

Multiple Factors Confounding Phylogenetic Detection of Selection on Codon Usage

Simon Laurin-Lemay,¹ Hervé Philippe,^{1,2} and Nicolas Rodrigue^{*,3}

¹Department of Biochemistry and Molecular Medicine, Robert-Cedergren Center for Bioinformatics and Genomics, Faculty of Medicine, Université de Montréal, Montréal, QC, Canada

²Centre de Théorisation et de Modélisation de la Biodiversité, Station d'Écologie Théorique et Expérimentale, UMR CNRS 5321, Moulis, Ariège, France

³Department of Biology, Institute of Biochemistry, and School of Mathematics and Statistics, Carleton University, Ottawa, ON, Canada

*Corresponding author: E-mail: nicolas.rodrigue@carleton.ca.

Associate editor: Tal Pupko

Abstract

Detecting selection on codon usage (CU) is a difficult task, since CU can be shaped by both the mutational process and selective constraints operating at the DNA, RNA, and protein levels. Yang and Nielsen (2008) developed a test (which we call CUYN) for detecting selection on CU using two competing mutation-selection models of codon substitution. The null model assumes that CU is determined by the mutation bias alone, whereas the alternative model assumes that both mutation bias and/or selection act on CU. In applications on mammalian-scale alignments, the CUYN test detects selection on CU for numerous genes. This is surprising, given the small effective population size of mammals, and prompted us to use simulations to evaluate the robustness of the test to model violations. Simulations using a modest level of CpG hypermutability completely mislead the test, with 100% false positives. Surprisingly, a high level of false positives (56.1%) resulted simply from using the HKY mutation-level parameterization within the CUYN test on simulations conducted with a GTR mutation-level parameterization. Finally, by using a crude optimization procedure on a parameter controlling the CpG hypermutability rate, we find that this mutational property could explain a very large part of the observed mammalian CU. Altogether, our work emphasizes the need to evaluate the potential impact of model violations on statistical tests in the field of molecular phylogenetic analysis. The source code of the simulator and the mammalian genes used are available as a GitHub repository (<https://github.com/Simonll/LikelihoodFreePhylogenetics.git>).

Key words: synonymous substitution, nonsynonymous substitution, CpG hypermutability, model violation.

Introduction

Elucidating the evolutionary factors influencing codon usage (CU) in protein-coding genes is a challenging endeavor. One difficulty resides in the complex nature of the CU genotype–phenotype relation. The nature of this relationship is affected by features of the mutational process, the degeneracy structure of the genetic code, synonymous codon recognition by tRNA molecules, and multiple other constraints operating at the DNA, RNA, and protein levels.

Unequivocally, the most important feature highlighted since the beginning of CU studies (Grantham et al. 1980), is the positive correlation found between the CU of highly expressed genes and the isoaccepting-tRNA pool composition (Ikemura 1985; Bulmer 1987; Akashi 1995; Sharp et al. 1995; Duret 2002). Presumably, the coevolution of CU and tRNA pool composition ensures the accuracy and efficiency of translation of highly expressed genes (reviewed in Quax et al. 2015).

The CU of highly expressed genes is usually postulated to result from a selection process on synonymous mutations, as investigated in large comparative frameworks across bacteria (Sharp et al. 2005) and eukaryotes (dos Reis and Wernisch

2009), and several more specific contexts, such as in *Drosophila* (Shields et al. 1988; Akashi 1994; Biene and Eyre-Walker 2006; Lawrie et al. 2013), in *Caenorhabditis elegans* (Duret and Mouchiroud 1999), in *Daphnia pulex* (Lynch et al. 2017), in vertebrates (Doherty and McInerney 2013), in the Euarchontoglires from mammals (Yang and Nielsen 2008), as well as in humans (Lavner and Kotlar 2005). While there might be mutational features that could partially account for the match between CU and the tRNA pool composition, “mutational pressures alone cannot explain why the more frequent codons [...] are those that are recognized by more abundant tRNA molecules” (Hershberg and Petrov 2008), probably because selection might be more effective at modulating the tRNA pool composition. Nonetheless, until a plausible mechanism is proposed, there generally remains “an understandable reluctance to accept selection at synonymous sites” (Chamary et al. 2006).

Exogenous gene expression is one area in which these issues have practical ramifications. Indeed, practitioners in this field remain without a clear rationalization of the methods for their activity: “A future challenge in studying the relation between coding sequences and protein production is

to perform a thorough comparative analysis of all currently known, and yet to be discovered, features of coding sequences that influence the translation process." (Quax et al. 2015). To date, two strategies are available to them for increasing the efficiency of exogenous genes expression: the first consists of expressing additional tRNA genes in the host cell to match the CU of the exogenous gene to be expressed; and the second consists of modifying the CU of the exogenous genes (reviewed in Quax et al. 2015). For instance, when using a bacterial system, one cannot achieve a worthwhile production of some desired human protein without modifying the codons used for encoding the desired amino acid sequence (Doble and Gummadi 2007). Nonetheless, there are still many instances in which the CU modifications for a particular sequence proceed by trial-and-error (Webster et al. 2017), without discernible reasons for the final CU that optimizes protein production.

Probabilistic models of molecular evolution have the potential to tease apart the determining factors of CU. The *mutation-selection* models (reviewed by McCandlish and Stoltzfus 2014), and particularly the phylogenetic ones, are well suited for testing hypotheses related to coding sequence evolution. The distinguishing features of these models is that they specify a substitution process with distinct parameterizations for the manner in which genetic variation is generated and for the fixation probability of genetic variants. In some cases, the models have specifically included considerations of selection on CU (McVean and Vieira 2001; Nielsen et al. 2007; Rodrigue et al. 2008; Yang and Nielsen 2008; Rodrigue and Philippe 2010; Pouyet et al. 2016).

The most well known of these models of codon substitution are those of Yang and Nielsen (2008). In their work, they propose a likelihood ratio test (LRT) to detect selection on CU, which we refer to as the CUYN test. The LRT is performed using two competing mutation-selection models: the null model is built with selection acting only on amino acid usage, assigning the same fitness to each degenerate codon encoding a particular amino acid; and the alternative model assigns a distinct fitness to each codon, and thus accounts for selection acting on both amino acid usage (by assigning a higher/lower fitness overall to the codons that encode a particular amino acid) and CU (by assigning a distinct fitness to each codon of an amino acid).

Yang and Nielsen (2008) found that much of the mammalian genes they tested rejected the null model, suggesting pervasive selection on CU. However, population genetics principles suggest that for organisms with a small effective population sizes, like mammals, selection is too inefficient to distinguish small effects conferred by certain synonymous mutations (reviewed in Chamary et al. 2006; Charlesworth 2009; Lynch et al. 2016). In contrast, selection is expected to be efficient within highly expressed proteins and in groups of fast-growing organisms (Sharp et al. 2010). Consequently, mammalian CU is expected to be mainly determined by mutation bias (reviewed in Chamary et al. 2006).

