Codon substitution models (CSMs) are commonly fitted to alignments of homologous protein-coding sequences with the objective of determining whether sites in the gene underwent positive selection. Under the standard paradigm such evidence is often assumed to be enough to conclude the gene evolved adaptively. CSMs are commonly validated using simulated alignments. A central theme of this dissertation is use of relatively realistic alignment-generating processes grounded in mutation-selection (MS) theory (Chapter 1). The MS framework permits sites to be evolved each on their own site-specific fitness landscape defined by a vector of fitness coefficients for the twenty amino acids. A novel MS alignment-generating process was used to show that evidence for variation in site-specific rate ratios (a.k.a. heterotachy) with episodic positive selection can be produced by episodic adaptive changes in site-specific fitness coefficients, consistent with the standard paradigm, but also by a second previously unrecognized process that I call non-adaptive shifting balance. This finding undermines sophisticated CSMs specifically designed to infer episodic adaptation by detecting heterotachy with episodic positive selection (Chapter 2).

Processes that tend to generate similar patterns in data are said to be confounded. Confounding can lead to a novel statistical pathology that I call phenomenological load. A series of novel CSMs fitted to alignments generated under a version of MS uniquely formulated to mimic real data were used to demonstrate that phenomenological load can lead to false biological conclusions. These analyses were accompanied by a novel method to assess the potential impact of phenomenological load on any given model parameter (Chapter 3). Confounding of adaptive and non-adaptive processes that generate heterotachy can be avoided by abandoning positive selection as an indicator of adaptation and instead using evidence of changes in site-specific amino acid fitnesses. This approach was realized by constructing the phenotype-genotype branch-site model (PG-BSM), a descendant of traditional branch-site models that combines alignment data with a discrete phenotype (i.e., contextual information) under a unified statistical framework. The PG-BSM was validated using extensive simulations and produced plausible results when applied to real data (Chapter 4). This dissertation ends with a discussion of implications of my findings (Chapter 5).