æ

17. THE COSTS OF CELLULAR FEATURES

2 September 2021

Although it is common to view all cellular features as products of adaptive processes, it should now be clear that this starting point is generally no more than an outgrowth of a tradition in which natural selection is viewed as the only evolutionary force of significance. Estimates of the adaptive value of alternative manifestations of a trait (phenotypic variation) can be derived by measuring the fitness of individuals with different phenotypic values (Walsh and Lynch 2018). However, it is one thing to identify optimum phenotypic values, but quite another to determine the overall consequences of developing and exhibiting the trait at all. Virtually all structures and functions of cells require energetic expenditures for construction and maintenance, and the total benefit of a trait needs to be considered in the light of this baseline investment. Indeed, a cellular feature that appears to be adaptive in its current context need not have endowed any net benefits to the ancestor in which it first emerged, having instead become established in an effectively neutral manner, potentially guided by biased mutation pressure.

To understand when such scenarios are possible and in what population-genetic contexts, and to more broadly appreciate the relevant investments that cells make in different types of functions, a quantifiable measure of the costs of cellular attributes is essential. In principle, such costs might be measured as the decline in cell fitness were the trait to be expressed while conveying no benefits. In reality, however, such a measure is nearly impossible for many well-established traits for the simple reason that once integrated into critical cellular pathways, a trait will accumulate secondary side effects over evolutionary time. For example, modifications of gene-expression levels will often have side effects (e.g., promiscuous binding or aggregation) that are irrelevant to the basic construction / maintenance costs of gene products.

Thus, we require an indirect way of summarizing the baseline costs of simply expressing and maintaining a trait that does not require invasive manipulation of the cell, and beyond this, we require a way to quantify the fitness consequences of such costs. To provide a quantitative framework for addressing these issues, the conceptual link between cell biological expenditures and evolution will first be outlined. This will then be followed by applications to several key features exhibiting a substantial gradient in complexity across the Tree of Life, most notably the expansion of gene number and gene-architectural complexity and of the investment in membrane-bound organelles in the eukaryotic domain. These analyses will show that such elaborations need not have been driven by positive selection for cellular complexity, but rather may be inevitable passive consequences of the diminished efficiency of selection resulting from a reduction in effective population size. Al-

though such a scenario may leave the impression of the cumulative development of a long-term fitness drag on populations of reduced size, this need not be the case. This is because embellishments incorporated into the genomic / cellular real estate by nonadaptive mechanisms can also serve as novel substrate for future adaptive evolution by descent with modification.

The Bioenergetic Cost of a Cellular Feature

We start with the cost of a simple cellular feature, e.g., a noncoding RNA, a protein molecule or complex, or the membrane of a cellular inclusion. Regardless of their fitness benefits, all cellular traits entail some baseline expenditures on construction and maintenance. This motivates the need for a universal currency by which costs can be measured for a wide variety of traits in ways that generalize across all phylogenetic lineages, and for such purposes there seems to be little alternative than to rely on a measure of energy utilization. In the fields of ecology and traditional evolutionary biology, the maximization of energy flow through biomass production has long been viewed as a target of natural selection, given that all processes devoted to survival and offspring production require energy (e.g., Lotka 1922; Van Valen 1976, 1980).

Although analyses might be carried out with alternative limiting factors (e.g., carbon or some other key nutrient), the nature of limiting nutrients can vary among phylogenetic lineages and even within species growing in different environments, restricting the general utility of such metrics. For example, silicon availability can be critical to diatom growth, but nearly irrelevant to most other organisms. Likewise, nitrogen will rarely be limiting to a species capable of nitrogen fixation. Ultimately, regardless of the substrates and elemental building blocks being consumed, their entry into all aspects of cellular maintenance and biomass production must be coupled with energy, and most notably the operational and maintenance costs of cellular features can only be measured in terms of energy utilization.

As to the energetic currency to be used, there seems to be little alternative than the consumption of ATP molecules. Across the Tree of Life, hydrolysis of ATP to ADP is universally deployed in the vast majority of energy-requiring processes, including universally conserved pathways such as the TCA cycle, modes of amino-acid biosynthesis, and mechanisms of assembly of cellular polymers (e.g., DNA, RNA, and protein). Some processes involve the hydrolysis of other nucleotides such as CTP and GTP, but the energy release here can still be counted in terms of ATP equivalents, as can the use of coenzymes such as NADH and NADPH in electron-transport chains to produce ATP.

Thus, in keeping with previous efforts in microbial bioenergetics (Bauchop and Elsden 1960; Atkinson 1970; Stouthamer 1973; Tempest and Neijssel 1984; Russell and Cook 1995), all costs described below will be defined in units equivalent to numbers of ATP hydrolyses. Additional justification for this approach is that although the yields of microbes grown on alternative substrates (per unit carbon consumed) can vary substantially with the nature of the substrate, once the appropriate energetic conversions are made, the number of ATP hydrolyses necessary to build an offspring cell are found to be relatively constant (Chapter 8).

The direct cost of a trait can be subdivided into three components: 1) biosynthesis of the basic building blocks; 2) assembling of the building blocks into the full structure; and 3) maintenance (Figure 17.1). First, with respect to biosynthesis, nearly all cellular features are assembled from four types of monomeric building blocks: amino acids, nucleotides, lipids, and carbohydrates. If not provided by the outside environment, such molecules must be synthesized within the cell by processes requiring carbon skeletons and consuming energy (derived, for example, from transformations of glucose or acetate to precursor metabolites such as pyruvate and acetyl CoA and then to downstream products). When monomeric building blocks are available externally (a situation enjoyed by predators), the reliance on *de novo* biosynthesis will be diminished, but there will still be costs of resource acquisition and of various transformations of the precursors (e.g., branch-point metabolites). Second, the assembly cost of a cellular feature is the sum of requirements for construction from its monomeric building blocks. For example, protein assembly requires polymerization of the constituent amino acids, addition of post-translational modifications, and folding the amino-acid chain into the appropriate globular form. Finally, cellular traits almost always experience maintenance costs, e.g., accommodation of molecular turnover, and identification and elimination of cumulative errors.

The sum of costs noted above represents the baseline investment that must be made in a cellular feature regardless of its benefit to the host cell. Given the near universality of many biosynthetic pathways and enzyme-reaction mechanisms, all three sets of costs can often be calculated from information residing in the biochemistry, biophysics, and cell biological literature. Nonetheless, these three direct-cost components do not fully describe the consequences of a trait's presence for the cell. Even if the trait under consideration pays for itself by endowing the cell with increased fitness (in excess of the direct baseline costs), trait construction and maintenance will still impose a drain on resources that could have been allocated to other essential cellular functions. For example, when metabolic precursors that are generally processed for ATP production are instead allocated to the production of a focal trait (i.e., as carbon skeletons), the loss of availability for other purposes represents an opportunity cost (Figure 17.1). This follows from the fact that the production of an additional building block (or entire trait) adds to a cell's lifetime energy budget requirements, while subtracting from the prior pool of resources available for other functions, which must be replaced (generally by extending the cell cycle). Atkinson (1970) defined opportunity costs as the "prices of metabolites," and seemingly independent of him, Craig and Weber (1998) and Akashi and Gojobori (2002) used this approach to partition the costs of amino acids into components associated with the utilization of metabolites and subsequent investments in product synthesis.

Summing up, the total cost of synthesizing and maintaining a trait and diverting components from alternative usage is the sum of the direct and opportunity costs,

$$c_T = c_D + c_O, \tag{17.1}$$

where all costs represent the cumulative expenditures over the entire lifespan of the cell. The direct cost, c_D, reflects actual ATP hydrolysis (or related) reactions resulting in heat dissipation in the cell. However, c_O represents a diversion of metabolic precursors and will not be manifested in heat production, given that no ATP is

produced or consumed. It may be argued that c_O should not be included in cost analyses in situations where energy is a nonlimiting resource. If, for example, the energy extracted from the food source is in excess supply relative to a key elemental resource (e.g., carbon, nitrogen, phosphorus, or iron), c_O may largely be irrelevant to the fitness of the cell, reducing the energetic cost from a fitness perspective to $c_T \simeq c_D$. It remains unclear, however, how common such conditions arise, particularly when one considers that energy is used not only in the production of biomass, but is the key contributor to all activities involved in cell maintenance.

An observation relevant to this issue derives from studies of microbes growing on defined media with all carbon and energy being provided by a single compound and all other nutrients being in excess supply. In this case, the growth yield per carbon consumed increases linearly with the substrate heat of combustion (which is inversely related to the degree of oxidation) up until a threshold value, and thereafter levels off (Figure 7.8). This suggests that below a critical substrate value of $\simeq 10$ kcal/g carbon, growth is limited by energy, whereas a food supply above this threshold contains excess energy relative to carbon content required for growth. Notably, the most common substrate used in growth experiments with microbes, glucose, has a heat of combustion of 9.3 kcal/g carbon, close to the threshold at which growth is equally limited by carbon and energy. Very few commonly used substrates have heats of combustion much beyond the apparent threshold (values being 11.0, 13.6, and 14.8 for glycerol, ethanol, and methanol, respectively), providing further justification for a focus of energy as a basis for cost measures.

The Evolutionary Cost of a Cellular Feature

Given a measure of the absolute cost of a cellular feature, the overall implications depend on the cell's lifetime energy requirements. From a cell physiological perspective, the energetic cost of the feature must be scaled relative to the total cost of building and maintaining the cell (Chapter 8), and evolutionary considerations further require that this relative measure be converted to an appropriate metric of the impact on fitness.

Supposing the cell has a baseline total energy budget per cell cycle of C_T (which includes the costs of growth and maintenance, with a capital C denoting a whole-cell cost), the presence of the trait under consideration (c_T) influences the lifetime energy budget such that $C_T = C'_T + c_T$, where C'_T is the total energy budget in the absence of the trait. Note that the view here is that C_T is the sum of direct and opportunity costs, which includes the direct energetic expenditure of ATP on biosynthesis and the indirect diversion of energy-containing carbon skeletons (opportunity loss), as would be reflected in the consumption of glucose molecules consumed in a chemostat and their conversion to ATP equivalents.

Owing to the additional amount of resource acquisition necessary to complete the cell cycle, investment in the trait, c_T, is expected to increase the cell-division time by a factor equal to the investment ratio c_T/C_T (ignoring for the time being any direct advantages conferred by the trait). The baseline fitness effects of building and maintaining a trait then impose a negative selection disadvantage $\simeq -\ln(2)(c_T/C_T)$, with $\ln(2) \simeq 0.69$ simply scaling the rate to the continuous-growth scale (Foundations

17.1).

One caveat with respect to this definition is that it assumes that the addition of the trait does not somehow alter the cell's basic metabolic makeup in other ways that would modify the total baseline energy budget C_T. In some instances, there may be nonadditive side effects associated with a trait, as for example, when a novel protein promiscuously interacts with inappropriate substrates, aggregates with other cellular components, and/or excessively occupies cellular volume or membrane real estate. However, even if this does occur, the result given in Foundations 17.1 will be only slightly modified if the fractional alteration to C_T is small, which seems likely for cellular modifications involving just one or two genes. The general point then remains – as there will always be some baseline cost of expressing a trait, the net benefit of a trait needs to be derived by subtracting the construction/maintenance costs from the direct benefits accrued from increased survival and/or reproduction (Figure 17.2).

It further follows that if a trait is to be maintained by natural selection, it must pay for itself in terms of fitness enhancement – the baseline costs must be sufficiently small that the net benefit s_n is greater than the power of random genetic drift ($1/N_e$ for a haploid, and $1/2N_e$ for a diploid population). If this condition is not met, an existing functional trait will generally be vulnerable to loss by degenerative mutations. However, it also follows that in this domain of effective neutrality ($|s| < 1/N_e$), mutation pressure can lead to the slow accumulation of cellular modifications that impose a net drain on a cell's energetic capacity, as long as the incremental deleterious fitness effects are $< 1/N_e$. This is discussed below in the context of alterations to genome architecture and gene structure in various eukaryotic lineages.

Biosynthetic Costs of Nucleotides and Amino Acids

The majority of the operational infrastructure of cells consists of DNA, RNA, and protein, and although there are many energetic costs associated with assembly, processing, and maintenance of these molecules, the primary costs are associated with construction of the monomeric building blocks of the polymeric chains. Because the biosynthetic pathways for nucleotides and amino acids are highly conserved across the Tree of Life (Chapter 19), it is relatively straight-forward, albeit tedious, to estimate the direct and opportunity costs of these basic units (Foundations 17.2). As discussed above, such costs are quantified in terms of ATP usage, specifically the number of phosphorus atoms released via ATP hydrolyses, the primary source of energy in most cellular reactions. Biosynthetic pathway steps not involving ATP, but relying on different reactions including electron transfers resulting from coenzyme conversions (e.g., NADH to $NADH^+$, NADPH to $NADPH^+$, and $FADH_2$ to FAD) can be converted to ATP equivalents using conventions in biochemistry based on the known pathway for ATP production.

As a first-order approximation, the average opportunity and direct costs of building the four ribonucleotides (used in RNA molecules) are $\simeq$ 43 and 6 ATPs, respectively, with the total costs ranging from 43 for UTP to 55 for ATP. Deoxyribonucleotides (used in DNA), which are made from the former by ribonucleotide reductase (and thymidylate synthase for U $\rightarrow$ T), have the same opportunity costs,

but average direct costs elevated to 8 ATPs. Although the opportunity costs for both A:T and C:G bonds in double-stranded DNA are 85.5 ATP equivalents, the direct costs of the former are higher than the latter (19 vs. 17 ATPs), so that the total cost of an A:T bond is elevated by $\sim 2\%$.

The average opportunity and direct costs per amino acid are ~ 24 and 6 ATPs (Foundations 17.2), with the total average cost per residue being $\sim 61\%$ of the cost of an average nucleotide. However, contrary to the situation with nucleotides, all of which have similar biosynthetic costs, the total costs for different amino acids range six-fold, from 12 ATPs for glycine to 71 for tryptophan, in strong correlation with the molecular weights of the individual residues (Seligmann et al. 2003). It has been suggested that these cost differences among amino acids may be sufficiently high to be perceived by natural selection (Akashi and Gojobori 2002), and averaged across the Tree of Life, there is indeed a nearly 30-fold reduction in the use of the most expensive relative to the cheapest amino acids (Krick et al. 2014). However, for selection to be sufficiently strong to drive differences in amino-acid utilization, the cost differential between alternative amino acids at single codon sites must be sufficiently large (relative to the cell's entire energy budget) to overcome the background noise associated with random genetic drift.

