

Preface

Since the discovery of the main principles of Natural Selection by Charles Darwin, researchers have been pursuing the identification of the mechanisms that regulate organism's adaptations. It took an extraordinary sense of observation by Darwin and 150 years of research to realize about two of the main principles characterizing natural selection: the slow pace at which new adaptations emerge and that adaptive evolution, that is the emergence and fixation of new traits improving organism's fitness, is subtle and many times impossible to spot. The principles emerging from the study of natural selection have fascinated evolutionary biologists for over a century. The fingerprints of this force of adaptation has been sought by many and the existence of a designing hand behind it has been dismissed by the evolutionary morphological "monstrosities" emerging in nature. Another leap in the understanding of natural selection was made by the use of molecular data emerging from the sequencing of genes and the realization that mutations in protein-coding genes are of stochastic nature and that only those with neutral effects on protein's function survive in the genome. Kimura formalized in a mathematically elegant manner how important neutral mutations are in the generation of genetic variability. Today we know that such genetic variability is a reservoir of possible future adaptations as many variants can be enriched in the population to a point so that when combined in a single genome can lead to novel adaptations.

The neutral theory that rests on the importance of neutral mutations in the generation of genetic variability has led to a better understanding of the complexities of natural selection. Such complexities comprise more than two types of mutations: Neutral mutations, adaptive changes, and those mutations that are neutral but bear potential for novel adaptations (also known as exaptations). This intermingled nature of natural selection has made it difficult to disentangle neutral evolution signatures from those leading to adaptations.

In recent years, a plethora of genomes have been sequenced and new genomes are being released at an unprecedented pace. In addition to this, protein structures are enriching our databases and a bewildering set of network analyses are being performed at many different complexity levels, including protein-protein interactions, metabolic reactions, transcriptomics, etc. Integrating this information to infer the signatures of natural selection requires highly sensitive, fast, and accurate methods and computational tools to deal with the burst of data, together with an absolute understanding of how natural selection works. Unfortunately, finding evidence of adaptive evolution remains a difficult task as methods to deal with data and the models to reproduce the most important parameters governing the action of natural selection are in their infancy. Despite limitations, currently it is imperative to have a complete understanding of the models and methods to read the history of genes and organisms.

Chapters that guide us into the methods designed for the identification of natural selection form this book. It is divided into 9 chapters. The first chapter is a brief introduction to the importance of identifying natural selection and how can signatures of natural selection be found taking into account that distribution of individual's

fitness effects in a fitness landscape. Basic concepts on genes' structure, mutations and nucleotide and codon biases are given in the second chapter. These concepts are important for later understanding the ways in which deviations from such structures can help identifying signatures of adaptive evolution. In particular, mutation and nucleotide bias is brought forward as two main parameters in helping to unearth signatures of selection. The basic models of molecular evolution are described in the third chapter. Statistical ways to compare the goodness-of-fit of the models to the data given certain hypothesis are also detailed. Brushstrokes on the neutral theory of molecular evolution are provided in the fourth chapter. In this chapter, we also distinguish two levels of sequence analysis: inter-species and intra-population. The link between natural selection and the origination of novel functions is given in Chapter 5. Indeed, we show that divergence at the amino acid sequence level can lead to functional divergence, often leaving phylogenetic patterned fingerprints. Two methods of functional divergence are discussed in this chapter, one based on a maximum-likelihood framework and another based on a non-parametric approach. Chapter six calls to caution in the application of methods to identify natural selection as methods should account for recombination, which could have important biasing effects in the estimations of the parameters of selection. Once methods to identify natural selection have been examined, chapter seven puts forward the factors that influence the rate of evolution of protein-coding genes. Plausible hypotheses are presented in this chapter to explain the variation of protein's evolutionary rate. Selective pressures in the context of molecular networks are presented in Chapter 8. Finally, how do mutations interact and the consequences of these interactions in the emergence of adaptations at the molecular level are discussed and exemplified in Chapter 9.

This book collects the main principles underlying the action of natural selection and describes what methods are generally used to identify its action. In particular, a useful guide in the identification of the signatures of natural selection is given in this book, targeting students of the molecular biology field interested in evolution.

I would like to thank all the authors that have contributed to this book. I also wish to thank anonymous reviewers as well as the editors who made possible the publication of the book.

Mario A. Fares