Several important features of the mutational process are unaccounted for in most codon substitution models. Among

these, a phenomenon known as *CpG hypermutability*, whereby certain mutations occur at higher rates in the context of the states at adjacent positions in the sequence, is considered pervasive in mammalian genomes (reviewed in Hodgkinson and Eyre-Walker 2011). The resultants of CpG hypermutability are numerous. For instance, the most used synonymous codon for Alanine in human, GCC, is four times more represented than GCG (Coleman et al. 2008), probably because the latter codon includes a highly mutable context (i.e., CpG), and is therefore short-lasting. In addition, on the basis of the *relative synonymous codon usage* (RSCU) metric, the most frequent adjacent codon pair for Alanine-Glutamine is expected to be GCC-GAR. In fact, however, this pair is the most underrepresented in humans (Coleman et al. 2008), probably because of the instability of the CpG context found at the interface of these two sequential codons. Another important feature that is known to bias the generation of genetic variation within mammalian genomes is *GC-biased gene conversion* (Duret and Galtier 2009): a phenomenon produced from meiotic recombination favoring transmission of G: C alleles over A: T alleles. Hypermutability of CpG contexts and GC-biased gene conversion are considered responsible for the isochore structures found in vertebrates (Duret and Galtier 2009; Munch et al. 2014; Mugal et al. 2015). These phenomena may have an impact on the test proposed by Yang and Nielsen (2008), in a manner that is very difficult to foresee. Indeed, in light of their surprising results, Yang and Nielsen (2008) point out that "the sensitivity of the LRT to violations of the assumed mutation model is not well understood and merits further research."

In this work, we use simulations to evaluate the effect of model violations on the accuracy of the CUYN test. We explore several types of violations at the mutation level, including one with dependence across codons, related to CpG hypermutability. We also evaluate how site-heterogeneous preferences on amino acids can affect the test. Numerous false positives are obtained under these model violations. While not excluding other potential features that can affect CU bias, our findings suggest that CpG hypermutability alone could explain the results of the CUYN test. Moreover, the test is even more unlikely to be reliable when yet other features (e.g., biased gene conversion) become involved, such as would be the case for real data sets. Altogether, given the small effective population sizes of mammals, and the potentially strong effect of known model violations, we find the CUYN test to be ill-suited for the detection of CU in the mammalian context.

Phylogenetic Mutation-Selection Models of Codon Substitution

Our study involves comparisons across several different substitution models, which we detail in this section. Branch lengths are free parameters for all models used in this work, while the tree topology is fixed. The models we use herein assume a point-mutation process, going from one codon *a* to another *b*, which differ at the *c*th position. In other words, *c* is an index of value 1, 2, or 3, indicating which of the three

nucleotide positions is different between codons a and b . Stop codons are also disallowed from the state space, leading to a 61×61 rate matrix Q . The rate matrix is quite sparse, however, since entries corresponding to nucleotide doublet or triplet events are set to 0. All nonnull, nondiagonal entries of the matrix are specified from two overall sets of parameters: those controlling a mutation process, and those controlling selection.

At the mutation level, nucleotide propensity parameters are invoked, defined as $\varphi = (\varphi_n)_{1 \leq n \leq 4}$, with $\sum_{n=1}^4 \varphi_n = 1$.

When using the M[HKY] settings, one parameter is introduced (i.e., $\kappa > 0$) to account for unequal rates between transitions and transversions. When using the M[GTR] settings, the exchangeabilities of each unique pair of nucleotides, m and n , are defined as $\varrho = (\varrho_{mn})_{1 \leq m, n \leq 4}$, with $\sum_{1 \leq m < n \leq 4} \varrho_{mn} = 1$. For some models, the transition rates within the CpG context (i.e., C to T and G to A) are modulated via a multiplicative parameter, λ_{CpG} .

At the level of selection, in the most elaborate settings, the specification of the model involves a set of K vectors, with each having 20 entries corresponding to amino acid preferences (also called *profiles*), denoted $\psi = (\psi_l^{(k)})_{1 \leq l \leq 20, 1 \leq k \leq K}$. The specification also involves an allocation variable denoted $z = (z_i)_{1 \leq i \leq N}$, where N is the length of the gene (in codons); for a given site i , z_i returns an index from 1 to K , specifying the amino acid profile operating at that site. The scaled selection coefficient (see Yang and Nielsen 2008) associated to a nonsynonymous change from codon a to b at site i is given as

$$S_{ab}^{(i)} = \ln \left(\frac{\psi_{f(b)}^{(z_i)}}{\psi_{f(a)}^{(z_i)}} \right), \quad (1)$$

where $f(a)$ returns an index, from 1 to 20, of the amino acid encoded by codon a . The value of $S_{ab}^{(i)}$, in turn, defines a fixation factor, denoted $h(S_{ab}^{(i)})$, and calculated as

$$h(S_{ab}^{(i)}) = \frac{S_{ab}^{(i)}}{1 - e^{-S_{ab}^{(i)}}}, \quad (2)$$

which will then be used directly in the rate matrix. The set of K amino acid profiles, the allocation variable z , and the value of K itself, are random variables within a Dirichlet process device (Rodrigue et al. 2010). We refer to this setting as S[NCatAA]. Of course, one can dispense with the Dirichlet process device, and simply use a single category of amino acid preferences, in which case, we drop the index on sites, i , from the notation, and refer to it as S[1CatAA].

Another setting at the level of selection replaces a single 20-dimensional vector with one that includes 61 values instead, one for each codon. The scaled selection coefficients and fixation factors are computed as before, although in this case, they are required for both synonymous and nonsynonymous events. We refer to this as the S[1CatCodon] setting.

Finally, we include a parameter (i.e., ω_*) on nonsynonymous rates, aimed at capturing deviations from the mutation-selection balance (Rodrigue and Lartillot 2017). The different models obtained from the combinations of mutation and selection parts are thus as follows:

- M[HKY]-S[1CatAA]: the mutation part of this model has four parameters (3 degrees of freedom) controlling nucleotide propensities, as well as a parameter (1 df) to distinguish transitions from transversions. The selection part of this model has a single category of amino acid preference (20 parameters and 19 df). The model was first described by Yang and Nielsen (2008) as FMutSel0, and is detailed below:

$$Q_{ab} = \begin{cases} \varphi_{b_c}, & \text{if syn. tr.}, \\ \varphi_{b_c} \kappa, & \text{if syn. ts.}, \\ \varphi_{b_c} \omega_* h(S_{ab}), & \text{if nonsyn. tr.}, \\ \varphi_{b_c} \kappa \omega_* h(S_{ab}), & \text{if nonsyn. ts.}, \end{cases} \quad (3)$$

where “syn.” and “nonsyn.” are short for “synonymous” and “nonsynonymous,” “tr.” and “ts.” are short for “transversion” and “transition,” and b_c returns an index, from 1 to 4, of the nucleotide found at the c th position in codon b .