Why the focus on amino acids at single sites as units of selection rather than the full genome content? For selection to be effective at discriminating differences in the amino-acid contents of stretches of sequence longer than single codons, multiple codons for expensive types of amino-acid variants would have to be simultaneously linked within chromosomal segments, and recombination among sites would have to be sufficiently rare for such a segment to be a reliable target of selection on bioenergetic content. Given that most populations have levels of heterozygosity at silent sites < 0.02, that such levels are typically at least 50% smaller for amino-acid replacement sites, and that $\sim 75\%$ of nucleotide sites within a gene are replacement sites (Chapters 4 and 12), then for a gene of average length (~ 1000 bp), the number of amino-acid polymorphisms segregating simultaneously will generally be < 5, many of which by chance will involve amino acids with small cost differences. Moreover, because the rate of recombination per nucleotide site is roughly equal to that of mutation (Chapter 4), considerable decoupling of segregating amino-acid substitutions will occur over time. Thus, whether natural selection is capable of perceiving energetic-cost differences among amino acids merits further consideration, noting that the maximum cost differential at a particular amino-acid site is $(71 - 12) = 59$ ATP equivalents.

Consider first the situation for an *E. coli*-sized cell with volume 1 μm^3, which requires $\sim 3 \times 10^{10}$ ATP hydrolyses to build (Chapter 8). The mean number of protein molecules associated with an average gene in a cell of this size is ~ 1700, with almost all genes falling in the range of 10 to 10^4 protein copies/cell (Chapter 7). Thus, relative to the total cell budget, the maximum energetic impact of the substitution of a single amino-acid residue (assuming a highly expressed gene with 10^4 protein copies/cell) $\simeq (59 \times 10^4)/(3 \times 10^{10}) \simeq 2 \times 10^{-5}$. As this quantity times $\ln(2)$ translates into a selection coefficient (Foundations 17.1), then recalling that selection is effective in a haploid population if $2N_e s > 1$ (Chapter 8), the effective population size N_e need only exceed $\sim 4 \times 10^4$ for selection to promote this extreme an amino-acid change in a highly expressed gene. For a gene with just 10 protein

copies per cell, the critical N_e increases to 4×10^7, which is near the typical N_e for microbial populations.

This rough analysis suggests that the selective promotion of amino acids with low biosynthetic costs (assuming they do not compromise protein function) can indeed be quite effective in prokaryotes, although with diminishing strength in lowly expressed genes. Consistent with these arguments, Akashi and Gojobori (2002) found a nearly 10% decline in the average cost per amino-acid residue in proteins with increasing gene-expression levels in *E. coli* and *B. subtilis,* and a similar conclusion was reached for other bacterial species (Heizer et al. 2006; Raiford et al. 2012). Thus, at least in bacteria, amino-acid substitutions that may be neutral with respect to protein function may nonetheless be advanced via their relative metabolic demands.

Now consider a yeast-sized cell, ~ 100 μm^3 in volume, with a range of 100 to 10^6 protein copies per gene (Chapter 7). Given the near linear scaling of lifetime cellular energy budgets with cell volume (Chapter 8), the construction cost for this larger cell is $\sim 100\times$ that for a typical bacterium, but so is the upper limit for protein number. Thus, assuming similar protein length, the upper limit to the relative cellular expense (c_T/C_T) is the same as for *E. coli,* whereas the lower limit is $\sim 10\times$ smaller. The critical N_e values then become 4×10^4 to 4×10^8 for highly vs. lowly expressed genes, with the latter being near the upper bound seen in unicellular eukaryotes (Chapter 4). Similar calculations for a metazoan cell size of ~ 1000 μm^3 in volume yield a range of critical N_e values of 4×10^4 to 4×10^9 for the most highly to most lowly expressed genes. The latter critical point is orders of magnitude greater than what is observed in multicellular species.

Consistent with these disparities, indirect analyses suggest that selection may promote energetically cheap amino acids in highly expressed genes in unicellular eukaryotes, and perhaps even in some genes in multicellular eukaryotes (Swire 2007; Heizer et al. 2011). However, the overall pattern is substantially weaker than in bacteria – only a 1% reduction in the average amino-acid cost in highly vs. lowly expressed genes in *S. cerevisiae* (Raiford et al. 2008), and 3% in the flour beetle *Tribolium castaneum* (Williford and Demuth 2012).

These observations suggest that in eukaryotes, selection for usage of energetically cheap amino acids approaches effective neutrality for a large fraction of genes with lower expression, and increasingly so in organisms that are larger in size. Even this, however, is a highly conservative conclusion, in that the preceding computations were carried out with the most extreme bioenergetic cost difference between amino acids. All but five amino acids have ATP costs in the range of 12 to 36 ATPs (Foundations 17.2). Amino-acid cost differences on the order of 6 ATPs (as opposed to 61) require effective population sizes to be 10-fold higher than those noted above for selection to be efficient, greatly reducing the likelihood of selective promotion of amino acids based on their energetic demands in species with large cell sizes. Thus, biased amino-acid usage in species with insufficient N_e to enable selection based on cost differences must have alternative explanation, such as mutation bias or simple functional constraints. Notably, in *E. coli* and yeast, amino-acid sites containing expensive residues tend to evolve more slowly than those harbored by the cheapest residues, suggesting that in these relatively high-N_e species, energetically costly residues are primarily relied upon for key structural or functional reasons (Seligmann et al. 2003; Barton et al. 2010).

A further issue of concern is the nature of amino-acid acquisition. Although it is often the case that most (and in some cases all) amino acids are synthesized within the cell, many species (e.g., metazoans) are unable to synthesize one or more amino acids and must acquire them from external food sources. Thus, it is of interest that the patterns of reduced usage of expensive amino acids noted above also occur in microbes that are auxotrophic (unable to synthesize) for such residues (Swire 2007; Raiford et al. 2012). Although there is no direct cost of biosynthesis of externally acquired amino acids, Swire (2007) argues that there is still an opportunity cost – an amino acid taken up by a heterotroph can either be directly incorporated into a protein or degraded to produce ATP that can be utilized in other cellular processes. Thus, given that the amount of energy extracted from the breakdown of an amino acid is about the same as the energy for building one, that most of the total cost of amino-acid biosynthesis involves opportunity loss, and that energy must be expended for amino-acid uptake, the differences in cost for directly acquired vs. internally synthesized amino acids may be relatively minor. Resolving this matter is of interest as the evolutionary loss of a biosynthetic pathway is expected to occur when the payoffs of direct biosynthesis are sufficiently small relative to the cost of the constructing and maintaining the pathway.

An Empirical Shortcut to Cost Estimates

There are two major challenges of applying the preceding theory to nonmodel organisms. First, for those who have worked through the book-keeping contortions in Foundations 1.2 and 1.5 at the close of this chapter, the complexity of estimating the costs of various cellular components from known biosynthetic pathways will be apparent. Moreover, although many such pathways are highly conserved cross the Tree of Life, variants do exist within and among species, and for the vast majority of organisms, the precise nature of the underlying reactions may be uncertain even for the simplest of building blocks. In addition, for the wide variety of biological compounds for which biosynthetic mechanisms are completely unknown, pathway analysis is not an option.

Second, as outlined in the preceding section, to make a connection to evolutionary theory, all costs of cellular components need to be normalized by the entire cellular energy budget. Effective methods for estimating the latter, outlined in Chapter 8, have been applied to a wide variety of microbes, leading to a general expression for the cost of building and maintaining a cell as a function of cell volume. However, the methods involved are not broadly utilizable, as they require the culturing of organisms on a defined medium in chemostats at a wide range of growth rates. For long-lived organisms, or those that consume other organisms with unknown chemical compositions, implementation of such a strategy is highly impractical.

Thus, there is a need for approximate methods that can be applied in the absence of detailed knowledge of an organism's biochemical pathways or growth features. One possible approach invokes the Kharasch and Sher (1925) formula for the degree of reduction of an organic compound, N_E as defined in Equation 7.14, as this is almost perfectly correlated with the heats of combustion of a wide range

of organic compounds and hence provides a measure of the energy content of a substance. N_E is a simple function of the number of carbon, hydrogen, and oxygen atoms in a compound, and from the standpoint of the costs of building blocks, a key missing element would seem to be phosphorus, which enters into the energetic values of a number of precursor compounds, nucleic acids, and phospholipids, etc. However, modification of the expression for N_E to

$$N_{\rm ATP} = 4N_C + N_H - 2N_O + 10N_P, \tag{17.2}$$

where the N terms to the right are numbers of the subscripted element in the molecular formula for the compound, provides an excellent first-order approximation to the total costs for a wide range of cellular compounds (Figure 17.3). This means that, from the standpoint of cell biology, the total cost of an organic substance, in terms of ATP equivalents, can be closely approximated using only the atomic numbers in the chemical formula. In principle, it also implies that as a first-order approximation, the total cost of building a cell, C_T, might be measurable using simple information on the elemental composition for C, H, O, and P alone.

To gain some appreciation for the potential utility of this approach, recall that chemostat results suggest that the cost of constructing a cell in a species with average volume 1 μm^3 $\simeq 2.7 \times 10^{10}$ ATPs (Chapter 8). For a cell with this average size, the expected newborn size is 0.67 μm^3, which then grows to size 1.33 μm^3 at the time of cell division. From Equation 7.1, a volume increase of 0.67 μm^3 is equivalent to a dry weight increase of $\simeq 0.00039$ ng, approximately 50% of which is carbon (Chapter 7). Noting that carbon has a molecular weight of 12 g, and applying Avogadro's number for the number of molecules / mol leads to $N_C \simeq 10^{10}$ atoms of carbon required for the construction of a newborn cell. Averaging over elemental analyses for two bacteria and two yeasts, Duboc et al. (1985) estimated C : H : O : P molar ratios of 1.00 : 1.73 : 0.57 : 0.02. Applying the elemental numbers to Equation 17.2 then leads to a total cost estimate of 4.7×10^{10}, which is $\sim 1.7\times$ the direct estimate.

Given the numerous sources of inaccuracies in both estimates, this level of disagreement should not be viewed as too serious, particularly because for most downstream evolutionary analyses, an estimate of C_T to order-of-magnitude accuracy is generally sufficient. Equation 17.2 is actually expected to underestimate more direct estimates of construction costs, as this indirect approach only estimates the ATP equivalents tied up in the actual standing biomass in a cell. Not included are the costs of acquisition of resource molecules (nutrient import through channels), of assembling building blocks into higher-order structures, of molding such structures into their appropriate forms, etc. However, as noted in Lynch and Marinov (2015, 2017) the sum of these additional costs is generally small relative to the costs of synthesizing the basic building blocks.

Despite its promising practical utility, a theoretical basis for Equation 17.2 remains to be developed. However, a crude understanding for why $N_{\rm ATP}$ provides a reasonable approximation to cell-structural costs derives from the fact that $N_E = 24$ for glucose, and that throughout it has been assumed that a glucose molecule has an energetic content equivalent to 30 ATPs (based on known biochemical pathways; Fundamentals 17.2). With the latter transformation, this suggests that at least at the whole-cell level, multiplication of the right side of Equation 17.2 by 1.25 should yield the predicted number of ATPs. If the cell were capable of extracting fewer than

30 ATPs per glucose molecule, the weighting factor would be reduced, becoming 1.0 for the case in which the glucose equivalents in ATPs is just 24 (and 0.5 in the case of just 12).

In prior attempts to draw connections like that noted above, Williams et al. (1987) ignored phosphorus, whereas Duboc et al. (1985) advocated a weighting factor of 5 based on arguments related to the fate of phosphorus in the burning of a P-containing substance. However, both approaches lead to poorer fits of the actual data for the cellular building blocks outlined in Figure 17.3, especially for precursor molecules, and the accounting for nitrogen and/or sulfur content does not improve the situation. Thus, for obtaining estimates of c_T for individual cellular components, the scaling applied in Equation 17.2 is clearly preferable. On the other hand, because phosphorus is such a small component of overall cellular biomass, none of these scaling factors has much impact in estimating the total cost of a cell, C_T; in the preceding example, ignoring N_P entirely only reduces the estimate of N_{ATP} by 3%.

The Energetic Cost of a Gene

As a first application of the preceding ideas, consider the total cost of a maintaining and operating a gene, which involves up to three levels of investment. Even for an unexpressed genome segment, there are DNA-level costs in terms of replication and chromosome maintenance (which in eukaryotes, include the cost of nucleosomes around which the DNA is wrapped). Transcription and transcript processing impose additional costs, and for protein-coding genes, there are still more costs in terms of amino-acid biosynthesis and polypeptide processing. The sum of investments at these three levels constitute the total cost of any genomic sequence.

Individual genes face a "use it or lose it" challenge. If the net fitness advantage of a gene is smaller than the power of random genetic drift, it will be vulnerable to passive inactivation by the accumulation of degenerative mutations in an effectively neutral fashion, and if there is a strong enough net selective disadvantage of the remnant pseudogene, physical removal will be accelerated by directional selection for deletion mutations. On the other hand, excess and even nonfunctional DNA can often accumulate in genomes by insertion mechanisms, as in the case of the expansion of mobile-genetic elements. Thus, biased mutation pressure alone can facilitate such genomic expansion provided the cost of the excess genomic material is smaller than the power of random genetic drift.

One of the mysteries of genome evolution concerns the number of genes contained within genomes and the mechanisms responsible for the lineage-specific expansions of such numbers in eukaryotes, especially in multicellular species (Lynch 2007a; Lynch et al. 2011; Chapter 24). The genomes of most prokaryotes contain < 5000 protein-coding genes, whereas most eukaryotes harbor $> 10^4$ genes, with the genomes of most multicellular eukaryotes (metazoans and land plants) containing 15,000 to 30,000. Most significantly, the expansion of total genome sizes from prokaryotes (generally 1 to 10 million base pairs) to unicellular eukaryotes (generally 10 to 100 million base pairs) to multicellular eukaryotes (hundreds to thousands of million base pairs) is much less a consequence of an increase in gene number than

of the proliferation of introns, mobile-genetic elements, and other sorts of genomic insertions. Most prokaryotic genomes are highly streamlined, typically containing $< 5\%$ intergenic DNA and devoid of introns and mobile-genetic elements, whereas the genomes of multicellular eukaryotes often contain $< 5\%$ coding DNA and harbor massive numbers of large introns, mobile elements, and other forms of DNA insertions.

This raises the question as to whether expansions in genome size and gene-architectural complexity are driven by adaptive processes as opposed to being inevitable consequences of the increased power of random genetic drift in organisms of larger size. A common view is that that there is an intrinsic advantage to both cellular complexity and multicellularity (Chapter 24), but such a stance is nothing more than an assumption (Lynch 2007b; Booth and Doolittle 2015), presumably fostered by one multicellular species having come to dominate the biological and intellectual world.

There is, however, no direct evidence that what we regard as complexity has been directly promoted by natural selection. If the advancement of complexity is a goal of selection, given that all extant organisms are temporally equidistant from the last universal common ancestor, the more astounding observation is the extreme phylogenetic rarity of complex multicellularity involving large numbers of cell types, which is represented by only two eukaryotic lineages, metazoans and land plants.

