly consequences. Although few isolating barriers have been elucidated at the molecular level in any species, in one of the major engines of speciation research, the genus Drosophila, negative interactions between heterospecific Nups have a direct role in hybrid incompatibility, with the causal genes having evolved at highly elevated rates, apparently driven by positive selection (Presgraves 2007; Presgraves and Stephan 2007; Tang and Presgraves 2009). It remains unclear whether such divergence has been driven by
centromeric drive (Chapter 6), by the kinds of pathogen interactions noted above, or by a combination of both.

Finally, we note that under the assumption that the rate of export of transcripts from the nucleus is limited by the surface area of the nuclear envelope, Cavalier-Smith (1978, 2005) suggested the need for a strong coordination between nuclear and cell volumes. Drawing from observations (mostly in land plants; Price et al. 1973) of an association between genome size and nuclear volume, his nucleoskeletal hypothesis postulates that organisms with large cells evolve large genome sizes as a means to support a large nuclear membrane. Under this view, DNA has a pure structural role, independent of its coding content, with a larger nuclear envelope leading to an associated increase in the number of pores, which in turn supports an enhanced flow of mRNAs required to maintain the needs imposed by large cell size. The limited amount of comparative data suggests a roughly constant scaling of total nuclear pore number with nuclear size, with a pore density generally between 5 and 15/µm² (Figure 14.9).

There are, however, a number of observations that shed doubt about the validity of the nucleoskeletal hypothesis. First, it is unclear that transport rates through pores (as opposed to association rates between cargoes and transporters) is the limiting factor in material transport. Empirical studies suggest the latter, with transporter efficiency being greatly compromised by off-binding to nonspecific substrates (Riddick and Macara 2005; Timney et al. 2006). Second, as pointed out in Chapter 6, nuclear volume does not appear to be regulated by the amount of DNA in a cell. Third, increases in genome size in organisms with larger cell sizes may simply be an indirect consequence of the latter experiencing higher levels of random genetic drift and being more vulnerable to the passive expansion of excess DNA (Chapter 8). The most notable source of genome expansion is mobile-element activity (Lynch 2007), a highly mutationally hazardous enterprise for building a nuclear support structure.

The central issue with the nucleoskeletal hypothesis is that the data do not support such a general relationship between genome and cell sizes (Figure 14.9). Although some groups of eukaryotes do arguably exhibit an increase in genome size with cell volume (Figure 14.9), the slope of the scaling relationship is far below the value of 1.0 expected if nuclear volume is determined by bulk DNA. A number of eukaryotic groups show no positive scaling, and this is also true when considering eukaryotes as a whole. In contrast, there is a weak but significantly positive scaling between genome size and cell volume for both heterotrophic and photosynthetic bacteria, neither of which have nuclear envelopes. The latter pattern is largely due to the fact that bacteria with larger cells generally have genomes with larger numbers of genes.
Literature Cited


at cell growth temperatures. Biochemistry 38: 13275-13278.


Yang, W., J. Gelles, and S. M. Musser. 2004. Imaging of single-molecule translocation through


Figure 14.1. Schematics of the structures of the two major classes of lipid molecules. The glycerophospholipid is saturated as the fatty-acid tails contain only single carbon-carbon bonds. The sphingolipid has a single double carbon-carbon bond, denoted by the double line. A third common group of membrane lipids consists of a diverse array of sterols (not shown), which lack head groups and intercalate between the fatty-acid tails of the amphiphilic components of membranes.
Figure 14.2. **Left**) Lipid bilayers, with the head groups in red and the fatty-acid tails in black. In the lower left, some individual molecules have C=C (unsaturated) bonds, yielding slightly kinked tails and a more open membrane. **Right**) The width of the tail region relative to the head group determines the tendency of a membrane to curve inwardly vs. outwardly.