- M[GTR]-S[1CatAA]: This model is nearly the same as M[HKY]-S[1CatAA], but has 6 distinct parameters (5 df) controlling the relative rate of each (unordered) pair of nucleotides. The model was also first described in Yang and Nielsen (2008), and is detailed in equation (4):

$$Q_{ab} = \begin{cases} \varrho_{a_c b_c} \varphi_{b_c}, & \text{if syn.}, \\ \varrho_{a_c b_c} \varphi_{b_c} \omega_* h(S_{ab}), & \text{if nonsyn.} \end{cases} \quad (4)$$

- M[HKY + λ_{CpG}]-S[1CatAA]: Similarly as before, this model only differs from equation (3) in having an additional parameter, λ_{CpG} , controlling the mutation rate of transitions in the CpG context. The model is detailed below:

$$Q_{ab} = \begin{cases} \varphi_{b_c}, & \text{if syn. tr.}, \\ \varphi_{b_c} \kappa, & \text{if syn. ts. non - CpG}, \\ \varphi_{b_c} \kappa \lambda_{CpG}, & \text{if syn. ts. CpG}, \\ \varphi_{b_c} \omega_* h(S_{ab}), & \text{if nonsyn. tr.}, \\ \varphi_{b_c} \kappa \omega_* h(S_{ab}), & \text{if nonsyn. ts. non - CpG}, \\ \varphi_{b_c} \kappa \omega_* h(S_{ab}) \lambda_{CpG}, & \text{if nonsyn. ts. CpG.}, \end{cases} \quad (5)$$

where “CpG” refers to a hypermutability context (assuming $\lambda_{CpG} > 1$) type of event.

- M[GTR + λ_{CpG}]-S[1CatAA]: Similarly as before, this model only differs from equation (4) in having one parameter, λ_{CpG} , controlling the mutation rate of the CpG context. The model is detailed in equation (6):

$$Q_{ab} = \begin{cases} \varrho_{ac} \varphi_{bc} \varphi_{b_c}, & \text{if syn. tr.,} \\ \varrho_{ac} \varphi_{bc} \varphi_{b_c}, & \text{if syn. ts. non - CpG,} \\ \varrho_{ac} \varphi_{bc} \varphi_{b_c} \lambda_{CpG}, & \text{if syn. ts. CpG,} \\ \varrho_{ac} \varphi_{bc} \varphi_{b_c} \omega_* h(S_{ab}), & \text{if nonsyn. tr.,} \\ \varrho_{ac} \varphi_{bc} \varphi_{b_c} \omega_* h(S_{ab}), & \text{if nonsyn. ts. non - CpG,} \\ \varrho_{ac} \varphi_{bc} \varphi_{b_c} \omega_* h(S_{ab}) \lambda_{CpG}, & \text{if nonsyn. ts. CpG.} \end{cases} \quad (6)$$

- M[HKY]-S[1CatCodon]: In contrast to M[HKY]-S[1CatAA], which in effect assigns all codons encoding a particular amino acid the same preference, this model has as distinct parameter for each codon (60 df). The model was first described by Yang and Nielsen (2008) as FMutSel, and is detailed in equation (7):

$$Q_{ab} = \begin{cases} \varphi_{b_c} h(S_{ab}), & \text{if syn. tr.,} \\ \varphi_{b_c} \kappa h(S_{ab}), & \text{if syn. ts.,} \\ \varphi_{b_c} \omega_* h(S_{ab}), & \text{if nonsyn. tr.,} \\ \varphi_{b_c} \kappa \omega_* h(S_{ab}), & \text{if nonsyn. ts.} \end{cases} \quad (7)$$

- M[GTR]-S[1CatCodon]: Similarly as before, this model only differs from the previous one in having 6 parameters controlling relative rates of nucleotide exchange, as described in Yang and Nielsen (2008), and equation (8):

$$Q_{ab} = \begin{cases} \varrho_{ac} \varphi_{bc} h(S_{ab}), & \text{if syn.,} \\ \varrho_{ac} \varphi_{bc} \omega_* h(S_{ab}), & \text{if nonsyn.} \end{cases} \quad (8)$$

- M[HKY]-S[NCatAA]: Of the same form as M[HKY]-S[1CatAA], this model allows for heterogeneity across sites for the amino acid preference by including multiple categories of amino acid preference, as described by Rodrigue et al. (2010), and equation (9):

$$Q_{ab}^{(i)} = \begin{cases} \varphi_{b_c}, & \text{if syn. tr.,} \\ \varphi_{b_c} \kappa, & \text{if syn. ts.,} \\ \varphi_{b_c} \omega_* h(S_{ab}^{(i)}), & \text{if nonsyn. tr.,} \\ \varphi_{b_c} \kappa \omega_* h(S_{ab}^{(i)}), & \text{if nonsyn. ts.} \end{cases} \quad (9)$$

- M[GTR]-S[NCatAA]: This model extends the previous one, to 6 parameters controlling relative rates of nucleotide exchange. This is the model studied in Rodrigue et al. (2010), and given in equation (10):

$$Q_{ab}^{(i)} = \begin{cases} \varrho_{ac} \varphi_{bc} \varphi_{b_c}, & \text{if syn.,} \\ \varrho_{ac} \varphi_{bc} \varphi_{b_c} h(S_{ab}^{(i)}) \omega_*, & \text{if nonsyn.} \end{cases} \quad (10)$$

Finally, we also used two intermediate parameterizations to the M[HKY] and the M[GTR] settings. The first of these is referred to as M[Tr], since it introduces distinct parameters to transversion exchangeabilities (i.e., where exchangeability parameters for A and C, A and T, C and G, and G and T

are distinct, but where a single exchangeability parameter is shared for transition events involving A and G and those involving C and T: Posada 2008), and the second is denoted M[Ts], since it includes distinct parameters for transition exchangeabilities (where a single exchangeability parameter is shared for events involving A and C, A and T, C and G, and G and T, whereas those involving A and G, as well as C and T, have distinct exchangeability parameters: Tamura and Nei 1993).

Results and Discussion

CUYN Test on Observed Data

We applied the CUYN test on 137 placental mammalian gene alignments (see Materials and Methods). At a level of 5%, all of the 137 genes analyzed showed a significant LRT (fig. 1A). Yang and Nielsen (2008) and Kessler and Dean (2014) also found a large proportion of genes with significant LRTs, but not 100%. This is likely due to the fact that the genes we study are longer and include more species, resulting in more evolutionary signal (i.e., substitutions) available for the models to learn their parameters, and therefore increases the statistical power of the test.

Protocols and Validations of the CUYN Test

In order to perform a verification of the CUYN test, we conducted several sets of simulations using parameter values obtained from the use of various mutation-selection models on 16 genes among the 137 previously subjected to the CUYN (see Materials and Methods). As a negative control, we generated a set of simulated alignments without selection on CU using parameter values obtained under the null models (i.e., M[HKY]-S[1CatAA] and M[GTR]-S[1CatAA]). The rate of false positives, at a significance level of 5%, is close to expectations for the analyses with M[HKY] mutation-level parameterization (fig. 1B: 5.2%). Unexpectedly, almost twice the amount of false positives (table 1: 8.8%) is obtained when using M[GTR] parameterization. To explore this surprising result, we increased the lengths of the branches of the tree over which our simulations were conducted by multiplying them with a parameter $\lambda_{TBL} = 50$. In these conditions, the rate of false positive approaches the expectation (table 1: 5.7%), suggesting that the overall signal present in the data when $\lambda_{TBL} = 1$ is weak, even if we used a richer taxon sampling than Yang and Nielsen (2008), and hence limiting the statistical accuracy of the CUYN test when using the M[GTR] setting.

As a positive control, we generated a set of simulated alignments with selection on CU using the parameter values obtained under the alternative models (i.e., M[HKY]-S[1CatCodon] and M[GTR]-S[1CatCodon]). In this case, all the simulated alignments tested show significant evidence for selection on CU. Altogether, for these control simulations, the CUYN test generally performs correctly, by detecting signatures of selection on CU when they are indeed present in the data, and not rejecting the null models in the absence of selection on CU.