To help explain the gradient from extreme streamlining of genomes in prokaryotes to the extraordinary bloating of genomes in metazoans and land plants, we now draw from a wide variety of observations from cell biology and biochemistry to evaluate how expensive a gene (or segment of DNA) is from an energetic perspective. Through its phenotypic manifestations, a gene may have a multiplicity of advantages, but the energetic costs associated with replication, maintenance, and expression represent a minimum burden that must be overcome to achieve a net selective advantage large enough to ensure gene survival (Figure 17.2). The following subsections present first-order approximations of the cumulative costs of a gene at the genomic, transcriptional, and protein levels.

Chromosome-associated costs. We first consider the baseline cost of harboring a segment of DNA, regardless of its expression level. Genome replication requires the synthesis of two new DNA strands from each parental double-helix DNA molecule. From Foundations 17.2, the average total cost per deoxyribonucleotide is 52 ATPs. There are numerous additional costs of a gene at the DNA level, a major one being the unwinding of the parental double helix, which requires ~ 1 ATP per nucleotide. All other replication-related costs – opening of origins of replication, clamp loading, proofreading, production of the RNA primers used for replicate-strand extension, ligation of Okazaki fragments, and DNA repair – are an order of magnitude or so smaller. Thus, noting the double-stranded nature of DNA, as a first-order approximation, the total cost of replicating a gene L_n nucleotides in length is $\simeq 100L_n$.

There is, however, one additional major chromosome-level cost specific to eukaryotes – the highly ordered, dense coverage of nucleosomes, each of which contains two heterotetrameric histone complexes followed by a linker histone. Throughout eukaryotes, each nucleosome wraps ~ 147 bp, and with an average linker length between nucleosomes of 33 bp, there is on average one nucleosome per 180-bp in-

terval. Weighting by the cost of synthesizing the amino acids that comprise histone proteins and the cost of translating such proteins, the total nucleosome-associated cost $\simeq 190L_n$ ATPs (using slightly modified building costs, from Lynch and Marinov 2015), more than the DNA itself.

Taking all of the above issues into consideration, the total chromosome-level cost of a bacterial DNA segment (in units of ATP hydrolyses) is

$$c_{\mathrm{DNA},b} \simeq 100L_n, \tag{17.2a}$$

whereas for a haploid eukaryote,

$$c_{\mathrm{DNA},h} \simeq 290L_n, \tag{17.2b}$$

and doubling the preceding cost for a diploid eukaryote yields

$$c_{\mathrm{DNA},d} \simeq 580L_n. \tag{17.2c}$$

These results provide a quantitative basis for understanding the evolutionary mechanisms underlying the dramatic differences in gene structure and genomic architecture between prokaryotes and eukaryotes. Because replication is essentially a one-time investment in the life of a cell, the maximum fractional contribution of the DNA-level cost of a gene to a cell's total energy budget occurs at minimum cell-division times, and because maintenance costs are proportionally small under such conditions (Chapter 8), the former can be approximated as c_{DNA}/C_G, where C_G is total cost of constructing a cell. Thus, a prokaryotic cell with a representative volume of 1 μm^3 (and associated cellular construction cost of $\sim 3\times10^{10}$ ATPs) has a replication-associated cost of DNA $\simeq (3\times10^{-9})L_n$, which implies a fractional drain on the total cellular energy budget of (3×10^{-9}) for a 1-bp insertion and (3×10^{-6}) for a gene-sized insertion of 1000 bp. Thus, because free-living prokaryotes typically have effective population sizes $\sim 10^8$ (Chapter 4), when growing at maximum rates, such organisms experience efficient enough selection to remove insertions as small as 10 bp (and even 1-bp when N_e approaches the apparent upper bound of 10^9).

In contrast, for a unicellular eukaryote with a moderate-sized 100 μm^3 cell containing a haploid genome (e.g., a yeast), the fractional cost of DNA is $\simeq 10^{-10}L_n$, yielding maximum relative chromosome-level costs of 10^{-9} and 10^{-7} for 10- and 1000-bp segments of DNA, respectively. Because unicellular eukaryotes often have $N_e < 10^8$, sometimes ranging down to 10^6, these results imply that insertions of small to moderate size in such species will frequently be undiscernible by natural selection based on energetic effects alone.

For a larger cell size of 2500 μm^3, more typical of a multicellular eukaryote, and a diploid genome, the relative cost of DNA declines to $\simeq 10^{-11}L_n$, so even a 10^5-bp segment of DNA has a relative cost of just 10^{-6}. The effective population sizes of invertebrates tend to be in the neighborhood of 10^6, with that of some vertebrates (including humans, historically) ranging down to 10^4. Thus, even though the chromosome-level cost of a DNA insertion in a diploid multicellular eukaryote is $\sim 5\times$ that in a prokaryote, the disparity in total cellular energy budgets dilutes the effect, such that the power of random genetic drift is sufficient to overwhelm the ability of selection to prevent the accumulation of quite large insertions on the basis of energetic costs at the chromosome level.

These results provide a mechanistic explanation for the highly streamlined genomes of prokaryotes relative to eukaryotes. As outlined in Foundations 17.3, however, there is an additional cost of excess DNA, unassociated with bioenergetics – all excess DNA, even when nonfunctional, is dangerous in that it increases the substrate for mutations to gene malfunction (Lynch 2007a). On a per-nucleotide basis, this too is typically a weak evolutionary cost, strong enough to be perceived in many microbes but often effectively neutral in eukaryotes (multicellular species in particular).

Transcription-associated costs. Although the costs of transcription are numerous, and not all of them can be quantified with certainty, the major contributors are well understood. Thus, it is again possible to achieve order-of-magnitude estimates of the investments required to produce individual transcripts, adhering to the strategy summarized in Lynch and Marinov (2015).

Because transcripts are typically degraded (and must be replaced) on time scales much shorter than cell-division intervals, the total cost of transcription per cell cycle depends on the lifespan of a cell (T). If we consider a cell containing an average number of transcripts N_r at birth and a degradation rate per transcript of δ_r, during its lifetime, a cell must produce N_r additional surviving transcripts to create a daughter cell at the same steady-state level as well as an additional $\delta_r N_r T/\ln(2)$ replacement molecules to offset molecular degradation (Foundations 17.4). Here, it will be assumed that a set of N_r transcripts necessary to provision the equivalent of a daughter cell require *de novo* synthesis of nucleotides, whereas the remaining $\delta_r N_r T/\ln(2)$ replacement molecules are produced from recycled ribonucleotides.

Several forms of transcription-associated costs are general across prokaryotes and eukaryotes, but only two of these are quantitatively significant enough to be of concern here. The primary investment is the synthesis of ribonucleotides, with an average cost of ∼45 ATP per base (Foundations 17.2). Adding in the 2 ATP equivalnets required for each chain-elongation step, the total cost of *de novo* ribonucleotide synthesis associated with a gene with transcript length L_r is then $\simeq 47 N_r L_r$. The second major cost involves the replacement of degraded transcripts within the lifespan of the cell, and here the total expenditure is simply taken to be the two ATPs that must be expended per nucleotide for each chain-elongation step, which leads to a cost of $2\delta_r N_r L_r T/\ln(2)$.

A third cost, associated with helix unwinding, is $< 5\%$ of that associated with ribonucleotide recycling; and a fourth cost is associated with aborted transcripts (as not all transcription-initiation events lead to completed transcripts), but because such events generally occur within the first ten nucleotides, this cost is even smaller than that for helix unwinding. Still smaller is the cost of activating and initiating transcription. Thus, to a close approximation, the sum of the two predominant costs of transcription, expenditure on ribonucleotide synthesis and chain elongation, closely approximate the total cost of transcribing a gene in the lifetime of a bacterial cell (in units of ATP),

$$c_{\mathrm{RNA},b} \simeq N_r L_r (50 + 2.9\delta_r T). \tag{17.3a}$$

Several additional energy-consuming features of transcription are incurred by eukaryotes, but only two of them are quantitatively relevant here. First, eukaryotic

mRNAs are terminated by extended poly(A) tails, with initial lengths of ~ 250 nucleotides. Taking into consideration the full costs of As necessary for the standing pool of mRNAs and the two ATPs per bp necessary for chain elongation associated with excess degraded transcripts, the total cost of poly(A) tails is $\sim 250N_r(45 + 2.9\delta_r T)$. Second, as noted above, eukaryotic DNA is populated by regularly spaced nucleosomes, and in order for RNA polymerases to proceed, the DNA must be unwrapped from these, and this and other related processing entails a total energetic cost $\simeq 0.25N_r L_r \delta_r T$. For intron-containing genes, there is a small additional cost of splicing, and there are also costs associated with transcript termination, $5'$ mRNA capping, phosphorylation cycles associated with the RNA polymerase II, and nuclear export, but all of these are quite small relative to the two costs noted above.

Summing the eukaryotic-specific components with Equation 17.3a, the total cost associated with transcription for a eukaryotic gene is

$$c_{\mathrm{RNA},e} \simeq N_r[(12,500 + 50L_r) + (725 + 3.2L_r)\delta_r T], \tag{17.3b}$$

where L_r is the length of the primary transcript (before splicing). Note that in Equations 17.3a,b the total cost associated with transcription is subdivided into two components, the first defining the baseline requirement for building a cell, and the second being a linear function of the cell-division time.

Observations from single-cell methodologies provide quantitative insight into some of the key parameters in these formulations. As noted in Chapter 7, standing numbers of transcripts per gene (N_r) are generally quite small. For example, for *E. coli,* average $N_r \simeq 5.0$ (with a range of 0 to 100 among genes) (Lu et al. 2007; Li and Xie 2011). The mean N_r is 10 per gene in *S. cerevisiae* (Lu et al. 2007; Zenklusen et al. 2008), and the median is ~ 20 in mammalian cells (Islam et al. 2011; Schwanhäusser et al. 2011; Marinov et al. 2014). In all cases, there is a broad distribution around the mean, so that genes with numbers of transcripts deviating 10-fold from the mean are not uncommon (Golding et al. 2005; Raj et al. 2006; Taniguchi et al. 2010; Csárdi et al. 2015).

Estimates of transcript-decay rates suggest that the half-lives of mRNAs are typically much shorter than cell-division times. In *E. coli,* $\sim 80\%$ of mRNAs have decay rates (δ_r) in the range of 7 to 20/hour, with a median of 12/hour (Bernstein et al. 2002; Taniguchi et al. 2010). The mRNAs in *Bacillus subtilis* have a median decay rates of 8/hour (Hambraeus et al. 2002); and in *Lactococcus lactis,* mean and median mRNA decay rates are in the range of 3 to 7/hour, decreasing with decreasing cellular growth rates (Dressaire et al. 2013). For eukaryotes, median mRNA decay rates range from 3 to 6/hour in *S. cerevisiae* (Wang et al. 2002; Neymotin et al. 2014), and average 0.1/hour in mouse fibroblast cells (Schwanhäusser et al. 2011). To a first-order approximation, these results suggest that $\delta_r T$ is generally within the range of 10 to 100, which from Equations 17.3a,b further implies that mRNA decay typically inflates the total cost of transcription by a factor 2 to $6\times$ that expected on the basis of the steady-state number of transcripts per cell, $50N_r L_r$ and $N_r[(12,500 + 50L_r)]$, for bacteria and eukaryotes, respectively.

By comparison with the preceding results for chromosome-level costs, it can be seen that the costs at the level of transcription will often be several fold higher. Considering microbes for example, and noting that the length of a transcript is very close to the length of a gene ($L_n \simeq L_r$), the ratio of Equations 17.3a and

17.2a, $N_r(0.5 + 0.029\delta_r T)$, shows that the ratio of these two costs always exceeds 1.0, provided the steady-state number of transcripts is > 2, and can be several-fold higher for genes with higher expression levels and/or multiple transcript half-lives per cell cycle.

For the yeast *S. cerevisiae,* $> 95\%$ of genes are intron-free, and the remaining 5% contain only a single small intron, so as a first-order approximation, it can again be assumed that genes and transcripts have essentially the same lengths $L_n \simeq L_r = 2000$ bp. As noted above, the mean number of mRNAs per gene per cell is $\overline{N}_r \simeq 10$, and with an average decay rate of 4.5/hour and a doubling time $\simeq 1.5$ hours under optimal growth conditions $\delta_r T \simeq 7$. From Equation (17.3b), the cost of transcription for a typical yeast gene is then on the order of $10 \cdot \{(12,500+100,000)+[(725+6,400)\cdot 7]\} \simeq 1.6 \times 10^6$ ATPs. By contrast, from Equation 17.2b, the chromosome-level cost of a gene in this species $\simeq 0.6 \times 10^6$ ATPs.

Finally, we consider the situation for a typical human gene, where the median number of mRNAs per gene $\simeq 20$, and the average mRNA decay rate is ~ 1.4/day. Assuming a cell-division time of one day and an average primary transcript length of 47 kilobases (owing to the large burden of introns, which are transcribed before being spliced), and ignoring the small contribution from the cost of splicing ~ 7 introns per gene, the cost of transcription per gene is then on the order of $20 \cdot \{(12,500 + 2,350,000) + [(725 + 150,400) \cdot 1.4]\} \simeq 5 \times 10^7$ ATPs. Total gene lengths are difficult to define in metazoans, but a 50% inflation relative to the pre-mRNA (~ 70 kilobases) is not unreasonable. Equation 17.2c then implies a typical cost at the chromosome level of 4×10^7 ATPs. Taken together, all of these results suggest that transcription-associated costs in mammalian cells are typically of the same order of magnitude as those at the chromosome level, although the former can greatly exceed the latter for highly expressed genes.

Translation-associated costs. The conceptual approach employed in the preceding section can be extended to the protein level by again assuming that the cost of production of the steady state number of proteins must be covered by *de novo* synthesis of amino acids, with the excess molecules needed to compensate for protein decay being acquired from salvaged amino acids. Although several sources of costs underlie protein production and subsequent management, the overwhelming contributions are associated with just three functions, the biochemical details of which are summarized in Lynch and Marinov (2015).

First, the cost associated with the production of the standing level of protein for a particular gene necessary for an offspring cell is $N_p L_p \bar{c}_{\rm AA}$, where N_p is the number of protein molecules per newborn cell, L_p is the number of amino acids per protein, and $\bar{c}_{\rm AA}$ is the average total cost of synthesis per amino-acid residue (assumed to be equivalent to 30 ATP hydrolyses, based on Foundations 17.2). Second, the total cost associated with chain elongation of all proteins produced during the cell cycle is $4N_pL_p[1+(\delta_p T/\ln(2))]$, where δ_p is the rate of protein decay, and the 4 results from the 2 ATPs required for activating the cognate tRNA, an additional 1 for transferring the tRNA to the ribosome, and 1 more for the movement of the ribosome to an adjacent mRNA triplet). Third, degradation of proteins imposes an approximate cost of $N_pL_p\delta_pT/\ln(2)$ ATP hydrolyses. Additional costs small enough be ignored are associated with translation initiation and termination, post-translational mod-

ification, and protein folding. Summing up the three primary expenses, the total protein-level cost of a gene in both bacteria and eukaryotes is

$$c_{\rm PRO} \simeq N_p L_p (34 + 7\delta_p T), \tag{17.4}$$

where again the first term represents a one-time cost incurred regardless of the length of the cell cycle, and the second term grows linearly with the cell-division time owing to the cumulative costs of protein turnover and replacement.

