BIOINFORMATICS APPLICATIONS NOTE Vol. 16 no. 4 2000
Pages 400–401

Detecting hypermutations in viral sequences with an emphasis on G → A hypermutation

Patrick P. Rose and Bette T. Korber

Theoretical Biology and Biophysics, Los Alamos National Laboratories, MS K710 Los Alamos, NM 87544, USA

Received on October 25, 1999; revised on November 30, 1999; accepted on December 7, 1999

Abstract

Summary: *This program compares sequence sets to a reference sequence, tallies G → A hypermutations, and presents the results in various tables and graphs, which include dinucleotide context, summaries of all observed nucleotide changes, and stop codons introduced by hypermutation.*

Availability: *www.hiv.lanl.gov/HYPERMUT/hypermut.html*

Contact: *pxr@t10.lanl.gov or btk@t10.lanl.gov*

Occasionally among viral sequences an extreme form of mutation is observed, with extensive, monotonous G → A substitutions (Vartanian *et al.*, 1991; Wain-Hobson, 1996); this is defined as hypermutation. The program HYPERMUT identifies hypermutated sequences in a sequence alignment. It was first written in the context of a study of HIV sequence evolution during anti-viral therapy (Zhang *et al.*, 1999). The current version provides detailed analysis, graphical output and a web-based interface. www.hiv.lanl.gov. Hypermutation is not limited to HIV, but has been observed in many other viruses, including CAEV (Wain-Hobson *et al.*, 1995); HBV (Gunther *et al.*, 1997; Ngui *et al.*, 1999); and even TTV, a DNA virus (Ball *et al.*, 1999).

Hypermutation can result in premature stops (Cheynier *et al.*, 1997; Wain-Hobson, 1996), when Tryptophan (UGG) changes to a stop codon (UAG, UAA or UGA) and are often riddled with stop codons and not viable. There is ample opportunity *in vivo* for recombination, however, which would allow short stretches of hypermutated sequences to become incorporated into a viable background, and provide an additional mechanism for viral diversification (Fitzgibbon *et al.*, 1993). Two mechanisms have been proposed to explain hypermutation. The first is a faulty RT which misincorporates As for Gs (Fitzgibbon *et al.*, 1993). There is a naturally high A content (35% A) in HIV which may be the result of a tendency for HIV RT to make such substitutions even under normal circumstances. The second is an imbalance in nucleotide pools, in particular a suboptimal dCTP concentration (Vartanian *et al.*, 1991). Hypermutation can be forced *in vitro* and *in vivo* by deliberate introduction of asymmetric dNTP concentrations (Julias and Pathak, 1998).

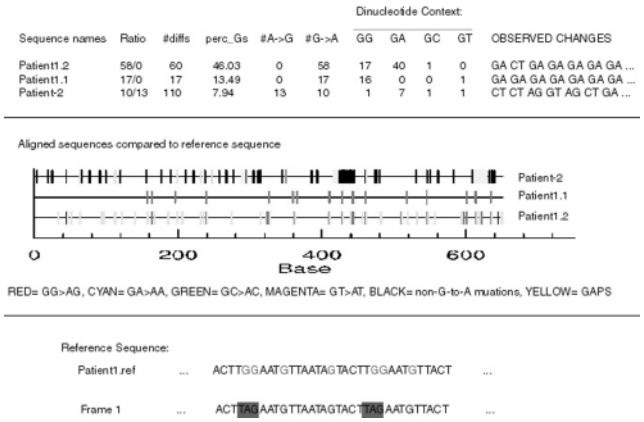

Fig. 1. The output of HYPERMUT is shown for two hypermutated sequences compared to a reference sequence from the same patient (patient1.ref), and to an additional non-hypermutated sequence from another patient. (The example sequences were from the study Zhang *et al.*, 1999). Top: summary of the substitutions between the reference sequence and the other sequences. The 'observed changes' list the substitutions between the sequence and the reference sequence, and would help a user pick out a hypermutation pattern other than G → A. Center: the hash marks indicate the physical location of each mutation in the sequence, colored hashes are G → A mutations, with dinucleotide context represented by four distinct colors; black hashes are non-G-to-A mutations; and yellow hashes mark indels. The comparison of patient 1 with patient 2 shows non-hypermutated sequence comparison, dominated by black hashes. Comparison with sequence 1.1 shows a hypermutated sequence where 13.5% of the Gs in the reference sequence were mutated, and the substitutions were almost exclusively in the dinucleotide context GpG. Sequence 1.2 shows a hypermutated sequence where the dinucleotide context is dominated by the dinucleotide context GpA. Bottom: illustrations from a fragment of the output where hypermutation results in stop codons. The top sequence is the reference sequence, with all Gs highlighted. The red boxes in the lower sequence indicate stop codons that were introduced by hypermutation in sequence 1.2.

