Genview and Gencode: a pair of programs to test theories of genetic code evolution

T. Andrew Ronneberg, Stephen J. Freeland* and Laura F. Landweber

Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA

Received on October 4, 2000; revised and accepted on November 23, 2000

ABSTRACT

Summary: Genview and Gencode are tools for testing the adaptive nature of a genetic code under different assumptions about patterns of genetic error and the nature of amino acid similarity. Genview provides a user friendly, point-and-click interface by which a user may reproduce and extend analysis of the adaptive properties of the standard genetic code or any of its secondary derivatives. Genview is a graphical user interface (GUI) program which runs on Linux, Unix® and Microsoft Windows® platforms and is based on the GTK+ toolkit. Genview outputs ASCII configuration files which are interpreted by Gencode to perform an analysis. Gencode is available for the same platforms as Genview.

Availability: Online documentation can be found at http://rnaworld.princeton.edu/genview/documentation. The source code can be downloaded from http://rnaworld.princeton.edu/genview/

Contact: sfreelan@princeton.edu

INTRODUCTION

A variety of recent research publications have presented evidence that the arrangement of codon/amino-acid assignments within the standard genetic code results from Darwinian selection to minimize the phenotypic impact of genetic error (Haig and Hurst, 1991; Freeland and Hurst, 1998a,b; Ardell, 1998; Freeland et al., 2000b). However, this interpretation remains debated (Knight et al., 1999; Di Giulio and Medugno, 2000; Di Giulio, 2000; Freeland et al., 2000a) and the analyses presented to date are far from exhaustive.

The extent of code adaptation is well suited to computational analysis by quantifying a code’s susceptibility to genetic errors and comparing this value to the distribution of values obtained for alternative codes. In a Gencode analysis, the phenotypic impact of errors is specified by a $\Delta_{\text{code}}$ function, which is determined by the average difference between amino acids whose codons differ by a single nucleotide. This function is calculated for the code of interest and for sample(s) of alternative plausible codes. Genview allows the user to specify the precise $\Delta_{\text{code}}$ function in terms of specific patterns of mutation and mistranslation and to define the rules by which alternative codes are generated in order to explore a genetic code’s ability to minimize the impact of mutational and translational errors.

PROGRAM OVERVIEW

The user begins by specifying the code structure to be analyzed. The user may define any genetic code by reassigning any codon or group of codons to any of the 20 amino acids specified by the canonical genetic code, to the translation termination (‘TER’) identity, or to one of five user-defined amino-acids (see Figure 1). The use of nonstandard amino acids may shed light on why the ‘universal’ genetic code and all naturally occurring variants only encode the canonical 20 amino acids.

Two kinds of similarity measures can be used to test the deleterious impact of genetic errors. The first type is based on experimentally-determined physiochemical properties like polarity, size, and iso-electric point (Haig and Hurst, 1991). The second type are substitution matrices, which are calculated from the frequency of amino acid substitutions in deep branching homologous proteins (Ardell, 1998; Freeland et al., 2000b).

Genetic errors may take many forms. Genview allows the user to explore the impact of different error biases on any genetic code arrangement. For example, one well-known bias is the tendency for transition mutations (G$\leftrightarrow$A, C$\leftrightarrow$T) to occur more frequently than transversion mutations (e.g. A$\leftrightarrow$C, U$\leftrightarrow$G). The user may specify a range of transition–transversion biases, or perform a more refined analysis by specifying the frequency of mutation from each nucleotide to the other three nucleotides (see Graur and Li (2000) for an introduction to nucleotide substitution patterns). In addition, the user may specify the probability that translational mistakes...
Fig. 1. Altering codon assignments. When a codon is selected, a menu will appear allowing the user to change the amino acid codon assignment. Any change to the amino acid codon assignment will change the assignment for all amino acids in the group.

will occur at each of the three codon positions. Typically, previous analyses have generated and tested large samples of biologically plausible alternative codes in order to estimate the probability that chance alone would produce a code of equal adaptiveness (Haig and Hurst, 1991; Freeland and Hurst, 1998a,b). Genview allows the user to specify the number of alternative codes generated or, alternatively, to use a heuristic search algorithm to estimate the optimal pattern of codon assignments under a given set of assumptions. This latter option may be used to derive ‘percentage distance minimization’ estimates: an alternative measure of code adaptation (Di Giulio, 1999; Freeland et al., 2000b).

In addition, the user may specify the rules by which the alternative codes are generated. For example, codon assignment is randomized at the level of a ‘codon group’ which specifies a subset of codons with the same amino acid assignment. In the default setting, there are 21 groups, in which all codons specifying the same amino acid are initially assigned to the same group (the TER codons are also assigned to a group). The user may wish to split some of these groups, for example to divide the Ser codons into two groups, so that their amino acid assignments are randomized independently. Groups may be split to the level of individual codons. In addition, the user can specify which groups can exchange amino acid identities. These ‘meta-groups’ are useful for generating variant codes that are constrained by biosynthetic relationship of their amino acids (Freeland and Hurst, 1998b; Freeland et al., 2000b).

Lastly, the user can specify codons that have a fixed amino acid identity in the generated variants. The full scope of possible analyses is described in the User Guide that accompanies the program.

CALCULATIONS

The time it takes to run a Gencode analysis depends almost exclusively on the sample size requested for each point in parameter space multiplied by the number of such points to be tested. For reference, an analysis of 100,000 variant codes with transition/transversion biases ranging from 1 to 20 and modular powers ranging from 1 to 5 (i.e. generating and measuring errors within 10 million alternative codes, as discussed in the accompanying documentation) takes approximately 4 h to run on a 333 MHz Intel PII machine.

REFERENCES