Insight into the relative magnitudes of the two terms in Equation 17.4 requires information on protein-degradation rates, which can be obtained from results from high-throughput proteomics. Most notably, the decay rates of proteins are typically much lower than those of their cognate mRNAs. In the bacterium *Lactococcus lactis,* the vast majority of protein decay rates are in the range of 0.04 to 6.0/hour, with the median being 0.1 to 0.9/hour depending on the growth rate (Lahtvee et al. 2014), and those for other bacteria are commonly in the range of 0.05 to 0.20/hour (Trötschel et al. 2013). In *S. cerevisiae,* the median and mean decay rate is ~ 1.4/hour, with most values for individual proteins falling in the range of 0.2 to 5.5/hour under optimal growth conditions (Belle et al. 2006), and the median declining to 0.1/hour in nutrient limiting conditions (Shahrezaei and Swain 2008; Helbig et al. 2011). In mouse fibroblast cells, the median decay rate of a protein is ~ 0.02/hour (with a range of 0.002 to 0.3/hour) (Schwanhäusser et al. 2011), and in a human cancer cell line, the range is from 0.04 to 1.3/hour (Eden et al. 2011). Given the known division times for the cell types noted above, $\delta_p T$ will generally be < 1, and seldom > 10. This implies that the second term in Equation 17.4, the cost of protein degradation, will generally be of the same order of magnitude of the *de novo* protein synthesis or smaller.

Cellular abundances of proteins (N_p) are much higher than those for their cognate mRNAs, with the average ratio of the two per gene being 450 in *E. coli,* 5100 in *S. cerevisiae,* and 2800 in mammalian fibroblasts (Ghaemmaghami et al. 2003; Lu et al. 2007; Schwanhäusser et al. 2011). How do these numbers translate into the total protein-level cost of a gene? To keep the computations simple but still accurate to a first-order approximation, it will be assumed here that $7\delta_p T \simeq 6$, so that $c_{\rm PRO} \simeq 40 N_p L_p$.

In *E. coli,* the average number of proteins per gene is ~ 2250, and the average protein contains ~ 300 residues, implying an average $c_{\rm PRO} \simeq 3 \times 10^7$ ATP/protein-coding gene. By comparison, the average chromosome-level cost is $\sim 10^5$, and as noted above, the average transcription-associated cost is only a few fold greater than 10^5. Thus, the vast majority of the energetic cost of a protein-coding gene in bacteria is associated with translation. In the case of *S. cerevisiae,* there is an average of $\sim 50,000$ proteins per genetic locus per cell, and the average protein length is $\sim 50\%$ greater than in *E. coli,* yielding an average total cost of translation of $\sim 8 \times 10^8$ ATP per protein-coding gene, which is again approximately two orders-of-magnitude greater than the summed costs at the chromosome and transcription levels.

Evolutionary implications. Although the full slate of data necessary to estimate the total cost of a gene are only available for a few species (Lynch and Marinov 2015),

these are fully in accord with the hypothesis that baseline selective consequences (s_c) of such costs tend to exceed the power of random genetic drift in microbes and then progressively become smaller than the power of genetic drift in larger eukaryotic species (Figure 17.4).

For almost all genes in the bacterium *E. coli,* s_c falls in the range of 10^{-6} to 10^{-3}, far above the likely minimum values that can be perceived by selection in this large-N_e species (Figure 17.4). If such genes were to find themselves in an environment where their functions were no longer useful, inactivating mutations would be strongly selected for. Within eukaryotes, small peaks of lowly expressed genes exist with roughly the same absolute costs of *E. coli* genes. However, owing to the increased total cellular energy budgets, s_c for many eukaryotic genes falls below 10^{-6} and in some cases to as low as 10^{-9}. This is significant because the reduction in the effective population sizes of such species increases the power of random genetic drift. For the majority of genes in the eukaryotic species *S. cerevisiae* (yeast), *C. elegans* (nematode), and *A. thaliana* (land plant) the costs at the chromosomal and transcriptional levels are below or near the drift barrier, implying that without translation most gene sized insertions will be essentially invisible to the eyes of selection. The major contribution that pushes s_c of some genes past the drift barrier in eukaryotes is the cost of translation.

A more general analysis over a large number of species indicates that the total cost of a gene (relative to the cell energy budget) declines with increasing cell size across the Tree of Life (Figure 17.5). Average estimates of all three cost measures in bacteria are generally substantially greater than those in eukaryotes, although there is continuity in the scaling between groups. In addition, as noted above, there is a consistent ranking of $s_{\mathrm{DNA}} < s_{\mathrm{RNA}} < s_{\mathrm{PRO}}$, with a one to two order-of-magnitude increase from the former to the latter.

These results suggest that by reducing the contribution of single genes to a cell's total energy budget, evolutionary increases in cell size (and the associated increased power of random genetic drift in larger organisms) promote a shift in the selective environment such that gene-sized insertions in eukaryotes, particularly in multicellular species, are commonly effectively neutral from a bioenergetic perspective. The energetic cost of a DNA segment of even just a few nucleotides (even if nontranscribed) can be perceived by selection in a typical bacterial population with $N_e \simeq 10^8$. In contrast, insertions of even thousands of kb often impose a small enough energetic burden relative to the overall requirements of eukaryotic cells to be immune to selection.

Although costs at the RNA level are frequently greater than those of at the DNA level, these are often still not large enough to overcome the power of random genetic drift in eukaryotic cells. This means that many nonfunctional DNAs that are inadvertently transcribed in eukaryotes (especially in multicellular species) still cannot be opposed by selection. On the other hand, with the cost at the protein level generally being much greater than that at the RNA level, segments of DNA that are translated can sometimes impose a large enough energetic costs to be susceptible to selection, even in multicellular species.

These observations are relevant to the idea that an enhanced ability to generate energy, made possible by the origin of the mitochondrion, was a prerequisite for the evolution of increased gene numbers, protein lengths, protein folds, protein-protein

interactions, and regulatory elements in eukaryotic cells (Lane and Martin 2010). As already noted in Chapter 8, there is no dichotomous break in the size-scaling of the metabolic properties in prokaryotic vs. eukaryotic cells, and here we see that increased cell size does not induce a condition in which gene addition becomes an increasing selective burden, but quite the contrary. Although the absolute cost of a gene does increase with cell size, in terms of the fractional contribution to a cell's energy budget, which ultimately determines whether selection can oppose genome expansion, the scaled cost of an average gene decreases at the DNA, RNA, and protein levels.

Thus, population-genetic arguments based on both the mutational-hazard hypothesis (Foundations 17.2) and on the observed features of cellular energetics lead to the conclusion that passive increases in genome size are expected to naturally arise in organisms with increased cell sizes (which, by correlation, have reduced effective population sizes). This supports the view that variation in the power of random genetic drift has played a central role in the passive historical diversification of genome size and possibly cellular architecture across the Tree of Life (Chapter 24).

Finally, it should be noted that genes may have costs beyond those noted above. For example, there may be associated opportunity costs with respect to transcription and translation, as RNA polymerases, tRNAs, and ribosomes must be deployed in gene expression, reducing their availability to service other genes. Aggregation of proteins, associated with misfolding, can reduce the operation of key cellular functions, etc. However, experimental work suggests that the predominant cost of genes is indeed associated with the biosynthesis of the elemental building blocks rather than with toxic problems associated with harmful misfolding and protein-protein misinteractions (Stoebel et al. 2008; Plata et al. 2010; Eguchi et al. 2018), with the quantitative effects being in reasonable accord with the numbers cited above (Tomala and Korona 2013; Adler et al. 2014). Nonetheless, high gratuitous expression of proteins can lead to significant misfolding problems that induce secondary biosynthesis costs associated with the up-regulation of chaperones (Geiler-Samerotte et al. 2011; Frumkin et al. 2017).

The Cost of Lipids and Membranes

The major disparities in cellular structure between prokaryotes and eukaryotes involve internal membrane-bound organelles in the latter, with functions including sequestration and gated access to the genomic material, vesicle transport of a multiplicity of cargoes, power production, and platforms for molecular assembly. As outlined in Chapter 15, numerous features of these complex membrane systems appear to have evolved by repeated rounds of gene duplication and divergence. Given that the rate of *de novo* gene duplication in prokaryotes is comparable to that in eukaryotes (Lynch 2007a), and that some prokaryotes with organelles and internal membranes do exist, why is the typical internal layout of almost all prokaryotes devoid of membranes?

One possibility is that the evolution of internal membranes is the null state, driven by mutational bias and basic biophysical forces (Ramadas and Thattai 2013;

Mani and Thattai 2016), but that the construction and maintenance of such embellishments is energetically expensive enough that incremental changes are thwarted by the efficiency of natural selection in prokaryotes, but impervious to selection in larger eukaryotic cells. This could then lead to the emergence of internal cellular complexity as a simple consequence of nearly neutral, drift-like processes rather than by direct promotion by positive selection. An informative analogy to this sort of scenario from an evolutionary genomics perspective is the passive expansion of genome size by effectively neutral insertion processes (Lynch 2007a).

Because eukaryotic cells are typically larger than those of prokaryotes, often substantially so, there is an increase in the absolute investment of lipids based on the cell membrane alone. However, given a constant shape, the surface area of a cell increases with the square of cell length, whereas the volume increases with the cube of length (Table 8.1), so the relative investment in the external membrane declines with cell volume. Nonetheless, with the additional investment in membrane-bound organelles in eukaryotes, a larger fraction of cellular biomass is allocated to lipids than in prokaryotes, although not enormously so – the mean fractional contribution of lipids to total dry weight is ~ 0.06 for bacteria, ~ 0.08 for yeast species, and ~ 0.15 for unicellular photosynthetic algae (Chapter 7).

To understand the considerable bioenergetic costs of membrane production, we require information on the numbers of lipid molecules required for membrane production over the life of the cell as well as the cost of biosynthesis of such molecules. Again, the total number of molecules of a particular type required in cell's lifetime can be subdivided into a fixed quantity, equivalent to the number of molecules that comprise a newborn cell, N_l, and a time-dependent maintenance quantity associated with molecular turnover, $\delta_l N_l T/\ln(2)$, where δ_l is the molecular decay rate per lipid molecule, and T is the cell-division time.

Owing to the absence of information on the rate of lipid-molecule turnover, it will be necessary here to rely on N_l as an estimate of the minimum lifetime requirement for lipids. However, as there is no evidence that membrane lipids are rapidly degraded, the bias of the resultant estimates is not expected to be large. N_l can be determined by dividing cellular membrane areas by the head-group areas of membrane lipids, most of which are within 10% of an average value of $a_l = 0.65$ nm^2 (Nagle and Tristram-Nagle 2000; Petrache et al. 2000; Kucerka et al. 2011). It is also essential to know the thickness of the bilayer (h), as this determines the areas of the inner and outer layers. The thickness of a bilayer is approximately twice the radius of the head-group area, which $\simeq 0.5$ nm in all cases, plus the total length of the internal hydrophobic tail domain (Lewis and Engelman 1983; Mitra et al. 2004), which $\simeq 3.0$ nm, and so sums to $h \simeq 4.0$ nm. There are slight increases in bilayer thickness with the length of the fatty-acid chain deployed (Rand and Parsegian 1989; Wieslander et al. 1995; Rawicz et al. 2000), as each single and double carbon-carbon bond adds ~ 0.15 and 0.13 nm, respectively.

To gain some appreciation for the number of lipid molecules per cell, now consider a spherical cell with radius r, which implies a surface area for the outer side of the bilayer of $4\pi r^2$, and $4\pi(r-h)^2$ for the inner layer. The summed area is then $4\pi[r^2+(r-h)^2]$, which $\simeq 8\pi r^2$ for $h \ll r$. Dividing by a_l gives the total number of lipid molecules in the bilayer. With $a_l = 0.65$ nm^2 and $h = 4.5$ nm, for a bacterial-sized cell with radius $r = 1000$ nm (1 μm), there are then $\sim 4 \times 10^7$ total lipid molecules in the

cell membrane. This increases to $\sim 4\times10^9$ for a 10 μm sphere and $\sim 4\times10^{11}$ for a 100 μm sphere. For Gram negative bacteria (which include *Escherichia, Pseudomonas, Helicobacter,* and *Salmonella,* there are two cell membranes, which roughly doubles these numbers. These values need to be discounted somewhat to account for the occupancy of membranes by proteins; so, for example, if 50% of the membrane surface is occupied by proteins, the preceding values would need to be diminished by 50%. Nonetheless, it remains clear that the number of lipids in the cell membrane alone is almost always greater than the number of nucleotides in a cell's genome, often by orders of magnitude.

Costs of individual molecules. We now consider the total (direct plus opportunity) costs of biosynthesizing individual lipid molecules, the details of which are presented in Foundations 17.5. Most cellular membranes are predominantly comprised of glycerophospholipids (Table 17.1), which despite containing a variety of head groups (e.g., glycerol, choline, serine, glycerol, and inositol), all have biosynthetic costs per molecule (in units of ATP hydrolyses) of

$$c_L \simeq 283 + [30\cdot(n_L-16)] + (5\cdot n_U), \tag{17.5a}$$

$$c_L \simeq 291 + [32\cdot(n_L-16)] + (5\cdot n_U), \tag{17.5b}$$

in bacteria and eukaryotes respectively, where n_L is the mean fatty-acid chain length (number of backbone carbons), and n_U is the mean number of unsaturated carbons per fatty-acid chain. Although variants on glycerophospholipids are utilized in a variety of species (Guschina and Harwood 2006; Geiger et al. 2010), these are structurally similar enough that the preceding expressions should still provide good first-order approximations. The direct costs, which ignore the opportunity loss of ATP-generating potential from the diversion of metabolic precursors, are

$$c'_L \simeq 88 + [10\cdot(n_L-16)] + (5\cdot n_U), \tag{17.6a}$$

$$c'_L \simeq 104 + [12\cdot(n_L-16)] + (5\cdot n_U), \tag{17.6b}$$

in bacteria and eukaryotes, respectively.

For most lipids in biological membranes, $14 \le n_L \le 22$, and $0 \le n_U \le 6$, so the total cost per lipid molecule is generally in the range of $c_L \simeq 220$ to 500 ATP, although the average over the pooled population of lipids deployed in species-specific membranes is much narrower. Cardiolipin, which rarely comprises more than 20% of membrane lipids is exceptional because it derives from a fusion of two phosphatidylglycerols, and has total and direct costs of $\sim$ 570 and 190 ATPs/molecule. Thus, on an individual molecule basis, lipids are an order of magnitude more expensive than the other two major monomeric building blocks of cells (nucleotides and amino acids).