The program is designed to highlight and quantify

G → A hypermutations among the background of mutations that regularly occur. The overall analysis compares every sequence in an alignment to the first sequence. Several tables and graphs are produced, which focus on particular issues of interest. Dinucleotide context is an important feature of G → A hypermutation, with GpA and GpG being most susceptible (Wain-Hobson, 1996). HYPERMUT provides tallies of the four types of dinucleotide context of G → A mutations and a detailed graph depicting the types of mutations with their relative physical location in the sequence (Figure 1). To detect hypermutation against a background of normal mutations, the G → A/A → G ratio and a summary of all mutations is provided (Figure 1). As stop codons are frequently introduced through hypermutation, the program tallies all stop codons in a specified reading frame and highlights their position in the sequence relative to G → A substitutions. A user can restrict the analysis to a region of interest within the alignment as hypermutation may be limited to a region within the sequence.

Sequences are treated as character arrays where each nucleotide is separately compared with the same location in the reference sequence. The program identifies the changes, compares them with their neighboring locations for dinucleotide and codon context, lists them in a table and passes the coordinates on to XYPlot for the graphics. The user is provided an interactive setting with options and can submit aligned sequences in any common format. Graphs or tables resulting from the analysis can be downloaded in multiple image formats. XYPlot coordinates of the graphs can be obtained to integrate into other graphic-design software.

Acknowledgements

This work was supported by an interagency agreement between the DOE and NIH for the HIV sequence database, 3-Y01-AI-70001-13. We are grateful to Jean Carr, Francine McCutchan, and Gerald Learn for helpful discussion, Satish Pillai and Brian Gaschen for help with the web design, and Linqi Zhang for providing sequences.

References

Ball,J.K., Curran,R., Berridge,S., Grabowska,A.M., Jameson,C.L., Thompson,B.J., Irving,W.L. and Sharp,B.M. (1999) TT virus sequence heterogeneity in vivo: evidence for co-infection with multiple genetic types. *J. Gen. Virol.*, **80**, 1759–1768.

Cheynier,R., Gratton,S., Vartanian,J.P., Meyerhans,A. and Wain-Hobson,S. (1997) G → A hypermutation does not result from polymerase chain reaction. *AIDS Res. Hum. Retroviruses*, **13**, 985–986.

Fitzgibbon,J.E., Mazar,S. and Dubin,D.T. (1993) A new type of G → A hypermutation affecting human immunodeficiency virus. *Aids Res. Hum. Retroviruses*, **9**, 833–838.

Gunther,S., Sommer,G., Plikat,U., Iwanska,A., Wain-Hobson,S., Will,H. and Meyerhans,A. (1997) Naturally occurring hepatitis B virus genomes bearing the hallmarks of retroviral G → A hypermutation. *Virology*, **235**, 104–108.

Julias,J.G. and Pathak,V.K. (1998) Deoxyribonucleoside triphosphate pool imbalances *in vivo* are associated with an increased retroviral mutation rate. *J. Virol.*, **72**, 7941–7949.

Ngui,S.L., Hallet,R. and Teo,C.G. (1999) Natural and iatrogenic variation in hepatitis B virus. *Rev. Med. Virol.*, **9**, 183–209.

Vartanian,J.P., Meyerhans,A., Asjo,B. and Wain-Hobson,S. (1991) Selection, recombination, and G → A hypermutation of human immunodeficiency virus type 1 genomes. *J. Virol.*, **65**, 1779–1788.

Wain-Hobson,S. Retroviral G → A Hypermutation. In Myers,G., Korber,B.T., Foley,B.T., Jeang,K.-T., Mellors,J.W. and Wain-Hobson,S. (eds), *Human Retroviruses and AIDS 1996*. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.

Wain-Hobson,S., Sonigo,P., Quyader,M., Gazit,A. and Henry,M. (1995) Erratic → hypermutation within a complete caprine arthritis-encephalitis virus (CAEV) provirus. *Virology*, **209**, 297–303.

Zhang,L., Ramratnam,B., Tenner-Racz,K., He,Y., Vesanen,M., Lewin,S., Talal,A., Racz,P., Perelson,A.S., Korber,B.T., Markowitz,M. and Ho,D.D. (1999) Quantifying residual HIV-1 replication in patients receiving combination antiret. *N. Engl. J. Med.*, **340**, 1605–1613.