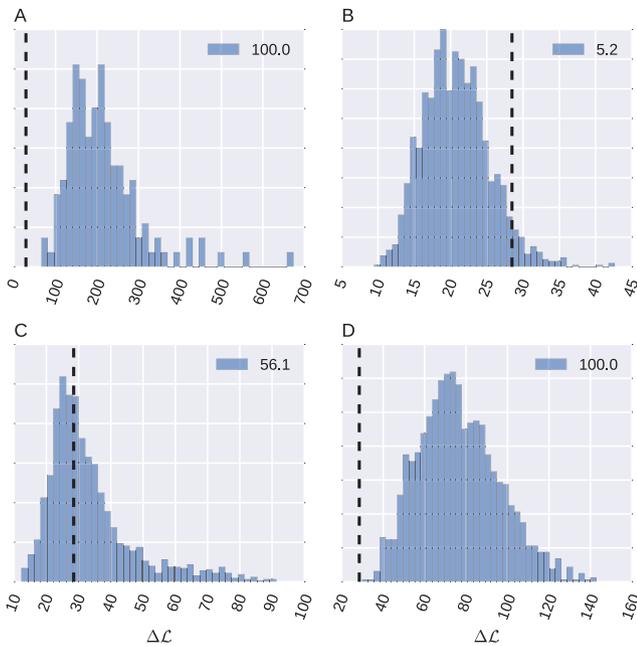


Fig. 1. Histograms of the log-likelihood differences ($\Delta\ell$) computed using the null and alternative hypotheses (i.e., $M[HKY]-S[1CatAA]$ and $M[HKY]-S[1CatCodon]$, respectively) on (A) 137 mammalian genes and on (B–D) simulated alignments. (B) Distribution of log-likelihood differences computed on the simulated alignments (100 replicates per set of parameter values), generated from parameter values obtained under $M[HKY]-S[1CatAA]$ on 16 genes. (C) Distribution of log-likelihood differences computed on the simulated alignment (100 replicates per set of parameter values), generated from parameter values obtained under $M[GTR]-S[1CatAA]$ on 16 genes. (D) Distribution of log-likelihood differences computed on the simulated alignments (100 replicates per set of parameter values), generated from the parameter values obtained under $M[HKY+\lambda_{CpG}=5]-S[1CatAA]$ on 16 genes. The vertical line corresponds to the threshold of significance (i.e., 28.47) at 5% with 41 degrees of freedom (i.e., 60–19 parameters) according to the χ^2 distribution. The proportion of significant analyses is shown at top right.

Model Violations at the Mutation Level

We next explored simulations done without selection on CU, using the parameter values obtained from analyses of the same 16 genes previously used. We found that alignments simulated with $M[GTR]$ and analyzed with $M[HKY]$ mutation-level settings generated 56.1% false positives (fig. 1C). This is a surprising result, since it involves one of the simplest model violations one could test. Moreover, it does not support the frequent use of the $M[HKY]$ in the analyses conducted with the PAML package (Yang 2007). When using the intermediate settings to $M[HKY]$ and $M[GTR]$, where transition exchangeabilities are set equal and transversion exchangeabilities are kept distinct, the $M[Tr]$ setting, or where transition exchangeabilities are kept distinct and transversion exchangeabilities are set equal, the $M[Ts]$ setting, we find that the heterogeneity of transversion rates has more impact than the heterogeneity of transition rates (i.e., 43.1% of false positives vs. 19.4%, respectively).

To study the properties of the CUYN test when dealing with increasingly information-rich data, we again increased

Table 1. Comparison of the Proportion of False Positives Detected When Computing CUYN Test on Simulated Alignments Generated Using Parameter Values Inferred from Various Mutation-Selection Models.

Models Used To Generate the Simulated Alignments	Mutational Model Used to Compute CUYN Test	CUYN Test (% positives)
$M[GTR]-S[1CatAA]$	$M[GTR]$	8.8
$M[GTR]-S[1CatAA]$	$M[HKY]$	56.1
$M[HKY]-S[1CatAA]$	$M[GTR]$	8.4
$M[HKY]-S[1CatAA]$	$M[HKY]$	5.2
$M[HKY]-S[1CatAA]+(\lambda_{TBL}=5)$	$M[HKY]$	4.2
$M[HKY]-S[1CatAA]+(\lambda_{TBL}=10)$	$M[HKY]$	5.3
$M[HKY]-S[1CatAA]+(\lambda_{TBL}=50)$	$M[HKY]$	5.7
$M[GTR]-S[1CatAA]+(\lambda_{TBL}=50)$	$M[GTR]$	5.7
$M[GTR]-S[1CatCodon]$	$M[GTR]$	100
$M[GTR]-S[1CatCodon]$	$M[HKY]$	100
$M[HKY]-S[1CatCodon]$	$M[GTR]$	100
$M[HKY]-S[1CatCodon]$	$M[HKY]$	100
$M[Ts]-S[1CatAA]$	$M[HKY]$	19.4
$M[Tr]-S[1CatAA]$	$M[HKY]$	43.1
$M[Ts]-S[1CatAA]+(\lambda_{TBL}=5)$	$M[HKY]$	24.8
$M[Tr]-S[1CatAA]+(\lambda_{TBL}=5)$	$M[HKY]$	74.1
$M[Ts]-S[1CatAA]+(\lambda_{TBL}=10)$	$M[HKY]$	27.0
$M[Tr]-S[1CatAA]+(\lambda_{TBL}=10)$	$M[HKY]$	77.1
$M[Ts]-S[1CatAA]+(\lambda_{TBL}=50)$	$M[HKY]$	33
$M[Tr]-S[1CatAA]+(\lambda_{TBL}=50)$	$M[HKY]$	85.3
$M[HKY+\lambda_{CpG}=0.1]-S[1CatAA]$	$M[HKY]$	99.9
$M[HKY+\lambda_{CpG}=0.5]-S[1CatAA]$	$M[HKY]$	67.1
$M[HKY+\lambda_{CpG}=2]-S[1CatAA]$	$M[HKY]$	76.9
$M[HKY+\lambda_{CpG}=4]-S[1CatAA]$	$M[HKY]$	99.8
$M[HKY+\lambda_{CpG}=5]-S[1CatAA]$	$M[HKY]$	100
$M[HKY+\lambda_{CpG}=8]-S[1CatAA]$	$M[HKY]$	100
$M[HKY+\lambda_{CpG}=16]-S[1CatAA]$	$M[HKY]$	100
$M[HKY]-S[NCatAA]$	$M[HKY]$	20.6
$M[HKY]-S[NCatAA]+(\lambda_{TBL}=5)$	$M[HKY]$	42.4
$M[HKY]-S[NCatAA]+(\lambda_{TBL}=10)$	$M[HKY]$	65.9
$M[HKY]-S[NCatAA]+(\lambda_{TBL}=50)$	$M[HKY]$	97.2

λ_{TBL} , multiplicative parameter used to dilate the branches length; λ_{CpG} , multiplicative parameter used to modulate the transition rates of the CpG context.

the lengths of the branches of the tree over which our simulations were conducted with a multiplicative parameter λ_{TBL} , taking on values 5, 10, and 50 (table 1). In one set of simulations, we increased the tree length when using $M[Tr]$ and $M[Ts]$ mutation-level specifications. We observed that even if the false positive rate increases with dilated branch lengths (e.g., 33% for $\lambda_{TBL}=50$ vs. 19.4% for $\lambda_{TBL}=1$ when using the $M[Ts]$) it does not reach the false positive rate detected when analyzing simulations made with the $M[GTR]$ (fig. 1C: 56.1%), suggesting that the main factor determining the number of false positives is the relative complexity of the model used to generate the simulated alignments compared to the model used for its analysis. This is confirmed by the fact that the accuracy of the test does not change for trees of greater length when $M[HKY]-S[1CatCodon]$ is used against $M[HKY]-S[1CatAA]$ (table 1: 4.2–5.7%).

