An algorithm for mapping positively selected members of quasispecies-type viruses

ABSTRACT

Selection mapping is a fundamental improvement over earlier methods (e.g., dN/dS) that identify positive selection at codons but do not identify which amino acids at these codons confer selective advantage. Using QUASI's selection maps, we characterize the selected mutational landscapes of influenza A H3 hemagglutinin, HIV-1 reverse transcriptase, and HIV-1 gp120.

INTRODUCTION

Background Antigenic drift and the generation of viral quasispecies Some RNA viruses form a quasispecies--a set of related viral variants that coexist in field populations and even within single infected individuals (reviewed in). The emergence of immunologically distinct members of a viral quasispecies through mutation and subsequent immune selection is called "antigenic drift." Antigenic drift is thought to be important in human immunodeficiency virus (HIV) infection and the continuing seasonal influenza epidemics because immunity generated against one viral quasispecies member selects for escape variants. Attributed in part to antigenic drift are the moderately high failure rate and the short-lived efficacy of influenza vaccines, the failure of synthetic foot-and-mouth disease virus vaccines, and the inability of recombinant HIV vaccines to provide complete protection against field strains of the virus. The hemagglutinin (HA) envelope surface glycoprotein--the major neutralizing determinant of influenza A--is a classic example of an antigenically drifting protein. Walter Gerhard and colleagues demonstrated that the immune pressure exerted by monoclonal antibodies (Abs) selects for HA escape mutants in model systems. Later, Dimmock and colleagues showed that polyclonal anti-sera also select for escape mutants. Similarly, much of the observed variability of glycoprotein 120 (gp120), the principal surface antigen of HIV, is thought to reflect antigenic drift. The correlation of intra-patient viral diversity with immune response strength has been cited as evidence that the immune response is a selective factor in HIV antigenic drift. Phylogenetic analyses describe divergence within a viral population, and these methods have been used to infer the selective advantages of viral variation. A more direct indication of the selective advantage gained through variation is an observed overabundance of replacement mutations relative to silent mutations in viral proteins. Such analyses of gp120 and its V regions indicate that replacement mutations are generally over-represented in this protein and thus appear to confer selective advantage to HIV-1. In more detailed analyses, several groups tested individual codons for replacement mutations that are, as an aggregate, overabundant. However, none of these methods determine which replacement mutations are actually positively selected. Also, when replacement mutations of varying fitness are lumped together, positively selected mutations may remain undetected among negatively selected mutations. To overcome these limitations, we have developed a "selection mapping" algorithm. The cornerstone of selection mapping is the testing of each observed replacement mutation at each codon to identify those particular replacement mutations that are overabundant relative to silent mutations at that codon. Such replacement mutations are determined to be positively selected. Negatively selected variants are recognized as "noise" and are thereafter ignored. Here, we use the selection mapping method to identify the positively selected variants of influenza A HA (H3 serotype),

HIV-1 reverse transcriptase (RT), and HIV-1 gp120.

CONCLUSION

We have developed an algorithm for using sequence data to map the positively selected mutations of viral quasispecies. We have used this method to map the positively selected variants of influenza A HA, HIV-1 RT, and HIV-1 gp120. Other obvious targets for selection mapping are the hepatitis C and foot-and-mouth disease viruses. We believe that potentially the most illuminating use of selection mapping may be the comparison of viral subpopulations to determine which variants are advantageous under different selective pressures. For example, selection mapping of HIV isolates with different cellular tropisms will allow the determination of mutations that are positively selected depending on the host cell type. Also, we may use selection mapping to analyze HIV breakthrough infections to determine if vaccines prevented the HIV quasispecies from inhabiting normally advantageous regions of the quasispecies sequence space. Finally, we propose that the positively selected viral variants (as opposed to all viral variants) should be included in future, highly multivalent vaccines designed to compensate for B-cell-selected antigenic drift.