Application of the preceding expressions to known membrane compositions indicates that the biosynthetic costs of eukaryotic lipids are somewhat higher than those in bacteria (Table 17.1), e.g., for a diversity of eukaryotic species, the average total cost per lipid molecule in the plasma membrane is $\sim$ 7% higher than in bacteria. The former estimate is also essentially the same as obtained for whole eukaryotic cells, although the average cost of mitochondrial lipids is especially high. These

elevated expenses in eukaryotes are a consequence of two factors: 1) the added cost of mitochondrial export of oxaloacetate to generate acetyl-CoA necessary for lipid biosynthesis in the cytoplasm; and 2) the tendency for eukaryotic lipids to have longer chains containing more desaturated carbons.

Table 17.1. Bioenergetic costs for the synthesis of lipid molecules. Data are provided for species with lipid composition measurements of the parameters needed to solve Equations 17.5a-17.6b, along with additional information on the contribution from cardiolipin. PL denotes glycerophospholipid, and C cardiolipin, with the cost for the remaining small fraction per molecule being taken to be the average of the preceding two. Total cost denotes the opportunity plus direct cost per molecule incorporated into the membrane. Mean costs are obtained by weighting the PL and cardiolipin costs by their fractional contributions. Standard deviations among species are given in parentheses. Modified from Lynch and Marinov (2017).

	PL Cost		Composition		Mean Cost	
Source	Total	Direct	PL	C	Total	Direct
Bacteria, whole cell	299 (22)	94 (8)	0.89 (0.09)	0.09 (0.06)	326 (14)	99 (6)
Euks., whole cell	326 (21)	124 (9)	0.95 (0.03)	0.04 (0.03)	346 (19)	128 (8)
Euks., plasma memb.	338 (16)	125 (7)	0.95 (0.05)	0.03 (0.03)	348 (19)	124 (7)
Euks., mitochondrion	345 (42)	129 (18)	0.85 (0.08)	0.11 (0.05)	376 (37)	134 (17)

Total cellular investment. We are now in position to estimate the total-cell bioenergetic cost associated with membrane lipids. Recalling the surface occupancy of individual lipid molecules and the bilayer nature of membranes, the biosynthetic cost of a membrane with surface area A (in units of μm^2, and ignoring the unknown contribution from lipid turnover) is

$$C_L \simeq 2A \cdot \bar{c}_L/(0.65 \times 10^{-6}), \tag{17.7}$$

where $\bar{c}_L$ is the average total cost of a lipid molecule (e.g., as given in Table 17.1). Dividing C_L by the total cost of building a progeny cell (Chapter 4) yields the proportion of a cell's total growth budget allocated to a membrane.

There are only a few cell types for which the internal anatomy has been scrutinized well enough to estimate the allocation to a cell's energy budget for the full set of membrane types. However, the data uniformly suggest that a substantial fraction ($\sim$ 20 to 60%) of a cell's growth budget is allocated to membrane lipids (Table 17.2). As expected, based on the surface area:volume relationship, the plasma membrane constitutes a diminishing cost with increasing cell size, from $>$ 16% in bacterial-sized cells with volumes $<$ 2 μm^2 to $<$ 7% in moderate-sized eukaryotic cells. The majority of the total cost of membrane production in eukaryotic cells is associated with internal membranes, e.g., a 7-fold inflation of the membrane budget of the green alga *Dunaliella,* leading to a substantial addition relative to that expected for a prokaryotic cell plan.

Enough information is available on the total investment in mitochondrial lipid membranes that a general statement can be made for this particular organelle. Over the eukaryotic domain, the total surface area of mitochondria (inner plus outer

membranes, summed over all mitochondria, in μm^2) scales with cell volume (V, in units of μm^3) as $A \simeq 3V$ (Figure 17.6a). Applying this to Equation 17.7, with the average total cost of mitochondrial lipids ($\simeq$ 376 ATP/ molecule; Table 17.1), and letting the total growth requirements of a cell be $(27 \times 10^9)V$ from Chapter 3, yields a measure of the relative cost of mitochondrial membrane lipids of 0.13, essentially independent of eukaryotic cell size.

These results indicate that the construction of mitochondrial membranes, critical to ATP synthase operation in eukaryotes, amounts to a $\sim$ 13% drain on the cellular energy budget beyond what would be necessary had ATP synthase remained in operation on the plasma membrane (with mitochondria being absent). These calculations are first-order approximations for rapidly growing cells, for which the contributions of cell maintenance and lipid-molecule turnover to the total cellular energy budget will be minor. For slowly growing cells, the costs will be higher or lower depending on whether the cost of mitochondrial-membrane maintenance is above or below that for total cellular maintenance. However, the central point remains – the costs of mitochondrial membranes represent a substantial baseline price, not incurred by prokaryotes, associated with relocating bioenergetics to the interior of eukaryotic cells.

Although the data are not as extensive, this approach can be extended to show that different cost scalings exist for other types of internal membranes. For example, across the Tree of Life, the outer surface area of the nuclear envelope scales $\simeq 2.7V^{0.5}$ (Figure 17.6b). Assuming the average cost of a lipid in these organelles to be about the same as for the entire cell ($\sim$ 346 ATPs, from Table 17.1), and recalling that the nuclear envelope has a double membrane, the total cost of the nuclear envelope (scaled to the total cell budget) is $0.1V^{-0.5}$. Thus, the fractional cost of the nuclear envelope (relative to a cell's total energy budget) declines with increasing cell volume, from $\sim$ 3% for a small eukaryote with $V = 10$ μm^3 to $\sim$ 0.3% when $V = 1000$. Although there are insufficient data to estimate scaling relationships for other organelle surface areas, multiple studies make clear that in all but the smallest eukaryotic cells, the summed contributions from membranes associated with the endoplasmic reticulum, golgi, and assorted vesicles and vacuoles exceed those associated with the nuclear envelope, often by more than tenfold (Table 17.2). Thus, adding in the mitochondrial investment, it appears that 10 and to 20% of a eukaryotic cell's energy budget is typically invested in internal membranes.

There are a number of other costs associated with membranes and their processing, but as in the case of nucleic acids and proteins, the major costs appear to be associated with the biosynthesis of the basic building blocks noted above. Consider, for example, the cost of molding membranes into specific shapes. The problem is most simply evaluated for spherical vesicles, for which the bending energy is ~ 400 K_BT. This energetic requirement is independent of vesicle size because although there is more surface area to bend in larger vesicles, the curvature is reduced, resulting in canceling of the two effects (Phillips et al. 2012). Knowing the rate of membrane flux for a cell then allows a rough estimate of the total bending energy required per unit time. For example, the entire cell membranes of mammalian fibroblast and macrophage cells are interiorized by pinocytosis in about 0.5 to 2.0 hours (Steinman et al. 1976). The same is true for *Dictyostelium* (slime mold) cells (Thilo and Vogel 1980). Knowing the surface area of the cell and the average surface

area of a vesicle, one can then estimate that about 1000 vesicles must be produced per minute. Assuming that the bending energy is acquired via ATP hydrolyzing reactions in the production of protein-coating cages (Chapter 15), and that an energy of $\sim 16\ K_BT$ is associated with each ATP hydrolysis, the total energy demand for bending $\simeq 400 \times 1000/16 = 25,000$ ATP/minute. The cost of membrane fusion is even smaller, being equivalent to $\sim 20\ K_BT$, or about one ATP hydrolysis, per fusion event (François-Martin et al. 2017). Using Equation 8.2a, for a *Dictyostelium*-sized cell with $V = 600\ \mu\text{m}^3$, the total bending energy associated with endocytosis is found to be quite small, $\simeq 0.001\%$ of the cell's total maintenance requirements.

Table 17.2. Contributions of membranes to total cellular growth costs (in units of ATP equivalents). Cell volumes (Vol) and total membrane surface areas (SA) are in units of μm^3 and μm^2, respectively. Fractional contributions to the total cell growth requirements are given for the plasma membrane (Pm), mitochondrial membranes (inner + outer, Mt), nuclear envelope (Nu), endoplasmic reticulum and golgi (ER/G), vesicles and vacuoles (V), and Total. The fraction of the total cell growth budget allocated to membranes is obtained by applying Equation 17.7, using the species-specific lipid biosynthesis costs (Table 17.1) where possible (and otherwise applying the averages for eukaryotic species), and normalizing by the allometric equation for ATP growth requirements given by Equation 8.2b. The SA given for *E. coli* is twice that of the cell, as this species has two membranes; the results for the two algae, *O. tauri* and *D. salina,* do not include the investment in plastid membranes. From Lynch and Marinov (2017).

			Fractional contributions to total cell growth:					
Organism	Vol	SA	Pm	Mt	Nu	ER/G	V	Total
Bacteria:								
Staphylococcus aureus	0.29	2.1	0.240					0.240
Escherichia coli	0.98	8.6	0.337					0.337
Bacillus subtilis	1.41	6.0	0.161					0.161
Eukaryotes:								
Ostreococcus tauri	0.9	14	0.364	0.030	0.149	0.033	0.036	0.612
Saccharomyces cerevisiae	44	211	0.066	0.061	0.034	0.022	0.023	0.206
Dunaliella salina	591	2326	0.028	0.035	0.014	0.065	0.065	0.207

Summary

- An understanding of the baseline energetic costs of constructing and maintaining cellular features is essential to determining whether certain aspects of a cell's biology are liable to accumulate by biased mutation pressure, despite being mildly deleterious, and also to evaluating the relative investments of total cellular budgets to alternative functions.

- The assessment of cell budgets in units of energy, rather than elemental constitution, is desirable because the latter contributes only to structural costs, and even then in not in all cases, and not to maintenance or operational activities. In

addition, most organic food substrates are more limiting with respect to energy than carbon content. The number of ATP hydrolyses provides a natural measure of energy expenditure, given that the energy carried by this molecule (and others closely related) is the universal currency of bioenergetics across the Tree of Life.

- The total energetic cost of any cellular feature can be viewed as the sum of three components: the direct costs consisting of investments in construction, operation, and maintenance; and opportunity costs that represent losses to other cellular functions owing to diversion of resources (e.g., carbon skeletons) to the trait of interest.

- The evolutionary fitness cost of a cellular feature, owing to baseline investment in construction and maintenance, exclusive of downstream phenotypic advantages, is directly proportional to the cost relative to the total cellular energy budget.

- The energetic costs of a gene (or genomic segment) entail three levels of investment involving construction and maintenance: chromosomal, transcriptional, and translational, with translation-related costs typically dominating the expenditures associated with an active gene.

- Quantification of the costs of constructing cellular constituents requires information on the biosynthetic costs of basic building blocks such as nucleotides (used in RNAs and DNA), amino acids (proteins), and fatty acids (membranes). Although individual amino acids are less energetically expensive than nucleotides, there is a six-fold range of variation among the twenty amino-acid types. For highly expressed protein-coding genes in species with large effective sizes (e.g., bacteria), such differences can be significant enough for selection to discriminate among alternative amino-acid types based on energetic considerations.

- Relative to the total energy budget of a cell, the cost of a genomic insertion of just a few bases can be sufficient to be resisted by natural selection in prokaryotic species with high N_e, whereas in larger eukaryotic species with relatively small N_e the proportional effects of insertions as large as several kilobases can be invisible to the eyes of selection. This, along with the weak mutational hazard of excess DNA, helps explain the streamlined genomes of microbes vs. the bloated genomes of multicellular eukaryotes.

- The half lives of transcripts are typically much shorter than the lifespan of a cell, so these must be continually replaced to keep the cell at steady state. Such recycling generally at least doubles the cost of transcription relative to the expectations with no degradation, and the overall cost of transcription of a gene is typically slightly more than the cost at the chromosomal level.

- Owing to the much larger number of proteins per cell, and despite their longer half lives, the average cost of proteins per gene per cell cycle is on the order of 100× that of transcripts.

- Although the absolute costs of genes increase with increasing organism size, the proportional increase relative to entire cell energy budgets declines. Because effective population sizes also decline with increasing organism size, this means that increased organism size results in an increased vulnerability to the passive accumulation of extraneous genomic material, as well as excess transcription and translation.

- The biosynthesis of membranes constitutes a substantial fraction of the total energy budgets of cells. The costs of individual lipid molecules are 6 to 8× that of individual nucleotides, and depending on the size of the cell, there will generally be on the order of 10^7 to 10^{12} lipid molecules in the cell membrane alone, typically well in excess of the number of nucleotides per genome.

- The hallmark of eukaryotic cells, individualized membrane-bound organelles, imposes a substantial increase in energetic cost per cell, relative to prokaryotes (once cell size is accounted for), such that 10 to 60% of the total energy budgets of eukaryotic cells are associated with membranes. Independent of cell volume, $\sim$ 13% of eukaryotic cell budgets are associated mitochondrial membrane lipids. On the other hand, owing to their small sizes, and consequent high surface-area:volume ratios, the total relative investment in bacteria membranes can be on the order of that for eukaryotic cells.

Foundations 17.1. The relationship of bioenergetic costs to the strength of selection. Understanding the baseline fitness consequences of the total energetic investment in a trait requires a definition of the construction and maintenance effects on the cell's reproductive rate (exclusive of any downstream changes in fitness owing to phenotypic effects). The selective cost associated with investment is defined as

$$s = r - r', \tag{17.1.1a}$$

where $r = \ln(2)/\tau$ denotes an exponential rate of growth, with τ being the mean cell division time (or population doubling time; Chapter 9), and r' and r denoting the growth rates in the presence and absence of the attribute under consideration. Denoting the increase in cell-division time as $\Delta_\tau = (\tau' - \tau)$,

$$s = \ln(2)\left(\frac{1}{\tau} - \frac{1}{\tau'}\right) \simeq \frac{\ln(2)\cdot\Delta_\tau}{\tau}. \tag{17.1.1b}$$

Further assuming that the cost of the trait is much less than the lifetime energetic expenditure of a cell, $c_T \ll C_T$, so that the increment in cell-division time scales linearly with the proportional increase in investment,

$$\tau' \simeq \tau\left(1 + \frac{c_T}{C_T}\right). \tag{17.1.2}$$

Noting that $\Delta_\tau \simeq \tau c_T/C_T$ leads to

$$s \simeq \frac{\ln(2)\cdot c_T}{C_T}. \tag{17.1.3}$$

This shows that the intrinsic selective disadvantage associated with the bioenergetic cost of a trait scales directly with the proportional increase in the total energy demand per cell cycle (Lynch and Marinov 2015; Ilker and Hinczewski 2019).

Foundations 17.2. The biosynthetic costs of nucleotides and amino acids. Critical to understanding the energetic costs of nucleic acids and proteins are the biosynthetic costs of the basic building blocks from which these are built. Of the numerous published accounts for such costs (Atkinson 1970; McDermitt and Loomis 1981; Williams 1987; Craig and Weber 1998; Akashi and Gojobori 2002; van Milgen 2002; Wagner 2005; Barton et al. 2010; Arnold et al. 2015), most are presented without detail or reference to prior work, and none are in entire agreement. Thus, given the morass of technical literature on biochemical pathways underlying such computations, it is desirable to have steps involved in the cost computations laid out in enough detail that the reader can readily make modifications should the biochemistry interpretations be deemed suspect.