Next, we simulated data with a new model that further modulates the mutation rates of the CpG context found within and across codon boundaries, using the multiplicative

parameter λ_{CpG} (see Materials and Methods). This new CpG context-dependent mutability, coupled with the null selection setting (i.e., $M[\text{HKY} + \lambda_{\text{CpG}}]\text{-S}[1\text{CatAA}]$), generates $\sim 100\%$ of false positives when λ_{CpG} reaches a value of only 5 (fig. 1D), even when CpG hypermutability only explains $\sim 10\text{--}20\%$ of the substitutions. We note that a value of $\lambda_{\text{CpG}} = 5$ is probably an underestimate in mammals (Hodgkinson and Eyre-Walker 2011). On the other hand, hypomutability of CpG (i.e., $\lambda_{\text{CpG}} < 1.0$) generates a similar amount of false positives, suggesting that any context-dependent mutation pattern may generate a high level of false positives for the CUYN test.

Impacts of Model Violations at the Level of Selection

A low rate of false positives (table 1: 20.6%) was recovered when analyzing alignments generated with heterogeneous amino acid fitness across sites. However, our alignments are highly conserved, with a small number of multiple amino acid substitutions at a given position (i.e., up to 68% of the positions have less than three nucleotide substitutions), leaving little opportunity to exhibit site-specific amino acid preferences. Increasing the total tree length by 5, 10, and 50 using the multiplicative parameter λ_{TBL} leads to an increase in the number of false positives (table 1: 42.4%, 65.9%, and 97.2%, respectively). Thus, site-heterogeneous amino acid preference can mislead results of the CUYN test, particularly at deep evolutionary scales or for fast evolving proteins.

CpG Hypermutability Can Largely Explain CU

The hypermutability of CpGs appears to be the model violation having the greatest impact on the rate of false positives when using the CUYN test. Given that CpG hypermutability is significant in mammals (Hodgkinson and Eyre-Walker 2011), we suspect that the selection detected on CU by Yang and Nielsen (2008) is largely due to this phenomenon. We designed a crude experimental protocol to explore the ability of the null model, $M[\text{GTR}]\text{-S}[1\text{CatAA}]$, and the alternative model, $M[\text{GTR}]\text{-S}[1\text{CatCodon}]$, to reproduce the codon frequencies of the 16 gene alignments. To do so, we computed a distance between the mean RSCU retrieved from batches of sequences generated under the stationarity of the different models and that same statistic computed from the real sequences of our alignments (refer to Materials and Methods for details on the procedure). As expected, the alternative model performed much better than the null model at predicting the codon frequencies, by rendering an RSCU closer to that of the true alignment (fig. 2).

We also explored the impact of CpG hypermutability on CU by generating sequences using parameter values from the stationarity of the null model obtained for each of the same 16 genes along with various values of λ_{CpG} , ranging from 0.1 to 20 (i.e., $M[\text{GTR} + \lambda_{\text{CpG}}]\text{-S}[1\text{CatAA}]$). Focusing on the RSCU retrieved from the batches of sequences generated under the latter conditions, we sought to find the values of λ_{CpG} that minimized the distance with the observed CU, leading to a rough approximation of the maximum likelihood value of λ_{CpG} itself. We then compare the resulting distance with

those retrieved from the null and the alternative models (fig. 2).

The distance to the true data greatly decreases when invoking λ_{CpG} within the null model to account for CpG hypermutability. Interestingly, the CU induced by $M[\text{GTR} + \lambda_{\text{CpG}}]\text{-S}[1\text{CatAA}]$ model appears to be close to that of the $M[\text{GTR}]\text{-S}[1\text{CatCodon}]$ model (fig. 2). In two cases, the rough optimization of λ_{CpG} brings the sequences drawn from the stationary distribution closer to the observed RSCU. This is particularly significant, since the $M[\text{GTR}]\text{-S}[1\text{CatCodon}]$ model involves 41 additional jointly optimized parameters (relative to the null), whereas the $M[\text{GTR} + \lambda_{\text{CpG}}]\text{-S}[1\text{CatAA}]$ involves only a single additional parameter, estimated crudely.

The batches of sequences simulated from the stationary distribution are independent one from the other, and thus computing the mean RSCU on them is justified. The sequences of the true alignments, however, are not independent of each other, such that averaging RSCU over them ignores their phylogenetic inertia. Therefore, we investigated how the rough estimation procedure for λ_{CpG} using subsets of sequences from the true alignment when computing the mean RSCU, to the point of using a single sequence. As expected, using only one sequence picked from the true alignment leads to a high variance of the minimized distance between the true data and the simulated data (supplementary fig. S1, Supplementary Material online). This is obviously a result of the low ratio between the degrees of freedom involved in the computation of the RSCU statistics (i.e., 41 df) and the number of codon states available in a single sequence drawn from the true alignment. In spite of this high variance, the general tendency of the results is similar: the distance between true and simulated data is comparatively low when invoking λ_{CpG} . Results obtained when averaging with 1/3 and 2/3 of the sequences present in the true alignments (supplementary figs. S2 and S3, Supplementary Material online) show progressively decreasing variance, as expected, and lead to very similar values of λ_{CpG} when compared to the results obtained with the entire alignment.

Figure 2 suggests that the selection detected on CU in mammals could be due to CpG hypermutability. The λ_{CpG} estimated from our rough procedure leads to values > 1 , as expected, and ranges between 4 and 12. These are probably underestimates, as the procedure is conditional on other parameter values obtained from the plain null model (i.e., $M[\text{GTR}]\text{-S}[1\text{CatAA}]$) rather than a proper joint estimation under the model with dependence across sites. The number of transition substitutions occurring in the CpG context is potentially limited, as the null model is probably attributing aspects of the hypermutability related to CpG contexts to the ω_* parameter and to the transition rates of the GTR mutation model. The CU of few genes (e.g., *EDEM2*) were not significantly altered by introducing the λ_{CpG} parameter, suggesting that CpG hypermutability is not among the most important factors determining CU for those genes. Other potential model violations should be further investigated, such as GC-biased gene conversion, and its potential interaction with CpG hypermutability.

