Sample Introductions, 7/17/09

A good example

A major goal of evolutionary genetics is to understand the molecular changes underlying phenotypic variation within and between species. The sequencing of whole genomes has made it possible to study not just individual mutations between orthologous sequences, but large-scale differences in gene complements between species. Such comparative genomic studies have found large disparities among organisms in the number of copies of genes involved in distinct cellular and developmental processes (e.g., [1,2]) and have even revealed the loss of entire gene families from individual lineages (e.g., [3,4]). **(In these first sentences, the authors introduce the overall problem that their work addresses – how molecular changes relate to phenotypic variation – and how comparative genomics can help.)** Though these studies begin to offer some insight into the molecular basis for phenotypic evolution, the timescales considered are often too long to provide evidence for the role of any single change (but see, e.g., [5–8]). **(But – and here's the gap in knowledge their work addresses – these studies haven't been able to account for the role of any single change.)** The sequencing of the genomes of 12 Drosophila species—whose most recent common ancestor (MRCA) lived only 60 million years ago [9]—offers the ability to study changes in the genomic complement of genes at an unprecedented resolution. **(Here's where they first introduce their solution and how it can help, although I think "unprecedented resolution" is perhaps a little vague.)**

Changes in the number of genes and proteins devoted to specific biological processes may arise in a number of different ways. First, gene duplication along any lineage will increase the number of genes, resulting in gene families containing multiple copies that are partially or completely overlapping in function. These gene duplicates may subsequently diverge in function by taking on new roles or by dividing up old roles [10–12]. There are now numerous examples in Drosophila of individual gene families with duplicates differentiated in both protein sequences (e.g., [13–16]) and gene expression domains (e.g., [17]). A second reason for differences in gene complement among species is that genes may be lost along a lineage when disabling mutations in them are not selected against. Such gene losses can even be directly advantageous [18], consistent with the so-called “less is more” hypothesis of Olson and colleagues [19]. Finally, the de novo creation of genes through various processes (e.g., [20–22])—while certainly quite rare—may contribute to lineage-specific differences in the number and function of constituent proteins. **(This paragraph gives the reader background on the types of genetic changes to be discussed throughout the paper, in very understandable language.)**

To provide a Drosophila-wide perspective on gene family evolution, we applied two different computational methods that estimate the rate and number of gene gains and losses. **(Here they state their specific purpose – again, a tad vague. They then describe their methods below.)** The first is a likelihood approach that estimates the average rate of gene gain and loss, the number of gains and losses on each branch of a phylogeny, and assigns p-values to large changes [23]. The second is the nonparametric gene tree/species tree reconciliation approach [24–27], which counts the number of gains and losses on each branch of the phylogeny without a specific probability model. While previous estimates of genome-wide rates of duplication in D. melanogaster [28,29] have offered a snapshot of one of the major mechanisms contributing to genome evolution, our analyses afford a wider view of this process. **(In the previous sentence, they make a general statement about the benefit of their solution, and then back the**

statement with a brief summary of their major findings in the next sentences.) We show that genes have been gained and lost in all species at varying rates; that several hundred gene families exhibit significantly large expansions or contractions in number, suggestive of adaptive natural selection; and that approximately equal numbers of gene families have either been lost completely in a species or are present only in a subset of the species considered here, information that can be used to improve the annotation of the D. melanogaster genome. **(Another benefit)** Throughout the analyses we examine the effect that heterogeneity in both assembly and annotation quality among the 12 genomes can have on evolutionary inferences.

Another good example

Halobacterium salinarum is a halophilic archaeon that thrives in extremely saline environments with salt concentrations reaching 4 M or higher. The organism is perhaps most well known for its retinal-protein bacteriorhodopsin (BR), which is a light-driven proton pump. BR is the only known nonchlorophyll structure that allows photosynthesis [1]. It is currently being developed for applications in optical security [2], optical data storage [3], and holography [4]. Accordingly, H. salinarum's photosynthetic capabilities are its most well studied aspects. For example, the 3D structure of BR has been resolved, and its complete catalytic cycle elucidated at the molecular level (reviewed in [5]). However, despite the focus on BR, photosynthesis is not the only means by which H. salinarum can generate energy. Respiration [6],[7] as well as the fermentation of arginine [8],[9] is another mechanism utilized by the organism. This bioenergetic flexibility makes the archaeon a good model system for investigating the interplay between different energy production modes. H. salinarum is also one of the few reported organisms that can use potassium gradients for long term energy storage in a battery-like manner [10]. **(In this paragraph, the authors provide the background readers need to appreciate their concluding point: that Halobacterium is a good model system for their study.)**