As a unit of energetic currency, we will rely on ATP usage, specifically the number of phosphorus atoms released via ATP hydrolyses, the primary source of energy in most endergonic cellular reactions. There are, however, two complications to deal with. First, instead of ATP, CTP and GTP are utilized in a few cellular reaction steps (e.g., lipid biosynthesis), and these will be treated as equivalent to ATP. Second, electron transfers (through the electron-transfer chain) resulting from conversions of coenzyme NADH to $NADH^+$, NADPH to $NADPH^+$, and $FADH_2$ to FAD drive the delivery

of hydrogen ions (H^+) that contribute to the proton-motive force used to produce ATP. As there is a precise recipe for converting hydrogen ions flowing through ATP synthase into ATP production, coenzyme use can be converted to ATP equivalents, as now described.

ATP / coenzyme equivalents. Recall from Chapter 2 that ATP is produced by ATP synthase complexes via the loading of protons (H^+) into the subunits of the rotating c ring in the F_o complex, which sits in mitochondrial membranes of eukaryotes and in the plasma membrane of prokaryotes. Rotation of the ring requires that each subunit be loaded with a proton, and it is generally believed that each rotation universally leads to the production of 3 ATPs. In eukaryotes, each NADH drives the pumping of 10 protons via the electron-transport chain, while $FADH_2$ has a lower energy state and delivers only 6. Thus, if the c ring contains $n = 12$ subunits, $12/3 = 4$ protons are required to produce each ATP. This is the basis for the common textbook assertion that the ATP value of a NADH is on the order of 2.5 to 3.0, as $10/4 = 2.5$. However, the number of c subunits actually varies from $n = 8$ to 15 (e.g., $n = 10$ for yeast and *E. coli,* and 8 for mammals, with most species being near the lower end; Chapter 2).

For eukaryotes, there is an additional small complication in that mitochondrially produced ATP must be exported to the cytoplasm and ADP imported back. This requires the use of one extra H^+ per ATP exported, so the H^+ to ATP ratio in eukaryotes is $(n+3)/3$, and the ATP/NADH ratio becomes $10 \cdot 3/(n+3) = 30/(n+3)$, i.e., 2.7, 2.3, and 2.0, for $n = 8$, 10, and 12, respectively. For prokaryotes, ATP/ADP transport is not required, and electron-transport chains are shorter, with 8 rather than 10 protons released by an NADH-initiated chain reaction (Nicholls and Ferguson 2013), so the ATP/NADH ratio is on the order of $8 \cdot 3/n = 24/n$. In this case, $n = 8$, 10, and 12, lead to ratios of 3.0, 2.4, and 2.0, respectively, similar to those for eukaryotes.

Thus, under the assumption that $n = 8$ to 11 or so in most organisms, and acknowledging that the exact value of n (determination of which requires structural work) is generally unknown, an ATP/NADH conversion factor of 2.5 appears to be well-justified for both eukaryotes and prokaryotes, and to keep things simple, the following analyses will adhere to this value. Given that $FADH_2$ generates only 6 H^+, then $(6/10) \times 2.5 = 1.5$, which is commonly stated for the ATP/$FADH_2$ equivalence, seems also to be justified. Should the data for a particular taxon warrant modifications of these ratios, the following calculations are readily modifiable. An excellent summary of the issues underlying the confusing use of different conversion ratios in the literature is given by Silverstein (2005).

Costs of precursor molecules. All cellular building blocks are constructed out of carbon skeletons that are ultimately derived from an organic carbon source, here assumed to be glucose. Generally, however, organic molecules are sources of both carbon and energy. As a carbon resource progressively moves down energy-generating metabolic pathways, when the modified intermediate products are diverted into side pathways for biosynthesis, the precursor molecule is no longer available for downstream energy production, the loss of which is viewed as an opportunity cost. The following paragraphs provide verbal explanations for the quantitative derivation of the opportunity costs associated with eleven precursor molecules necessary for amino-acid and nucleotide biosynthesis. All of these precursors reside in key positions in the TCA cycle, glycolytic pathway, or pentose-phosphate shunt. The key steps and products generated are outlined in Figure 17.7, omitting numerous non-energy-generating steps that are covered in detail in all biochemistry textbooks. Using the above conversion factors, these are then transformed into ATP equivalents in the following table.

We adopt the strategy of Atkinson (1970), which starts with the assumption

that, under aerobic metabolism, after descending through glycolysis and the citric-acid (also known as TCA or Krebs') cycle (Figure 17.7), a single glucose molecule would yield a net 4 ATP, 10 NADH, and 2 $FADH_2$ molecules, for a total value of $4 + (10 \cdot 2.5) + (2 \cdot 1.5) = 32$ ATP equivalents. In the first steps of glycolysis, two phosphates are added, yielding the loss of two ATPs, so the biphosphorylated product (fructose-1,6-biphosphate, not shown) has actually gained energy and has a value of 34 ATPs (as the initial subtraction of two ATPs is no longer necessary). The first precursor molecules of interest are the 3-carbon metabolites dhap and g3p, which are derived by splitting the prior 6-carbon molecule in glycolysis, and each has a value of 34/2. Thereafter, the descending metabolites have increasingly small values, as ATPs and/or NADHs have been produced at higher steps in the chain. An exception arises upon arrival at the citric-acid cycle, as acetyl-CoA joins oaa, yielding a summed value of $10 + 10 = 20$ ATPs. The early production of 1 NADH in the cycle results in αkg having a reduced value of $20 - 2.5 = 17.5$ ATP equivalents, and further losses of four products reduces oaa to 10.

The remaining precursor molecules of interest are associated with the pentose-phosphate shunt, which diverts glucose derivatives from the glycolytic pathway to produce nucleotides (below). The first component in this pathway, ribose-5-phosphate (penP), is derived from g6p at the expense of 2 NADHs, and therefore has an energetic value of $33 - (2 \cdot 2.5) = 28$ ATP equivalents. Production of eryP is energy neutral, whereas pRpp arises after an ATP $\rightarrow$ AMP conversion, and therefore has a value elevated by 2 ATP equivalents.

The formula Total ATP Cost $= (1 \cdot \text{ATP}) + (2.5 \cdot \text{NADH}) + (1.5 \cdot \text{FADH}_2)$ yields the total costs for each precursor, summarized in the following table. Modifications can become necessary with sources of carbon other than glucose, e.g., see Akashi and Gojobori (2002) for acetate and malate (although some of their computations stray from the above scheme).

Precursor	Abbrev.	ATP	NADH	FADH2	Total
Ribose 5-phosphate	penP	5	8	2	28.0
5-Phosphoribosyl pyrophosphate	pRpp	7	8	2	30.0
Erythrose 4-phosphate	eryP	5	8	2	28.0
Dihydroxyacetone phosphate	dhap	3	5	1	17.0
Glyceraldehyde-3-phosphate	g3p	3	5	1	17.0
3-Phosphoglycerate	3pg	2	4	1	13.5
Phosphoenolpyruvate	pep	2	4	1	13.5
Pyruvate	pyr	1	4	1	12.5
Acetyl-CoA	acCoA	1	3	1	10.0
Oxaloacetate	oaa	1	3	1	10.0
α-ketoglutarate	αkg	2	5	2	17.5

Amino-acid biosynthesis. As an entrée into the logic of estimating the cost of a basic building block, we now turn to the amino-acids used in protein production. All of these are derived from the above-noted precursor molecules, and although variants exist, the downstream biosynthetic paths are conserved in most species (Chapter 19). Acknowledging that future researchers may wish to make modifications in some cases, we retain a focus on the situation in which entry into metabolism is based on glucose.

The basic calculations start with the costs of the precursor molecules used for carbon skeletons, which then descend down various modifying steps. Many of these steps require an investment loss in energy of at least the same order of magnitude as ATP, and so must enter into the final bookkeeping. In a few cases, a step is energy producing. Detailed information on these matters can be found in most biochemistry and/or bioenergetics text books, and a simple listing of relevant reactions and their

values in terms of ATP hydrolyses is given below without further explanation.

Reaction	ATP
1) NADH → NAD^+	2.5
1) NADPH → $NADP^+$	2.5
2) Trans-aminotransferase reaction	3.5
3) Glutamine → Glutamate	1.0
4) Aspartate → Fumarate	1.0
5) Acetyl-CoA → CoA-SH	2.5
6) ATP → AICAR	3.0
7) Succinyl-CoA → CoA	1.0
8) H_4 folate → N^5,N^{10}-Methylene H_4 folate	−2.5
9) N^5-Methyl H_4 folate → H_4 folate	5.0
10) N^{10}-Formyl H_4 folate → H_4 folate	1.0

The following provides a list of the precursors and the reaction steps involved in the synthesis of each amino acid. The total precursor costs, obtained by weighting each component by its total ATP content (above) and summing, constitute estimates of the opportunity costs for each amino acid. The direct costs are estimated by weighting each step in the downstream biosynthetic pathway with its associated energy and summing. Numbers in parentheses under Reactions denote the net number of reactions of the indexed type; if not noted, this number is 1, and a negative number implies an energy gain. All costs are in units of numbers of ATP hydrolysis equivalents.

Amino acid	Precursor	ATP	Reactions	Opportunity	Direct	Total
Alanine	pyr		1	12.5	3.5	16.0
Arginine	αkg	1	$1(2), 2, 4, 5$	17.5	13.0	30.5
Asparagine	oaa	2	$2, 3$	10.0	6.5	16.5
Aspartate	oaa		2	10.0	3.5	13.5
Cysteine	3pg		$1(-1), 2, 5$	13.5	3.5	17.0
Glutamate	αkg		1	17.5	2.5	20.0
Glutamine	αkg	1	1	17.5	3.5	21.0
Glycine	3pg		$1(-1), 2, 8$	13.5	−1.5	12.0
Histidine	pRpp		$1(-2), 2, 3, 6$	30.0	2.5	32.5
Isoleucine	pyr, oaa	2	$1(3), 2(2)$	22.5	16.5	39.0
Leucine	2 pyr, acCoA	2	2(2)	35.0	9.0	44.0
Lysine	pyr, oaa	1	$1(2), 2(2)$	22.5	14.0	36.5
Methionine	oaa, cysteine, -pyr	1	$1(2), 2$	9.5	15.5	25.0
Phenylalanine	2 pep, eryP	1	2	55.0	7.0	62.0
Proline	αkg	1	1(3)	17.5	8.5	26.0
Serine	3pg		$1(-1), 2$	13.5	1.0	14.5
Threonine	oaa	2	$1(2), 2$	10.0	10.5	20.5
Tryptophan	2 pep, eryP, pRpp, -pyr, serine, -g3p	1	3	69.0	2.0	71.0
Tyrosine	2 pep, eryP	1	$1(-1), 2$	55.0	2.0	57.0
Valine	2 pyr		$1, 2$	25.0	6.0	31.0

For some amino acids, there can be shifts to alternative biosynthetic pathways under certain environmental conditions, although it remains unclear whether the cost estimates outlined above are greatly altered (Du et al. 2018).

Nucleotide biosynthesis. The general strategy outlined for amino acids is readily extended to estimating the costs of nucleotide biosynthesis, again facilitated by the fact that the basic biosynthetic pathways are nearly universal across the Tree of Life. All nucleotide synthesis starts with the precursor pRpp (phosphoribosyl pyrophosphate) as a carbon skeleton. This five-carbon ring is an early derivative of glucose-6-phosphate, which if not diverted would eventually descend down the citric-acid cycle (Figure 17.7). Amino-acid molecules also contribute carbon skeletons, and here it is assumed that they must first be synthesized. As these could have been utilized for other purposes (or not produced at all, thereby saving energy), they (along with pRpp) will be viewed as opportunity costs. The direct costs include additional investments of ATP and other cofactors in the downstream steps of molding and modifying the precursor carbon skeletons en route to producing nucleotides.

Consider first the production of purines (adenine and guanine). From the table above, it can be seen that the diversion of pRpp deprives the cell of 30 ATP hydrolysis equivalents. One glycine molecule is also consumed in the early stages of biosynthesis, the production of which results in an opportunity loss of 12 ATPs. This leads to a total opportunity loss for purines of $30 + 12 = 42$ ATPs per utilized pRpp molecule (Figure 17.8).

As pRpp and glycine merge and descend down ten steps in the production of the final purine precursor (IMP), direct costs consume the equivalent of ~ 9 ATPs; these derive from: 4 ATP $\rightarrow$ ADP; 2 glutamine $\rightarrow$ glutamate; 2 N^{10}-formyl H_4 folate $\rightarrow$ H_4 folate; and 1 aspartate $\rightarrow$ fumarate reactions. Conversion of IMP to AMP consumes a GTP and an aspartate to fumarate exchange, for an additional direct cost of 2 ATPs in AMP production. Conversion of IMP to GMP involves 1 ATP $\rightarrow$ AMP reaction, production of 1 NADH, and 1 glutamine $\rightarrow$ glutamate exchange for an additional direct cost of 0.5 ATPs in GMP production.

Pyrimidine (C and T) production also starts with pRpp, but consumes an aspartate (rather than glycine) molecule with an opportunity cost of 13.5 ATPs, yielding a total opportunity cost of 43.5 ATPs. The route to the final pyrimidine precursor UMP results in the production of 1 NADH, for a direct cost of -2.5 ATP equivalents. Conversion of UMP to UTP consumes 2 ATPs, and then conversion of UTP to CTP consumes an additional ATP, and involves a glutamine $\rightarrow$ glutamate exchange, for an additional direct cost of 4 ATP equivalents. Conversion of UMP to TMP involves a N^5,N^{10}-Methylene H_4 folate $\rightarrow$ H_4 folate exchange, which is equivalent to a direct cost of 2.5 ATP equivalents, and further conversion to TTP requires a final expenditure of 2 ATPs.

The costs of the four ribonucleotides are summed up in the following table.

Nucleotide	Opportunity	Direct	Total
Adenine (ATP)	42.0	13.0	55.0
Guanine (GTP)	42.0	11.5	53.5
Cytosine (CTP)	43.5	1.5	45.0
Uracil (UTP)	43.5	−0.5	43.0
Thymine (TTP)	43.5	2.0	45.5

Foundations 17.3. The mutational hazard of excess DNA. All genes have a mutational target size equal to the number of nucleotide sites (at the DNA level) for which the nucleotide identity has the potential to influence fitness (Lynch 2007a). This will include most amino-acid replacement sites in the coding region, and to a much

lesser (but not always insignificant) extent silent sites (for which the nucleotide identity has no impact on the encoded amino acid, but may influence the rate of translation). Here, we consider the impact of slightly larger segments of gene-associated noncoding DNA that are relevant to successful gene expression. These include introns, which commonly populate eukaryotic genes, a variety of transcription-factor binding sites, and numerous other regulatory sequences in the $5'$ untranslated regions (UTRs) of transcripts.