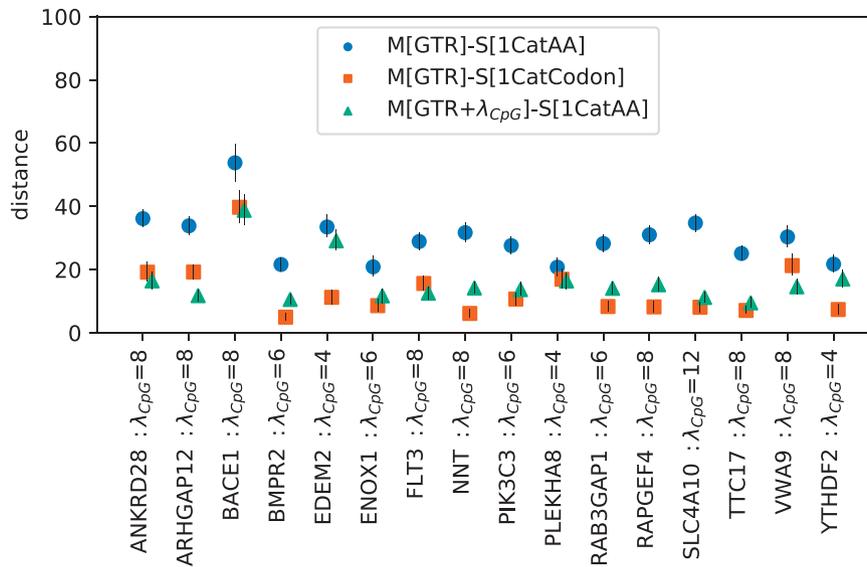


Fig. 2. Models comparison on the basis of their ability to predict CU (i.e., blue: M[GTR]-S[1CatAA], red: M[GTR]-S[1CatCodon] and green from the rough approximation procedure: M[GTR+ λ_{CpG}]-S[1CatAA]). The mean distances (i.e., dots) obtained for each specific analysis of the 16 mammalian genes are plot along with their associated error bars, that corresponds to 95% intervals, where the closest distances (2.5%) and the farthest distances (2.5%) were removed. The distances are computed between the mean RSCU retrieved independently from batches of sequences (i.e., 1,000 batches of 10 sequences) all generated under the stationarity of each specific model used, and the RSCU recovered from the true alignment. Only the results rendering the minimum mean distance from the rough optimization procedure are presented (i.e., triangles). The label of each gene analyzed is added as well as the values of λ_{CpG} that minimized the mean distance.

Low-Information Context of the CUYN Test on Mammals

Our study suggests that testing for CU with mammalian single gene alignments is a low-information context for ML inference. First, in analyzing the true data sets, most genes lead to parameter inferences at the boundaries of permissible values for nucleotide frequencies when in the M[GTR] settings (as detailed in the Materials and Methods). Second, when using real-data-obtained branch lengths, some simulation experiments uncovered an unexpectedly high rate of false-positives, which then approached expectations when conducting the simulation with increased branch lengths (i.e., increased statistical signal). Indeed, the attractiveness of ML inference comes from working with data sets that are sufficiently rich for large-sample theory to apply. With increasingly subtle features being introduced in the models (e.g., distinct exchangeabilities for each pair of nucleotides, or weak selection on CU), we risk falling into irregular conditions of ML estimation. A few recent studies have also uncovered such conditions (Baker et al. 2016; Mingrone et al. 2016). We note that Bayesian methods could be better-suited to dealing with such low-information settings, since they make no assumption of large-sample conditions; in the Bayesian framework, low-information contexts simply imply posterior distributions that are only very mild departures from prior distributions.

Conclusions and Future Directions

In spite of being formulated from mechanistic principles to account for CU, the parameters introduced in the alternative model of the CUYN test appear to be absorbing other

features of the evolutionary process than those intended, in a manner that is a statistically significant departure from the values expected under the null model. We have shown that violations on both the mutation and the selection aspects of the models can greatly impact the accuracy of the CUYN test, to a point where it may not be useful for the analysis of mammalian genes. We have shown that the frequent use of the M[HKY] setting should be avoided when the M[GTR] is available, since this oversimplified mutational parameterization alone can generate an important amount of false positives. We have also shown that two important aspects of evolution, CpG hypermutability, and, to a much lesser extent, site-heterogeneous preferences on amino acids, can mislead the CUYN test.

Simulation studies are key to evaluating the robustness of probabilistic inferences with respect to model violations that are well known to be pervasive in the data at hand. If the model-based test appears to be robust to violations (e.g., in some instances, site-heterogeneous selection on amino acids), its use is reassuring. If not (e.g., CpG hypermutability), the general reliability of the test is in doubt.

In some contexts, addressing model violations directly may be beneficial. In the case of CpG hypermutability, the most obvious modeling expansion from the work we have conducted here would be to include the λ_{CpG} parameter within the overall inference. While Monte Carlo methods for implementing such site-dependent models have been developed in Bayesian contexts (Rodrigue and Lartillot 2012), our crude approximation based on simulations also suggests the use of *Approximate Bayesian Computation* (Csillery et al. 2010). In any case, with such a formulation, it is hoped, the

parameters introduced by the alternative model would no longer need to absorb CpG effects, and would presumably be “freed” for their intended purpose. Or, the introduced parameters could absorb other model violations, still. Indeed, the issue is much more difficult when model violations are poorly characterized (e.g., selective constraints on mRNA secondary structure) or even unknown, but the results we present here suggest that this should be a major part of research efforts.

Materials and Methods

Data Preparation and Tree Inference

We queried the OrthoMaM database, version 9 (Douzery et al. 2014), to retrieve all gene alignments where all of the 43 species available in the database were present, which leads to a collection of 137 mammalian alignments. We then arbitrarily removed from these the 4 nonplacental mammals that are part of OrthoMaM, in order to focus our study on the 39 placental mammals. Each alignment was treated with HmmCleaner (Amemiya et al. 2013) and Gblocks (Talavera and Castresana 2007) to remove structural annotation errors and poorly aligned regions, respectively. Gblocks was used under a nonstringent setting, resulting in very few positions being removed (<1.5%). Our analyses were conducted with a fixed tree topology, as obtained under the CAT model (Lartillot and Philippe 2004) implemented in PhyloBayes-MPI (Lartillot et al. 2013) on the amino acid concatenation of the 137 alignments (supplementary fig. S4, Supplementary Material online). When compared to a recent review on the topic (Foley et al. 2016), our topology is almost identical, with the only difference being in the relationships at the base of Laurasiatheria, which precisely corresponds to very short internal branches. The impact of such topology variation is therefore expected to be negligible.

Inferring Model Parameters

Parameters were estimated on the 137 genes with PhyloBayes-MPI (Rodrigue and Lartillot 2014) for the M[HKY]-S[1CatAA], M[GTR]-S[1CatAA], M[HKY]-S[1CatCodon], M[GTR]-S[1CatCodon], M[HKY]-S[1NCatAA], and M[GTR]-S[1NCatAA] models. Parameters were estimated separately on the 137 genes with CodeML for the M[HKY]-S[1CatAA], M[GTR]-S[1CatAA], M[HKY]-S[1CatCodon], and M[GTR]-S[1CatCodon] models. With CodeML, we found that 121 genes gave extreme parameter values when applying the M[GTR]-S[1CatCodon], for instance, with one of the nucleotide propensity parameters at a value close to 0 or 1. Yang and Nielsen (2008) mentioned that in comparing M[HKY]-S[1CatCodon] and M[GTR]-S[1CatCodon] “[...] the estimates of codon-fitness parameters for the concatenated data under the 2 mutation models are very different (results not shown). This is the case even though both mutation models predicted very similar codon frequency parameters, which closely match the observed frequencies. Our estimates of the selection coefficients are affected by the mutation model. Thus, we found that the LRT is somewhat insensitive to the assumed mutation model but the estimates of codon fitness parameters are.” Obviously, the

mutation-level parameterization had to compensate in order to obtain similar codon frequencies. In order to keep computation-time of our simulation study manageable, we chose to work with the parameter values obtained from a subsample of alignments, and decided to work the 16 genes that did not exhibit this issue (this still implies 100×16 simulations for each condition, and several times as many ML inferences and Bayesian MCMC runs). Since the results based on CodeML and PhyloBayes-MPI for the M[HKY]-S[1CatAA], M[GTR]-S[1CatAA], M[HKY]-S[1CatCodon], and M[GTR]-S[1CatCodon] models were similar, we only present results based on CodeML (see supplementary tables S1 and S2, Supplementary Material online).