Saturated lipids only

Mixed saturated and unsaturated
Figure 14.3. Generic schematic of a vesicle-transport pathway with several stages. **Left**) Cargo capture: external cargo molecules (red) are initially bound by specific cargo-transport proteins (blue). Vesicle budding and coating: specific adaptor proteins (light green) bind to the cargo receptors, and in turn recruit vesicle coating proteins (dark green), which induce membrane curvature (dark grey). **Right**) Vesicle scission: coat proteins continue to be recruited, and the stem is eventually squeezed off with a concatamer of dynamin molecules (yellow coil). Vesicle uncoating: the coat proteins are removed, leaving the lipid-bound vesicle free to bind to a recipient membrane. Tethering: a specific RAB protein (blue) provides recognition between the vesicle and a tethering effector molecule (green), and vesicle and target SNARE proteins (orange and lavender) join to seal the final connection. Fusion: after docking with the recipient membrane, the cargo is unloaded.
Figure 14.4. Organelle paralogy hypothesis (Dacks and Field 2007; Mast et al. 2014). **Left** An ancestral communication mechanism between two molecular structures, e.g., an adaptor protein and its cargo receptor; a RAB protein and a tethering molecule; or a v-SNARE and a t-SNARE. (See Figure 14.3). **Center** Nested sets of duplications of the genes for both participants eventually leads to pairs with specialized functions. **Right** These specializations lead in turn to partitioning with respect to subcellular functions.
Figure 14.5. Above) The probabilities of the individual steps involved in the preservation of two duplicate genes, each with two independently mutable subfunctions. Below) The alternative fates of duplicated pairs of interacting genes, and their probabilities. In both cases, the fate probabilities are functions of the ratio of rates of subfunctionalizing to nonfunctionalizing mutations ($\mu_s/\mu_n$).
Figure 14.6. Schematics of the higher-order structure of the proteinaceous coats of eukaryotic lipid vesicles. **Top**) The basic structures of clathrin and COPI coats are homotrimeric subunits, whereas that for COPII vesicles is a heterodimer. These subunits then organize into lattices with distinct geometric shapes, the specific dimensions of which are defined by the lengths of the domains of the monomeric subunits. **Bottom**) A monomeric subunit of clathrin. Each linear domain of the arm consists of a long series of α helices, the numbers of which define the overall dimensions of the lattice. It remains to be seen whether such structures vary in any meaningful way phylogenetically. From Edeling et al. (2006) and Harrison and Kirchhausen (2010).
Figure 14.7. A view of the timing of the evolutionary diversification of the five known adaptor proteins and the COPI coat subunit. All of the nodes on the tree emerged prior to the last eukaryotic common ancestor (LECA), as all components are distributed throughout the entire eukaryotic phylogeny. Thus, the structure of the tree yields a hypothesis about the order of events in the functional diversification of adaptor proteins, e.g., implying a likely early role for COPI, but a relatively late recruitment of adaptor proteins to the trans-golgi network. From Hirst et al. (2011).
Figure 14.8. A hierarchical view of the nuclear envelope (NE) and the nuclear pore complex (NPC). Upper left) The outer layer of the NE is continuous with the endoplasmic reticulum. Perforating both sides of the NE are nuclear pores. Lower left) The NPC, through which cargos must be transported, is a complex structure consisting of hundreds of proteins encoded by multiple loci. Right) The central core of the NPC consists of eight spokes, each consisting of two columns, for a total of 16 columns per pore, all of which are organized into four cylindrical layers.
Figure 14.9. Left) Scaling of the total number of nuclear pores with nuclear volume, $33.1V^{0.78}$ ($r^2 = 0.94$). Because area scales with $V^{2/3}$, this implies that the mean number of pores per area scales only weakly with nuclear size, being proportional to $\sim V^{0.15}$. Right) The scaling of genome size with cell volume. For both heterotrophic bacteria and cyanobacteria, there is a significant positive scaling, with genome size (in megabases) being $4.1V^{0.21}$ ($r^2 = 0.24$) and $2.7V^{0.19}$ ($r^2 = 0.81$), respectively. However, in eukaryotes, only the relationship for chlorophytic algae is significant, and the overall relationship is weak as well.