What is the contribution of such elements for the vulnerability of a gene to mutational inactivation? Introns are transcribed into pre-mRNAs, and must be properly spliced out to yield a productive mature mRNA, with accurate recognition by the splicing machinery depending on motifs at both intron ends comprising a total ~ 20 to 30 nucleotides. Transcription-factor binding sites typically consists of motifs 8 to 16 bases in length, and other elements in UTRs are usually of this same size. Thus, to simplify discussion, we will consider an embellishment to a gene that magnifies the mutational target size by 10 nucleotides. Notably, even entirely nonfunctional DNA can magnify the vulnerability to deleterious mutations, as such material can acquire detrimental gain-of-function mutations, although the magnitude of such effects is difficult to quantify.

The mutation rate per nucleotide site per generation ranges from $\sim 10^{-10}$ in prokaryotes and many unicellular eukaryotes to 10^{-8} in vertebrates (Chapter 4), so the addition of a gene-structural embellishment that introduces 10-bp of sequence critical to gene function is equivalent to increasing the mutation rate to null alleles by 10^{-9} to 10^{-7}. This type of mutation-rate inflation for an embellished gene operates in a manner effectively identical to selection, as it is a measure of the excess rate of removal of such alleles from the population by conversion to null alleles.

How do these mutational hazards compare with the energetic costs of nucleotides? Returning to the results in the text, a 10-bp segment of DNA imposes an energetic penalty (relative to the total cost of building a cell) of $\sim 10^{-8}$ in a typical bacterium, $\sim 10^{-9}$ in a unicellular eukaryote, and $\sim 10^{-10}$ in a multicellular eukaryote. These rough calculations suggest that for bacteria, the primary selective disadvantage of excess DNA is associated with energetic costs (this, at most, being of the same order of magnitude as the mutational hazard). In contrast, the mutational cost starts to exceed the energetic penalty in unicellular eukaryotes, and greatly exceeds it in multicellular species. In the latter case, however, even the mutational cost of a gene-structural embellishment is insufficient to overcome the power of random genetic drift. Hence, from both the perspectives of energetics and genetics, we expect a gradient in genome size and gene-structural complexity from prokaryotes to multicellular eukaryotes, provided there is a mutational bias towards insertions.

Foundations 17.4. Numbers of molecules required in a cellular lifespan. From the time of birth to the time of cell division, for any particular cellular feature, a cell must accumulate new constituent molecules to a level consistent with the birth of a new cell, and in doing so will often have to acquire replacement molecules to balance any decay processes. Here, we consider the situation in which a newborn cell contains N_0 molecules of the type being considered (e.g., the number of transcripts or protein molecules associated with a particular gene, or the number of lipid molecules in a cell membrane). This number must then double to $2N_0$ molecules to allow for binary fission. Letting β be the rate of production of the molecule and δ be the rate of decay, so that $r = \beta - \delta$ is the net growth rate in cell size, then assuming exponential growth

$$\frac{dN}{dt} = rN \tag{17.4.1}$$

denotes the rate of increase in molecule number. This expression integrates to

$$N_t = N_0 e^{rt}, \tag{17.4.2}$$

so with the cell dividing when $N_t = 2N_0$, the cell-division time is $T = \ln(2)/r$.

The average number of molecules in the parental cell over its entire lifespan is obtained by integrating over Equation 17.4.2 from 0 to T,

$$\overline{N} = \frac{N_0}{T}\int_0^T e^{rt}\cdot dt = \frac{N_0(e^{rT}-1)}{rT} = \frac{N_0}{\ln(2)}, \tag{17.4.3}$$

with the final simplification following from the fact that $rT = \ln(2)$. The total number of molecules produced per cellular lifespan (N_p) is then the product of the average number of molecules ($\overline{N}$), the production rate per molecule (β), and the cell-division time (T),

$$N_p = \frac{\beta N_0 T}{\ln(2)} \tag{17.4.4}$$

which by using $\beta = r + \delta$ and $T = \ln(2)/r$ becomes

$$N_p = N_0\left(1 + \frac{\delta T}{\ln(2)}\right). \tag{17.4.5}$$

A simple interpretation of this expression is that during its lifespan, a cell must produce N_0 new, surviving molecules (to yield an offspring cell) and $\delta N_0 T/\ln(2)$ replacement molecules to offset molecular degradation/loss. Note that this second (maintenance) term increases linearly with the cell-division time.

Foundations 17.5. The biosynthetic costs of lipid molecules. Estimation of the total cellular expenditure on the synthesis of a lipid molecule requires separate consideration of the investments in the three subcomponents of these molecules: the fatty-acid tails, head groups, and linkers (Lynch and Trickovic 2020). As in the applications for amino acids and nucleotides, we will quantify such costs in units of the number of phosphorus atoms released via ATP hydrolyses. CTP, which is utilized in a few reaction steps in lipid biosynthesis, will be treated as equivalent to ATP, and electron transfers resulting from conversions of NADH to $NADH^+$ and NADPH to $NADPH^+$ will again be assumed equivalent to 2.5 ATPs, and $FADH_2$ to FAD to 1.5 ATPs (the resultant computations are slightly different than those in Lynch and Trickovic (2020) owing to slightly different assumptions about these conenzyme conversions).

The starting point for the synthesis of most fatty acids is the production of one particular linear chain, palmitate, containing 16 carbon atoms. Production of this molecule takes place within a large complex, known as fatty-acid synthase, and in bacteria biosynthesis of each molecule requires the consumption of 8 acetyl-CoA molecules, 7 ATPs, and reductions involving 14 NADPH. As noted in Foundations 17.2, each molecule of acetyl-CoA is equivalent to a net opportunity loss of 10 ATPs. Thus, the total cost of production of one molecule of palmitate is $(8\times 10) + (7\times 1) + (14\times 2.5) = 122$ ATP in bacteria. Fatty-acid production is slightly more expensive in nonphotosynthetic eukaryotes, where acetyl-CoA is produced in the mitochondrion and reacts with oxaloacetate to produce citrate, which must then be exported. Cleavage of oxalacetate in the cytosol yields acetyl-CoA at the expense of 1 ATP, and a series of reactions serve to return oxaloacetate to the citric-acid cycle in an effectively ATP

neutral way. Thus, the cost of palmitate increases to $122 + (8 \times 1) = 130$ ATP. Each additional pair of carbons added to this primary fatty-acid chain requires 1 additional acetyl-CoA and ATP, and 2 additional NADPHs, or an equivalent of 15 ATPs, and each subsequent desaturation of a bond consumes one NADPH, or 2.5 ATP equivalents.

To evaluate the total cost of a lipid molecule, we first consider the situation for glycerophospholipids, for which glycerol-3-phosphate serves as the linker between the fatty-acid chains and the headgroup. Conversion of glycerol-3-phosphate from dhap (within the glycolysis pathway, and having a cost of 17 ATPs; Figure 17.7) consumes 1 NADH, so the use of this molecule as a linker in a lipid molecule has an opportunity cost of $17 + 2.5 = 19.5$ ATPs. Linking each of the two fatty-acid tails requires 1 ATP, and linking the head group involves two CTP hydrolyses, yielding a total cost of $(19.5 + 2 + 2) = 23.5$.

All that remains is the cost of synthesis of the head group. In the case of phosphatidylglycerol, the head group is glycerol-3-phosphate, the cost of which is 19.5 ATP, as just noted, so the total cost of this molecule in a bacterium is $\simeq (2 \cdot 122) + 23.5 + 19.5 = 287$ ATP. From Foundations 17.2, the cost of a serine is 13.5 ATP, 6 fewer than for glycerol-3-phosphate, so the total cost of a phosphatidylserine is 281 ATP, and because ethanolamine and choline are simple derivatives of serine, this closely approximates the costs of both phosphatidylethanolamine and phosphatidylcholine. The headgroup of phosphatidylinositol is inosital, which is derived from glucose-6-phosphate (Figure 17.7), diverting the latter from glycolysis and depriving the cell of the equivalent of 33 ATPs, so the total cost of production of this molecule is 300.5 ATP. Finally, cardiolipin is synthesized by the fusion of two phosphatidylglycerols and the release of one glycerol, so taking the return from the latter to be 15 ATP, the total cost per molecule produced is $(2 \times 287) - 15 = 559$ ATP for bacteria (and 575 for eukaryotes), and the respective direct costs are 176 and 208 ATPs.

Estimation of the cost of biosynthesis of sphingolipids follows many of the steps just outlined. Construction of the linker molecule requires a palmitate molecule and the expenditure of 1 NADPH, for a total of 139 and 147 ATP in bacteria and eukaryotes, respectively. Then, a single fatty-chain is added, so assuming this is palmitate, this requires the expenditure of another 125.5 or 133.5 ATP. Finally, there are the costs of synthesizing and adding the head group, both of which are outlined in the preceding paragraphs.

Literature Cited

Akashi, H., and T. Gojobori. 2002. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA 99: 3695-3700.

Adler, M., M. Anjum, O. G. Berg, D. I. Andersson, and L. Sandegren. 2014. High fitness costs and instability of gene duplications reduce rates of evolution of new genes by duplication-divergence mechanisms. Mol. Biol. Evol. 31: 1526-1535.

Arnold, A., M. Sajitz-Hermstein, and Z. Nikoloski. 2015. Effects of varying nitrogen sources on amino acid synthesis costs in *Arabidopsis thaliana* under different light and carbon-source conditions. PLoS One 10: e0116536.

Atkinson, D. E. 1970. Adenine nucleotides as universal stoichiometric metabolic coupling agents. Adv. Enzyme Regul. 9: 207-219.

Barton, M. D., D. Delneri, S. G. Oliver, M. Rattray, and C. M. Bergman. 2010. Evolutionary systems biology of amino acid biosynthetic cost in yeast. PLoS One 5: e11935.

Bauchop, T., and S. R. Elsden. 1960. The growth of micro-organisms in relation to their energy supply. J. Gen. Microbiol. 23: 457-469.

Belle, A., A. Tanay, L. Bitincka, R. Shamir, and E. K. O'Shea. 2006. Quantification of protein half-lives in the budding yeast proteome. Proc. Natl. Acad. Sci. USA 103: 13004-13009.

Bernstein, J. A., A. B. Khodursky, P. H. Lin, S. Lin-Chao, and S. N. Cohen. 2002. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. Proc. Natl. Acad. Sci. USA 99: 9697-9702.

Booth, A., and W. F. Doolittle. 2015. Eukaryogenesis, how special really? Proc. Natl. Acad. Sci. USA 112: 10278-10285.

Chiyomaru, K., and Takemoto K. 2020. Revisiting the hypothesis of an energetic barrier to genome complexity between eukaryotes and prokaryotes. R. Soc. Open Sci. 7: 191859.

Craig, C. L., and R. S. Weber. 1998. Selection costs of amino acid substitutions in ColE 1 and colIa gene clusters harbored by *Escherichia coli* . Mol. Biol. Evol. 15: 774-776.

Csárdi, G., A. Franks, D. S. Choi, E. M. Airoldi, and D. A. Drummond. 2015. Accounting for experimental noise reveals that mRNA levels, amplified by post-transcriptional processes, largely determine steady-state protein levels in yeast. PLoS Genet. 11: e1005206.

Dressaire, C., F. Picard, E. Redon, P. Loubière, I. Queinnec, L. Girbal, and M. Cocaign-Bousquet. 2013. Role of mRNA stability during bacterial adaptation. PLoS One 8: e59059.

Du, B., D. C. Zielinski, J. M. Monk, and B. O. Palsson BO. 2018. Thermodynamic favorability and pathway yield as evolutionary tradeoffs in biosynthetic pathway choice. Proc. Natl. Acad. Sci. USA 115: 11339-11344.

Duboc, P., N. Schill, L. Menoud, W. van Gulik, and U. von Stockar. 1995. Measurements of sulfur, phosphorus and other ions in microbial biomass: influence on correct determination of elemental composition and degree of reduction. J. Biotechnol. 43: 145-158.

Eden, E., N. Geva-Zatorsky, I. Issaeva, A. Cohen, E. Dekel, T. Danon, L. Cohen, A. Mayo, and U. Alon. 2011. Proteome half-life dynamics in living human cells. Science 331: 764-768.

Eguchi, Y., K. Makanae, T. Hasunuma, Y. Ishibashi, K. Kito, and H. Moriya. 2018. Estimating

the protein burden limit of yeast cells by measuring the expression limits of glycolytic proteins. eLife 7: e34595.

François-Martin, C., J. E. Rothman, and F. Pincet. 2017. Low energy cost for optimal speed and control of membrane fusion. Proc. Natl. Acad. Sci. USA 114: 1238-1241.

Frumkin, I., D. Schirman, A. Rotman, F. Li, L. Zahavi, E. Mordret, O. Asraf, S. Wu, S. F. Levy, and Y. Pilpel. 2017. Gene architectures that minimize cost of gene expression. Mol. Cell 65: 142-153.

Geiger, O., N. González-Silva, I. M. López-Lara, and C. Sohlenkamp. 2010. Amino acid-containing membrane lipids in bacteria. Prog. Lipid Res. 49: 46-60.

Geiler-Samerotte, K. A., M. F. Dion, B. A. Budnik, S. M. Wang, D. L. Hartl, and D. A. Drummond. 2011. Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. Proc. Natl. Acad. Sci. USA 108: 680-685.

Ghaemmaghami, S., W. K. Huh, K. Bower, R. W. Howson, A. Belle, N. Dephoure, E. K. O'Shea, and J. S. Weissman. 2003. Global analysis of protein expression in yeast. Nature 425: 737-741.

Golding, I., J. Paulsson, S. M. Zawilski, and E. C. Cox. 2005. Real-time kinetics of gene activity in individual bacteria. Cell 123: 1025-1036.

Guschina, I. A., and J. L. Harwood. 2006. Lipids and lipid metabolism in eukaryotic algae. Prog. Lipid Res. 45: 160-186.

Hambraeus, G., C. von Wachenfeldt, and L. Hederstedt. 2003. Genome-wide survey of mRNA half-lives in *Bacillus subtilis* identifies extremely stable mRNAs. Mol. Genet. Genomics 269: 706-714.

Helbig, A. O., P. Daran-Lapujade, A. J. van Maris, E. A. de Hulster, D. de Ridder, J. T. Pronk, A. J. Heck, and M. Slijper. 2011. The diversity of protein turnover and abundance under nitrogen-limited steady state conditions in *Saccharomyces cerevisiae.* Mol. Biosyst 7: 3316-3326.

Heizer, E. M., Jr., D. W. Raiford, M. L. Raymer, T. E. Doom, R. V. Miller, and D. E. Krane. 2006. Amino acid cost and codon-usage biases in 6 prokaryotic genomes: a whole-genome analysis. Mol. Biol. Evol. 23: 1670-1680.