The metabolic network of an organism can be reconstructed from genomic, biochemical, and physiological data [11],[12],[13]. This network consists of the known and hypothesized reactions that take place within the organism, and is considered to be on a genome-scale when most or all of the genes with known metabolic function are included [14]. We, in a previous study, have reconstructed and proposed such a network for Halobacterium salinarum [15]. In addition to the immediate information gained from metabolic reconstructions, these networks can be analyzed to gain insights on emergent system properties through the use of appropriate computational methods. In this respect, the constraints-based framework has emerged as an important and convenient tool for modeling such systems because it does not require the detailed information typically required by full kinetic models. Rather, constraints-based models require only generally available physicochemical information such as stoichiometry, reversibility, energy balance, and, when available, reaction velocities [16],[17],[18]. **(This paragraph gives background on the features and advantages of the general approach they'll be taking: constraints-based modeling. In the next paragraph, they introduce their specific approach, which is flux balance analysis.)**

One of the methods available under the constraints-based framework is Flux Balance Analysis (FBA). Essentially, FBA uses linear optimization to find a flux distribution that maximizes a particular objective function, e.g., growth rate or ATP production [19],[20]. It has been shown that such optimality principles, within limits and under defined conditions, can describe the operation of metabolic networks, including the prediction of internal fluxes [21],[22]. Extensions

to FBA include hybrid models that introduce some degree of dynamics through the integration of time-variant input rates to the static model [15],[19],[20].

(Here's their purpose) Our aim in this study is two-fold. First, we set out to investigate the interplay between energy generation, nutrient utilization and biomass production under different bioenergetic modes. Second, we also analyzed the relationships between the different energy producing mechanisms of respiration, photosynthesis and fermentation themselves, which are typically examined individually. **(Here's the method)** To achieve these, we used a genome-scale metabolic network that connects the different aspects. Our results include several findings that are contrary to assumptions that are typically made; particularly with respect to the utilization of nutrients, and how the bioenergetic modes operate. **(Notice how the authors never explicitly state why their research is needed or what the problem is they hope to solve. However, they do allude to this when they say that several of their findings were contrary to the usual assumptions, a statement that certainly gets my attention.)** From a more methodological perspective, we also sought out to extend the existing framework for hybrid genome-scale metabolic models to handle biological systems where nutrient utilization and growth rates vary with time. Such changes in nutrient consumption, for example, can be the result of the differences between growth phases, or can arise from the interactions between the supplied metabolites. We demonstrate that the extended methodology not only accounts for such dynamics, but in several instances actually led to the identification of the underlying causes. **(Another teaser about the results that succeeds at piquing my interest.)**

A less good example

Recent advances in noninvasive imaging technology have allowed the creation of comprehensive whole-brain maps of the structural connections of the human cerebrum [1]–[7]. These maps have led to the quantitative characterization of various aspects of the network architecture of the brain, including degree distributions, small-world attributes, centrality and modularity. Comparative studies of structural and functional connectivity indicate that the presence of structural links between pairs of cortical regions is predictive of the occurrence of endogenously driven (resting-state) functional connectivity [4],[8],[9]. The mapping of structural connectivity has also enabled the construction of computational models of resting state activity [10],[11]. The direct comparison of empirically observed and computationally modeled resting state functional connectivity revealed a high degree of overlap, supporting the idea that large-scale structural brain networks do indeed shape and constrain endogenous patterns of functional connectivity [8].

The structural or functional robustness of networks has been investigated in a number of complex systems [12],[13], including biological networks [14]–[16] In the case of the brain, acute injuries from trauma, tumor, or stroke, as well as chronic or degenerative disturbances due to disease, correspond to node and edge deletions in the structural brain network. Many of the cognitive and behavioral effects of brain lesions are highly variable and their mechanistic origins remain difficult to discern. Nevertheless, lesions of specific brain regions are often associated with specific cognitive and behavioral disturbances, and lesions of some areas tend to have more severe effects than others [17]–[19]. Vulnerability analyses [20]–[24] of several non-human primate cortical networks suggest that lesion effects show regional specificity as well as non-local and distributed effects.

We describe a model of lesion effects in the human brain, based on a previously published map of structural connections [4] and a biophysical model of endogenous neural dynamics [8]. We investigate the effects of focal lesions (removing a spatially localized set of nodes and connections) on the endogenous dynamics of the remaining brain. We identify structural measures of brain connectivity that are predictive of the magnitude of the perturbations in the endogenous neural dynamics. We discuss our results in light of known behavioral and cognitive lesion effects. The computational and complex network approach taken in this paper provides a new link between localized structural damage of brain networks and global disruptions of dynamic interactions. **(The main issue I have with this introduction is that I'm not sure at the end why the model they've developed is needed. Are there drawbacks to the approaches others have used to make connections between the structural and functional attributes of the brain? Or is this one of the first methods of its kind? What new insights can their model offer into the effects of lesions, for example? In short, the authors cover in this introduction *what* they've done, but little to none of *why* they chose to do it, leaving readers to grasp the significance on their own.)**