Simulation Program

We wrote a simulation program that evolves sequences along a phylogenetic tree, generating DNA alignments with various mutation-selection models (see the previous section). Our simulation software can take as input the outputs of CodeML or of PhyloBayes-MPI when used with mutation-selection models.

In our implementation, sequences are evolved in a jump-chain manner, substitution by substitution, starting from the root and traversing all branches to the tips of the tree. We first sample a sequence from the stationary distribution as it would be under the site-independent model (with $\lambda_{\text{CpG}} = 1$). This sequence is then evolved for a large number of events, to reach stationarity under the site-interdependent framework, before being set as the starting state at the root of the tree. The simulation along each branch of the tree proceeds by first drawing a waiting time from an exponential distribution with a parameter corresponding to the sum of rates to all nearest-neighbors of the current sequence. If the waiting time drawn does not go beyond the length of the current branch, the nature of the event is drawn, with a probability proportional to the rate to each nearest-neighbor sequence. From this new state, and time-point along the branch, another waiting time is drawn as before, until the waiting time sampled goes beyond the length of the branch; once it does, the state at the descendant node is set to the current sequence, and the simulation procedure splits and continues independently along the daughter branches. This continues until sampling waiting times beyond the final terminal branches, thereby yielding the states in the simulated alignment set. The source code of the simulation software is available as a GitHub repository (<https://github.com/Simonll/LikelihoodFreePhylogenetics.git>; last accessed April 3, 2018).

Simulated Data Sets

We simulated data sets by using the parametric bootstrapping and the posterior predictive procedures. The different experimental conditions were defined by the models used. Experimental conditions were replicated at two levels: 100 simulated alignments were generated from the parameter values retrieved under the use of the different models on the 16 genes. The root position was set at 90% of the branch from the Afrotheria to the (Xenarthra + Laurasiatheria + Euarchontoglires). In some

cases, we manipulated the values of the model parameters prior to the simulation procedures: (1) we set the transversion rates to the average of nucleotide exchangeability parameters of transversions obtained from a GTR-based analysis (i.e., $M[Ts]$); (2) we did as in (1), but for the transition rates rather than transversion rates (i.e., $M[Tr]$); (3) we increased the total tree length by 5, 10, and 50 times using a multiplicative parameter, λ_{TBL} ; and (4) we modulated the mutation rate in the CpG context with a multiplicative parameter, λ_{CpG} , using different values (i.e., 0.1, 0.5, 1, 2, 4, 5, 8, 16).

Approximating the Observed CU

We developed a rough methodology to help build our intuition about how CpG hypermutability could impact CU. We drew sequences from the stationary distribution (i.e., 10,000 sequences per set of parameter values) to study how the different models predict the codon frequencies of the true alignments. The comparison of the CU obtained under the different models is achieved by retrieving the squared Euclidean distance (i.e., eq. 11) between the mean RSCU computed from 1,000 batches of 10 sequences drawn from each model stationary distribution and the mean RSCU obtained from the true alignment, or subsets thereof.

$$CU_{\text{dist}} = \sum_{a=1}^{61} (\log_2(z_a) - \log_2(y_a))^2, \quad (11)$$

where y and z are the mean codon state frequencies normalized by amino acid (i.e., RSCU) obtained from the batches of simulated sequences and the true sequence(s), respectively.

We investigated the effect of computing the distances when including different numbers of sequences randomly picked from the true alignment. This was achieved by randomly sampling various proportion of sequences (i.e., 1/39, 1/3, and 2/3) from each alignment studied; note that 39 sequences are present in each of the alignments studied. Our expectation was that the number of codon states present in a sequence, even if the sequences contain >500 codons, will limit the accuracy of the rough procedure we designed here, as the RSCU involves 41 degrees of freedom. We compared the ability of the null and the alternative models (i.e., $M[GTR]-S[1CatAA]$ and $M[GTR]-S[1CatCodon]$) to predict the CU for each of the 16 genes, but also performed the comparison with the CpG context model $M[GTR + \lambda_{CpG}]-S[1CatAA]$, incorporating different λ_{CpG} values (i.e., 0.25, 0.5, 0.75, 1, 2, 4, 8, 10, 12, 14, 16, 18, 20). We also tracked the proportion of CpG context substitutions occurring during the simulations.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

The authors would like to heartily thank Nicolas Lartillot for all the help provided during the project and commentary on the manuscript. We also thank three anonymous reviewers

for their constructive comments. This work was supported by the French Laboratory of Excellence project entitled TULIP (ANR-10-LABX-41; ANR-11-IDEX-0002-02), and by the Natural Sciences and Engineering Research Council of Canada. Computations were done on the Mammouth-série supercomputer from the Université de Sherbrooke, managed by Calcul Québec and Compute Canada. The operation of this supercomputer is funded by the Canada Foundation for Innovation (CFI), the ministère de l'Économie, de la science et de l'innovation du Québec (MESI) and the Fonds de recherche du Québec - Nature et technologies (FRQ-NT). S.L.L. is the recipient of a Fonds de la Recherche en Santé Québec (FRSQ) Graduate Scholarship.