Heizer, E. M., Jr., M. L. Raymer, and D. E. Krane. 2011. Amino acid biosynthetic cost and protein conservation. J. Mol. Evol. 72: 466-473.

Ilker, E., and M. Hinczewski. 2019. Modeling the growth of organisms validates a general relation between metabolic costs and natural selection. Phys. Rev. Lett. 122: 238101.

Islam, S., U. Kjällquist, A. Moliner, P. Zajac, J. B. Fan, P. Lönnerberg, and S. Linnarsson. 2011. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res. 21: 1160-1167.

Krick, T., N. Verstraete, L. G. Alonso, D. A. Shub, D. U. Ferreiro, M. Shub, and I. E. Sánchez. 2014. Amino acid metabolism conflicts with protein diversity. Mol. Biol. Evol. 31: 2905-2912.

Kučerka, N., M. P. Nieh, and J. Katsaras. 2011. Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature. Biochim. Biophys. Acta 1808: 2761-2771.

Lahtvee, P. J., A. Seiman, L. Arike, K. Adamberg, and R. Vilu. 2014. Protein turnover forms one of the highest maintenance costs in *Lactococcus lactis.* Microbiology 160: 1501-1512.

Lane, N., and W. Martin. 2010. The energetics of genome complexity. Nature 467: 929-934.

Lewis, B. A., and D. M. Engelman. 1983. Lipid bilayer thickness varies linearly with acyl chain length in fluid phosphatidylcholine vesicles. J. Mol. Biol. 166: 211-217.

Li, G. W., and X. S. Xie. 2011. Central dogma at the single-molecule level in living cells. Nature 475: 308-315.

Lotka, A. J. 1922. Contribution to the energetics of evolution. Proc. Natl. Acad. Sci. USA 8: 147-151.

Lu, P., C. Vogel, R. Wang, X. Yao, and E. M. Marcotte. 2007. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat. Biotechnol. 25: 117-124.

Lynch, M. 2007a. The Origins of Genome Architecture. Sinauer Assocs., Inc., Sunderland, MA.

Lynch, M. 2007b. The frailty of adaptive hypotheses for the origins of organismal complexity. Proc. Natl. Acad. Sci. USA 104 (Suppl.): 8597-8604.

Lynch, M., L.-M. Bobay, F. Catania, J.-F. Gout, and M. Rho. 2011. The repatterning of eukaryotic genomes by random genetic drift. Ann. Rev. Genomics Hum. Genet. 12: 347-366.

Lynch, M., and G. K. Marinov. 2015. The bioenergetic costs of a gene. Proc. Natl. Acad. Sci. USA 112: 15690-15695.

Lynch, M., and G. K. Marinov. 2017. Membranes, energetics, and evolution across the prokaryote-eukaryote divide. eLife 6: e20437.

Mani, S., and M. Thattai. 2016. Stacking the odds for Golgi cisternal maturation. elife 5: e16231.

Marinov, G. K., B. A. Williams, K. McCue, G. P. Schroth, J. Gertz, R. M. Myers, and B. J. Wold. 2014. From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. Genome Res. 24: 496-510.

McDermitt, D. K., and R. S. Loomis. 1981. Elemental composition of biomass and its relation to energy content, growth efficiency, and growth yield. Ann. Bot. 48: 275-290.

Mitra, K., I. Ubarretxena-Belandia, T. Taguchi, G. Warren, and D. M. Engelman. 2004. Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol. Proc. Natl. Acad. Sci. USA 101: 4083-4088.

Nagle, J. F., and S. Tristram-Nagle. 2000. Structure of lipid bilayers. Biochim. Biophys. Acta 1469: 159-195.

Neymotin, B., R. Athanasiadou, and D. Gresham. 2014. Determination of *in vivo* RNA kinetics using RATE-seq. RNA 20: 1645-1652.

Nicholls, D. G., and S. J. Ferguson. 2013. Bioenergetics. 4th Ed. Academic Press, New York, NY.

Petrache, H. I., S. W. Dodd, and M. F. Brown. 2000. Area per lipid and acyl length distributions in fluid phosphatidylcholines determined by ^{2}H NMR spectroscopy. Biophys. J. 79: 3172-3192.

Phillips, M. J., and G. K. Voeltz. 2016. Structure and function of ER membrane contact sites with other organelles. Nat. Rev. Mol. Cell. Biol. 17: 69-82.

Phillips, R., J. Kondev, J. Theriot, and H. Garcia. 2012. Physical Biology of the Cell, 2nd Ed. Garland Science, New York, NY.

Plata, G., M. E. Gottesman, and D. Vitkup. 2010. The rate of the molecular clock and the cost of gratuitous protein synthesis. Genome Biol. 11: R98.

Raiford, D. W., E. M. Heizer, Jr., R. V. Miller, H. Akashi, M. L. Raymer, and D. E. Krane. 2008. Do amino acid biosynthetic costs constrain protein evolution in *Saccharomyces cerevisiae*? J. Mol. Evol. 67: 621-630.

Raiford, D. W., E. M. Heizer, Jr., R. V. Miller, T. E. Doom, M. L. Raymer, and D. E. Krane. 2012. Metabolic and translational efficiency in microbial organisms. J. Mol. Evol. 74: 206-216.

Raj, A., C. S. Peskin, D. Tranchina, and D. Y. Vargas, and S. Tyagi. 2006. Stochastic mRNA synthesis in mammalian cells. PLoS Biol. 4: e309.

Ramadas, R., and M. Thattai. 2013. New organelles by gene duplication in a biophysical model of eukaryote endomembrane evolution. Biophys. J. 104: 2553-2563.

Rand, R. P., and V. A. Parsegian. 1989. Hydration forces between phospholipid bilayers. Biochim. Biophys. Acta 988: 351-376.

Rawicz, W., K. C. Olbrich, T. McIntosh, D. Needham, and E. Evans. 2000. Effect of chain length and unsaturation on elasticity of lipid bilayers. Biophys. J. 79: 328-339.

Russell, J. B., and G. M. Cook. 1995. Energetics of bacterial growth: balance of anabolic and catabolic reactions. Microbiol. Rev. 59: 48-62.

Schwanhäusser, B., D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach. 2011. Global quantification of mammalian gene expression control. Nature 473: 337-342.

Seligmann, H. 2003. Cost-minimization of amino acid usage. J. Mol. Evol. 56: 151-161.

Shahrezaei, V., and P. S. Swain. 2008. Analytical distributions for stochastic gene expression. Proc. Natl. Acad. Sci. USA 105: 17256-17261.

Silverstein, T. 2005. The mitochondrial phosphate-to-oxygen ratio is not an integer. Biochem. Mol. Biol. Educ. 33: 416-417.

Steinman, R. M., S. E. Brodie, and Z. A. Cohn. 1976. Membrane flow during pinocytosis: a stereologic analysis. J. Cell Biol. 68: 665-687.

Stoebel, D. M., A. M. Dean, and D. E. Dykhuizen. 2008. The cost of expression of *Escherichia coli* lac operon proteins is in the process, not in the products. Genetics 178: 1653-1660.

Stouthamer, A. H. 1973. A theoretical study on the amount of ATP required for synthesis of microbial cell material. Antonie Van Leeuwenhoek 39: 545-65.

Swire, J. 2007. Selection on synthesis cost affects interprotein amino acid usage in all three domains of life. J. Mol. Evol. 64: 558-571.

Taniguchi, Y., P. J. Choi, G. W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie. 2010. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science 329: 533-538.

Tempest, D. W., and O. M. Neijssel. 1984. The status of YATP and maintenance energy as biologically interpretable phenomena. Annu. Rev. Microbiol. 38: 459-486.

Thilo, L., and G. Vogel. 1980. Kinetics of membrane internalization and recycling during pinocytosis in *Dictyostelium discoideum.* Proc. Natl. Acad. Sci. USA 77: 1015-1019.

Tomala, K., and R. Korona. 2013. Evaluating the fitness cost of protein expression in *Saccharomyces cerevisiae.* Genome Biol. Evol. 5: 2051-2060.

Trötschel, C., S. P. Albaum, and A. Poetsch. 2013. Proteome turnover in bacteria: current status for *Corynebacterium glutamicum* and related bacteria. Microb. Biotechnol. 6: 708-719.

Uwizeye, C., et al. 2021. Morphological bases of phytoplankton energy management and physiological responses unveiled by 3D subcellular imaging. Nature Comm. 12: 1049.

van Milgen, J. 2002. Modeling biochemical aspects of energy metabolism in mammals. J. Nutr. 132: 3195-3202.

Van Valen, L. 1976. Energy and evolution. Evol. Theory 1: 179-229.

Van Valen, L. 1980. Evolution as a zero-sum game for energy. Evol. Theory 4: 289-300.

Wagner, A. 2005. Energy constraints on the evolution of gene expression. Mol. Biol. Evol. 22: 1365-1374.

Walsh, J. B., and M. Lynch. 2018. Selection and Evolution of Quantitative Traits. Oxford Univ. Press, Oxford, UK.

Wang, Y., C. L. Liu, J. D. Storey, R. J. Tibshirani, D. Herschlag, and P. O. Brown. 2002. Precision and functional specificity in mRNA decay. Proc. Natl. Acad. Sci. USA 99: 5860-5865.

Wieslander, A., S. Nordström, A. Dahlqvist, L. Rilfors, and G. Lindblom. 1995. Membrane lipid composition and cell size of *Acholeplasma laidlawii* strain A are strongly influenced by lipid acyl chain length. Eur. J. Biochem. 227: 734-744.

Williams, K., F. Percival, J. Merino, and H. A. Mooney. 1987. Estimation of tissue construction cost from heat of combustion and organic nitrogen content. Plant Cell Environ. 10: 725-734

Williford, A., and J. P. Demuth. 2012. Gene expression levels are correlated with synonymous codon usage, amino acid composition, and gene architecture in the red flour beetle, Tribolium castaneum. Mol. Biol. Evol. 29: 3755-3766.

Zenklusen, D., D. R. Larson, and R. H. Singer. 2008. Single-RNA counting reveals alternative modes of gene expression in yeast. Nat. Struct. Mol. Biol. 15: 1263-1271.

Figure 17.1. The distinction between direct and opportunity costs associated with synthesizing molecular building blocks (e.g., an amino acid). As the energy resource (e.g., glucose) is partially metabolized into precursor metabolites (carbon skeletons) necessary for synthesis of the building block, the additional energy that could have been captured from the complete metabolism of the resource is the opportunity cost. The conversion of precursor metabolites to some molecular building blocks can also consume electron carrier molecules such as NADH, which if not used in building-block synthesis would have generated ATP, and thus represent additional source of opportunity cost. The consumption of ATP in the biosynthetic process defines the direct cost of building-block synthesis. The assembly of macromolecules such as proteins and nucleic acids from building blocks requires additional post-synthesis investment, such as the costs of assembly (polymerization) and maintenance (e.g., associated with molecular turnover, or DNA repair).

Figure 17.2. The evolutionary distinction between the construction/maintenance cost of a trait and the direct benefits. s_c represents the reduction in fitness that would be expected from the presence of the trait in the absence of any direct benefits; it may be viewed as the selective advantage of a mutant relieved of the trait in an environment in which no advantages of the trait are experienced. s_d is a measure of the increase in fitness that would accrue in the absence of any assembly/maintenance costs. The difference $s_n = s_d - s_c$ denotes the net fitness advantage of the trait; if this value is negative, the trait is selectively disadvantageous despite any ecological benefits accrued. Clearly, a gene will be selected against if $s_d < s_c$, but $s_d > s_c$ is not a sufficient condition for gene preservation by natural selection, as the absolute value of s_n must exceed the power of random genetic drift, $1/N_e$ in a haploid species and $1/(2N_e)$ in a diploid, to be readily perceived by natural selection (Chapter 4).

Figure 17.3. Total costs associated with various cellular building blocks (estimated from biochemical pathways) as a function of the modified index for the degree of chemical reduction.

Figure 17.4. The distribution of the total costs for all genes (and their three components, associated with replication, transcription, and translation) in four species for which the detailed transcriptomic and proteomic data necessary for full analyses are available (from Lynch and Marinov 2015). The data are presented as frequency histograms, so that the peaks represent the mode of the costs over the full set of genes. The lower axis ($\log_{10}$) denotes the total (direct plus opportunity) costs, whereas the upper axis divides these numbers by the total cost of building a cell (from Chapter 8), yielding a measure of the selective cost of the maintenance and operation of a gene relative to the cell's total energy budget. The vertical dashed lines denote the drift barrier, with the middle line approximately demarcating $1/N_e$ for the species, and the two flanking lines simply providing order-of-magnitude margins for error. The fact that the relative costs of virtually all genes in *E. coli* are far to the right of the drift barrier implies that their baseline costs can easily be perceived by selection – if such genes do not pay their way by endowing the cell with benefits in excess of such costs, they would be rapidly purged by degenerative mutation and negative selection. Moving into eukaryotes, the effective population size declines with organism size, moving the drift barrier to the right, whereas the relative costs of genes move to the left owing to the predominant effect of an increase in the total cost of the cell. As a consequence, with increasing organism size, the baseline costs of genes tend to move to the left of the drift barrier, implying that gene-specific costs are too small, particularly at the DNA and RNA levels, to be perceived by selection.

Figure 17.5. Estimated costs (relative to the estimated total cellular energy budgets) for average genes in 44 species of prokaryotes and eukaryotes. For all components of cost, there is a clear negative scaling with increasing cell volume, both within and among groups, with no discontinuity in the scaling behavior between prokaryotes and eukaryotes. From Lynch and Marinov (2015); related analyses appear in Chiyomaru and Takemoto (2020).

Figure 17.6. Scaling of surface areas (A) of internal membranes in eukaryotes. **a)** Total for inner and outer mitochondrial membranes scales as $A = 3.0V^{0.98}$ ($r^2 = 0.92$), where A and V having units of μm^2 and μm^3, with little variation among functionally different types of organisms (phototrophs include land plants and various phylogenetic groups of phytoplankton). **b)** Total surface area of the nucleus scales with cell volume across phylogenetic groups as $A = 2.7V^{0.48}$ ($r^2 = 0.68$). Data from Lynch and Marinov (2017), Uwizeye et al. (2021), and a few other references.

Figure 17.7. Summary of the three major cellular pathways for the production of the three major carriers of energy and reducing power (ATP, NADH, and $FADH_2$), and the routes to the primary precursor molecules leading to the biosynthetic pathways for amino acids and nucleotides. Computations are given in units of ATP equivalents. Note that below g6p in glycolysis, there are actually two copies of each listed derivative, as the 6-carbon glucose has been split into two 3-carbon compounds, and hence the division by 2 at the dhap/g3p position. For simplicity, numerous intermediate steps extraneous to the computations are left out; details on these can be found in all basic biochemistry text books.

Figure 17.8. Key steps in the biosynthesis of ribonucleotides and deoxyribonucleotides, and summations of the energy demands (in units of ATP equivalents per molecule produced). Red and black denote opportunity- and direct-cost components.