**Codon Evolution: Mechanisms and Models**
**Gina M. Cannarozzi, Adrian Schneider (editors)**
**Oxford University Press, 2012**
**ISBN: 978-0-19-960116-5**
**Pages: 296**
**Price: £ 65.00**

The unifying principle in almost all of bioinformatics is sequence analysis: no matter if you are predicting the structure of proteins, analyzing the genetic variation in a population, or deciphering the evolutionary history of your favorite gene, your analysis hinges on looking at biological sequences and how they change over time, between species, or from gene to gene. Biological sequence analysis is the cornerstone.

It is no surprise, therefore, that a lot of effort is going into improving our tools for analyzing biological sequences in order to get as much and as accurate information as possible from our data. Especially today when large-scale sequencing projects are becoming commonplace in research groups all around the world, resulting in an unprecedented increase in available biological sequences, the need for proper computational tools for sequence analysis is greater than ever.

Comparative genomics and evolutionary studies can now include a huge number of species and genes, and obviously we want to get the most correct information about evolutionary relationships out of the available data. When looking at coding sequences we have access to information on both the DNA and protein level, and these signals can be combined by including the codon usage in protein coding genes. This can greatly improve your analysis as it is shown in numerous examples in the book *Codon Evolution: Mechanisms and Models*. Here, a selection of outstanding researchers present a thorough overview of this field covering both the theoretical underpinnings and practical applications. Understanding the evolution of codon usage over time, as well as the differences from species to species, we can get a much more complete understanding of sequence evolution. This has great impact on how to perform sequence analysis and, thus, on the field of bioinformatics as a whole.

The first part of the book, covering 12 chapters, describes different models of codon evolution. In chapter 1, A. Schneider and G.M. Cannarozzi introduce the subject matter and present notation and definitions used throughout the book. The chapter also briefly covers various widely used models, such as Markov models and maximum likelihood. This short introduction makes the book somewhat self-contained, although some background knowledge is useful to appreciate the details.

Chapter 2 by M. Anisimova describes parametric models of codon evolution and gives a thorough and well written introduction and overview of the field. This chapter covers a lot of ground from simply modeling codon frequencies through tests for selection to a discussion on modeling site dependencies. I enjoyed this chapter and found the review-like nature of it very useful. The next chapter by A. Schneider and G.M. Cannarozzi takes a different approach by describing empirical models of codon usage based on substitution matrices in the spirit of BLOSUM and PAM.

As has been the case in many other fields of bioinformatics, Bayesian statistics is also useful in the realm of codons, and in chapter 4 N. Rodrigue and N. Lartillot cover Monte Carlo approaches to codon substitution models. Using Markov chain Monte Carlo and simulated annealing under the well-known Metropolis-Hastings kernel (which has been used with success in other fields including structure prediction of RNA and protein, multiple alignment, and phylogeny), the authors present a framework for complex models where it would be impossible to numerically evaluate the likelihood function.

It is well-known that evolutionary rates (*e.g.* synonymous to non-synonymous substitutions) can vary between sites. Chapter 5 by H. Gu, K.S. Dunn and J.P. Bielawski presents the use of likelihood-based clustering to partition sites into distinct groups, each governed by a specific model, and they illustrate the utility of this approach on a large set of transmembrane proteins. In the following chapter, M. Anisimova and D.A. Liberles discuss how to detect natural selection in a statistical framework, and they give a very good introduction to the field. This is an interesting chapter, and especially the section on some of the common mistakes made in the field could become a useful resource. This ties in well with chapter 8, where G.A. Huttley and V.B. Yap show how important the assumptions in any given model are when estimating selection.

One of the most interesting chapters to me was chapter 7 by J.L. Thorne *et al.* Here, they review methods for comparing variation in protein coding genes between species within the realm of population genetics. In population genetics, most of the focus is on variation within a population but by taking a broader look and comparing inter-specific sequences, it becomes possible to look at mutations that became fixed long ago and which would not be visible within a single species.

After a very short chapter by M. Arenas and D. Posada on how to simulate the evolution of coding sequences, chapter 10 revisits the fact that we gain much more information by looking at codons rather than amino acids when analysing coding sequences. However, as S.A. Brenner points out, this leads to models that are hard to fully parameterize and he therefore discusses how to circumvent this problem by reducing the number of free parameters by grouping codons based on, for example, the observation that some codons are converted to other synonymous codons by purine to purine mutations. This discussion is important for accurately dating divergence times.

This leads nicely into the next chapter by B.S.W. Chang *et al.* which is a review of ancestral sequence reconstruction methods and models of divergence between clades. To finish off the first part of the book, chapter 12 by G. Aguileta and T. Giraud reviews studies on fungal genomes using codon models to investigate various aspects of their evolutionary history. This ties in well with the aforementioned increase in available sequence data due to next-generation sequencing technology.

The second and shorter part of the book describes different aspects of codon usage biases which is known to vary across species and between genes. The first chapter by A. Roth, M. Anisimova and G.M. Cannarozzi sets the stage by presenting an in-depth review of most (if not all) the various measures of codon bias that have been proposed to date. This is followed by a chapter by N.D. Rubinstein and T. Pupko who discuss the conservation of synonymous mutations, *i.e.* changes in codons that do not alter the encoded protein. The authors point to various reasons why synonymous mutations can affect fitness through *e.g.* translation efficiency, mRNA structure, or splicing signals.

The same theme is covered in chapter 15 where F. Supek and T. Šmuc discuss biases in codon usage and how to quantify these differences. They present both supervised and unsupervised methods for analyzing codon usage and show an application to genome data from archaea and bacteria. The following chapter by K. Zeng takes a population genetics view on codon usage bias by looking at synonymous polymorphisms within a group. This is an interesting review of two models where the author compares and contrasts the two.

The last chapter in the book by M.d.C. Santos and M.A.S. Santos discusses the various deviations from the standard genetic code and how these changes may occur naturally. This review is an interesting read and presents some interesting avenues for future research reaching back to the very root of the evolutionary tree connecting all life. All in all,

I find this book to be a thorough piece of work that covers a lot of ground and introduces a number of powerful tools that can be used when analyzing biological sequences. Since sequence analysis has such great importance in all fields of bioinformatics, and since evolution is the underlying principle in all of biology, getting a better understanding of coding sequence evolution on the codon level will be useful for most researchers in bioinformatics. This book highlights the role of codon usage in a number of different fields - protein structure, phylogeny, population genetics etc. - which shows how broadly applicable these models are, and how important it is to always have an evolutionary perspective on sequence analysis.

*Stinus Lindgreen*
*School of Biological Sciences,*
*University of Canterbury,*
*Christchurch, New Zealand*