References

- Akashi H. 1994. Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics* 136:927–935.
- Akashi H. 1995. Inferring weak selection from patterns of polymorphism and divergence at silent sites in drosophila DNA. *Genetics* 139(2):1067–1076.
- Amemiya C, Alfoldi J, Lee A, Fan S, Philippe H, MacCallum I, Braasch I, Manousaki T, Schneider I, Rohner N, et al. 2013. The African coelacanth genome provides insights into tetrapod evolution. *Nature* 496(7445):311–316.
- Baker J, Dunn K, Mingrone J, Wood B, Karpinski B, Sherwood C, Wildman D, Maynard T, Bielawski J. 2016. Functional divergence of the nuclear receptor NR2C1 as a modulator of pluripotentiality during hominid evolution. *Genetics* 203(2):905.
- Bierne N, Eyre-Walker A. 2006. Variation in synonymous codon use and DNA polymorphism within the *Drosophila* genome. *J Evol Biol* 19(1):1–11.
- Bulmer M. 1987. Coevolution of codon usage and transfer-RNA abundance. *Nature* 325(6106):728–730.
- Chamary J, Parmley J, Hurst L. 2006. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat Rev Genet* 7(2):98–108.
- Charlesworth B. 2009. Effective population size and patterns of molecular evolution and variation. *Nat Rev Genet* 10(3):195–205.
- Coleman JR, Papamichail D, Skiena S, Futcher B, Wimmer E, Mueller S. 2008. Virus attenuation by genome-scale changes in codon pair bias. *Science* 320(5884):1784–1787.
- Csillery K, Blum M, Gaggiotti O, Francois O. 2010. Approximate Bayesian Computation (ABC) in practice. *Trends Ecol Evol* 25(7):410–418.
- Doble M, Gummadi S. 2007. Biochemical engineering. New Delhi: Prentice-Hall of India.
- Doherty A, McInerney J. 2013. Translational selection frequently overcomes genetic drift in shaping synonymous codon usage patterns in vertebrates. *Mol Biol Evol* 30(10):2263–2267.
- dos Reis M, Wernisch L. 2009. Estimating translational selection in eukaryotic genomes. *Mol Biol Evol* 26(2):451–461.
- Douzery E, Scornavacca C, Romiguier J, Belkhir K, Galtier N, Delsuc F, Ranwez V. 2014. OrthoMaM v8: a database of orthologous exons and coding sequences for comparative genomics in mammals. *Mol Biol Evol* 31(7):1923–1928.
- Duret L. 2002. Evolution of synonymous codon usage in metazoans. *Curr Opin Genet Dev* 12(6):640–649.
- Duret L, Galtier N. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu Rev Genomics Hum Genet* 10:285–311.
- Duret L, Mouchiroud D. 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc Natl Acad Sci U S A* 96(8):4482–4487.
- Foley N, Springer M, Teeling E. 2016. Mammal madness: is the mammal tree of life not yet resolved? *Philos Trans R Soc Lond B Biol Sci* 371: 20150140 DOI: 10.1098/rstb.2015.0140.

- Grantham R, Gautier C, Gouy M, Mercier R, Pavé A. 1980. Codon catalog usage and the genome hypothesis. *Nucleic Acids Res.* 8(1):r49–r62.
- Hershberg R, Petrov D. 2008. Selection on codon bias. *Annu Rev Genet.* 42:287–299.
- Hodgkinson A, Eyre-Walker A. 2011. Variation in the mutation rate across mammalian genomes. *Nat Rev Genet.* 12(11):756–766.
- Ikemura T. 1985. Codon usage and transfer-RNA content in unicellular and multicellular organisms. *Mol Biol Evol.* 2(1):13–34.
- Kessler M, Dean MD. 2014. Effective population size does not predict codon usage bias in mammals. *Ecol Evol.* 4(20):3887–3900.
- Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol.* 21(6):1095–1109.
- Lartillot N, Rodrigue N, Stubbs D, Richer J. 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst Biol.* 62(4):611–615.
- Lavner Y, Kotlar D. 2005. Codon bias as a factor in regulating expression via translation rate in the human genome. *Gene* 345(1):127–138.
- Lawrie D, Messer P, Hershberg R, Petrov D. 2013. Strong purifying selection at synonymous sites in *D. melanogaster*. *PLoS Genet.* 9(5):e1003527.
- Lynch M, Ackerman M, Gout J, Long H, Sung W, Thomas W, Foster P. 2016. Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet.* 17(11):704–714.
- Lynch M, Gutenkunst R, Ackerman M, Spitze K, Ye Z, Maruki T, Jia Z. 2017. Population genomics of *Daphnia pulex*. *Genetics* 206(1):315–332.
- McCandlish D, Stoltzfus A. 2014. Modeling evolution using the probability of fixation: history and implications. *Q Rev Biol.* 89(3):225–252.
- McVean G, Vieira J. 2001. Inferring parameters of mutation, selection and demography from patterns of synonymous site evolution in *Drosophila*. *Genetics* 157:245–257.
- Mingrone J, Susko E, Bielawski J. 2016. Smoothed bootstrap aggregation for assessing selection pressure at amino acid sites. *Mol Biol Evol.* 33(11):2976.
- Mugal C, Arndt P, Holm L, Ellegren H. 2015. Evolutionary consequences of DNA methylation on the GC content in vertebrate genomes. *G3 (Bethesda)* 5:441–447.
- Munch K, Mailund T, Dutheil J, Schierup M. 2014. A fine-scale recombination map of the human-chimpanzee ancestor reveals faster change in humans than in chimpanzees and a strong impact of GC-biased gene conversion. *Genome Res.* 24(3):467–474.
- Nielsen R, Bauer DuMont V, Hubisz M, Aquadro C. 2007. Maximum likelihood estimation of ancestral codon usage bias parameters in *Drosophila*. *Mol Biol Evol.* 24:228–235.
- Posada D. 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol.* 25(7):1253–1256.
- Pouyet F, Bailly-Bechet M, Mouchiroud D, Gueguen L. 2016. SENCA: a multilayered codon model to study the origins and dynamics of codon usage. *Genome Biol Evol.* 8(8):2427–2441.
- Quax TE, Claassens N, Söll D, van der Oost J. 2015. Codon bias as a means to fine-tune gene expression. *Mol Cell* 59(2):149–161.
- Rodrigue N, Lartillot N. 2012. Monte Carlo computational approaches in Bayesian codon substitution modeling. In: Cannarozzi GM, Schneider A, editors. Codon evolution: mechanisms and models. Oxford: OUP. p. 45–59.
- Rodrigue N, Lartillot N. 2014. Site-heterogeneous mutation-selection models within the PhyloBayes-MPI package. *Bioinformatics* 30(7):1020–1021.
- Rodrigue N, Lartillot N. 2017. Detecting adaptation in protein-coding genes using a bayesian site-heterogeneous mutation-selection codon substitution model. *Mol Biol Evol.* 34(1):204–214.
- Rodrigue N, Lartillot N, Philippe H. 2008. Bayesian comparisons of codon substitution models. *Genetics* 180(3):1579–1591.
- Rodrigue N, Philippe H. 2010. Mechanistic revisions of phenomenological modeling strategies in molecular evolution. *Trends Genet.* 26(6):248–252.
- Rodrigue N, Philippe H, Lartillot N. 2010. Mutation-selection models of coding sequence evolution with site-heterogeneous amino acid fitness profiles. *Proc Natl Acad Sci U S A.* 107(10):4629–4634.
- Sharp P, Averof M, Lloyd A, Matassi G, Peden J. 1995. DNA-sequence evolution – the sounds of silence. *Philos Trans R Soc Lond B Biol Sci.* 349(1329):241–247.
- Sharp P, Bailes E, Grocock R, Peden J, Sockett RE. 2005. Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Res.* 33(4):1141–1153.
- Sharp P, Emery L, Zeng K. 2010. Forces that influence the evolution of codon bias. *Philos Trans R Soc Lond B Biol Sci.* 365(1544):1203–1212.
- Shields D, Sharp P, Higgins D, Wright F. 1988. Silent sites in drosophila genes are not neutral – evidence of selection among synonymous codons. *Mol Biol Evol.* 5:704–716.
- Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 56(4):564–577.
- Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol.* 10(3):512–526.
- Webster G, Teh A, Ma J. 2017. Synthetic gene design: The rationale for codon optimization and implications for molecular pharming in plants. *Biotechnol Bioeng.* 114(3):492–502.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24(8):1586–1591.
- Yang Z, Nielsen R. 2008. Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. *Mol Biol Evol.* 25(3):568–579.