HIV-1 *pol* mutation frequency by subtype and treatment experience: extension of the HIVseq program to seven non-B subtypes

Soo-Yon Rhee[a], Rami Kantor[b], David A. Katzenstein[a], Ricardo Camacho[c], Lynn Morris[d], Sunee Sirivichayakul[e], Louise Jorgensen[f], Luis F. Brigido[g], Jonathan M. Schapiro[a] and Robert W. Shafer[a] for the International Non Subtype B HIV-1 Working Group*

Objective: HIVseq was developed in 2000 to make published data on the frequency of HIV-1 group M protease and reverse transcriptase (RT) mutations available in real time to laboratories and researchers sequencing these genes. Because most published protease and RT sequences belonged to subtype B, the initial version of HIVseq was based on this subtype. As additional non-B sequences from persons with well-characterized antiretroviral treatment histories have become available, the program has been extended to subtypes A, C, D, F, G, CRF01, and CRF02.

Methods: The latest frequency of each protease and RT mutation according to subtype and drug-class exposure was calculated using published sequences in the Stanford HIV RT and Protease Sequence Database. Each mutation was hyperlinked to published reports of viruses containing the mutation.

Results: As of September 2005, the mean number of protease sequences per non-B subtype was 534 from protease inhibitor-naive persons and 133 from protease inhibitor-treated persons, representing 13.2% and 2.3%, respectively, of the data available for subtype B. The mean number of RT sequences per non-B subtype was 373 from RT inhibitor-naive persons and 288 from RT inhibitor-treated persons, representing 17.9% and 3.8%, respectively, of the data available for subtype B.

Conclusions: HIVseq allows users to examine protease and RT mutations within the context of previously published sequences of these genes. The publication of additional non-B protease and RT sequences from persons with well-characterized treatment histories, however, will be required to perform the same types of analysis possible with the much larger number of subtype B sequences.

AIDS 2006, **20**:643–651

From the [a]Division of Infectious Disease, Stanford University, Stanford, California, the [b]Division of Infectious Diseases, Brown University, Providence, Rhode Island, USA, the [c]Hospital Egas Moniz, Lisbon, Portugal, the [d]National Institute of Communicable Diseases, Johannesburg, South Africa, the [e]Chulalongkorn University, Bangkok, Thailand, the [f]The Danish HIV Cohort Study, Copenhagen, Denmark, and the [g]Instituto Adolfo Lutz, Sao Paulo, Brazil.

Correspondence to Ms S.-Y. Rhee, Division of Infectious Diseases, Room S-169, Stanford University, Stanford, CA 94305.

E-mail: syrhee@stanford.edu

* See the Appendix for other members of the International Non Subtype B HIV-1Working Group.

Received: 19 September 2005; revised: 3 November 2005; accepted: 17 November 2005.

ISSN 0269-9370 © 2006 Lippincott Williams & Wilkins

Keywords: HIV-1, antiretroviral therapy, drug resistance, HIV-1 subtype, reverse transcriptase, protease

Introduction

HIV-1 protease and reverse transcriptase (RT) sequencing is frequently performed to identify viruses likely to have decreased drug susceptibility as a result of either primary or acquired drug-resistance mutations. Sequencing results, however, are often difficult to interpret because of the high mutation rate and genetic variability of HIV-1. Although many HIV-1 protease and RT drug-resistance mutations have been well characterized, the significance of many other mutations in these genes is uncertain, particularly in non-subtype B HIV-1 isolates [1].

HIVseq (http://hivdb6.stanford.edu/asi/deployed/HIVseq.html) was designed in 2000 to accept user-submitted protease and RT sequences, compare them with a consensus reference sequence, and to use the differences (mutations) as query parameters for interrogating the HIV RT and Protease Sequence Database [2,3]. The program output listed the frequency of each observed mutation in treated and untreated persons and provided links to previous reports of these mutations. By allowing users to examine new sequences in the context of those previously published, the program facilitates the identification of sequences with unusual mutations, which can be immediately rechecked for sequence quality (e.g., electrophoretic intensity of observed peaks and background noise in the questionable region of the sequence). In addition, unexpected associations between specific mutations and previous drug treatment could be identified. For each gene, program output included one table for subtype B viruses and another table containing pooled data for all non-subtype B viruses. Although each of the subtypes differs from other subtypes genetically by as much as they differ from subtype B, insufficient sequence data have been available to determine mutation frequencies for each subtype.

The authors have recently collaborated to compile a large number of non-B subtype sequences from persons with well-characterized antiretroviral treatment histories [4]. To assist laboratories performing HIV-1 sequencing, data from this collaboration and from other published studies have been used to extend HIVseq to include non-subtype B viruses. Here, the process used to create subtype-specific mutation frequency datasets and the usefulness of these data to laboratories submitting individual sequences or researchers examining the complete dataset are described.

Methods and results

Mutation frequency data

HIV-1 subtype

Similarity plotting and bootscanning [5] using a window size of 400 and a step size of 40 nucleotides were performed using the Los Alamos HIV Sequence Database reference sequences for each of the nine pure subtypes and CRF01_AE and CRF02_AG [6]. Isolates that contained a combination of more than one subtype were excluded from analysis, unless subtypes A and G were detected in a pattern consistent with CRF02_AG. Because CRF01_AE *pol* sequences do not contain recombinant breakpoints, subtype assignment was based on the fact that *pol* CRF01_AE and pure A sequences are divergent. This approach had an accuracy of 96% when applied to the genes for protease and RT of 137 well-characterized subtype A, CRF01_AE, and CRF02_AG isolates with known subtypes based on *pol* and *gag* and/or *env*, with most errors resulting from the misclassification of subtype A protease sequences as CRF01_AE (data not shown).

Antiretroviral treatment history

Sequences were classified according to the treatment history of the persons from whom the sequenced isolates were obtained. Protease sequences were classified as protease inhibitor (PI) naive or treated based upon whether or not the person from whom the sequence was obtained had ever received a PI [regardless of history of use of RT inhibitors (RTI)]. RT sequences were classified as untreated, treated with nucleoside RTI (NRTI) or treated with non-nucleoside RTI (NNRTI). Sequences were classified as NRTI treated only if they were obtained from persons who had received one or more NRTI but no NNRTI. Persons having more than one isolate with the same mutation were counted only once.

To reduce the influence of possible transmitted resistance and omissions in treatment history, sequences were excluded from the drug class-naive category if they came from individuals with primary HIV-1 infection in regions of the world where antiretroviral therapy was in widespread use or if the sequences came from untreated persons containing two or more established non-polymorphic drug-resistance mutations.

Sequence quality control

Only sequences determined by dideoxyterminator sequencing were included. The consensus amino acid sequence was used to represent isolates when multiple clones were sequenced. To reduce the impact of sequencing errors, a sequence quality score was assigned to all sequences in the database. This score equaled the total number of stop codons, highly ambiguous nucleotides (B, D, H, V, N), and highly unusual mutations (defined as mutations occurring at a frequency of less than 0.05% in pooled treated and untreated group M sequences). Positions containing mixtures in which one of the components of the mixture was a stop codon or a highly unusual residue (e.g. a mutation at a highly conserved site such as the active site of an enzyme)

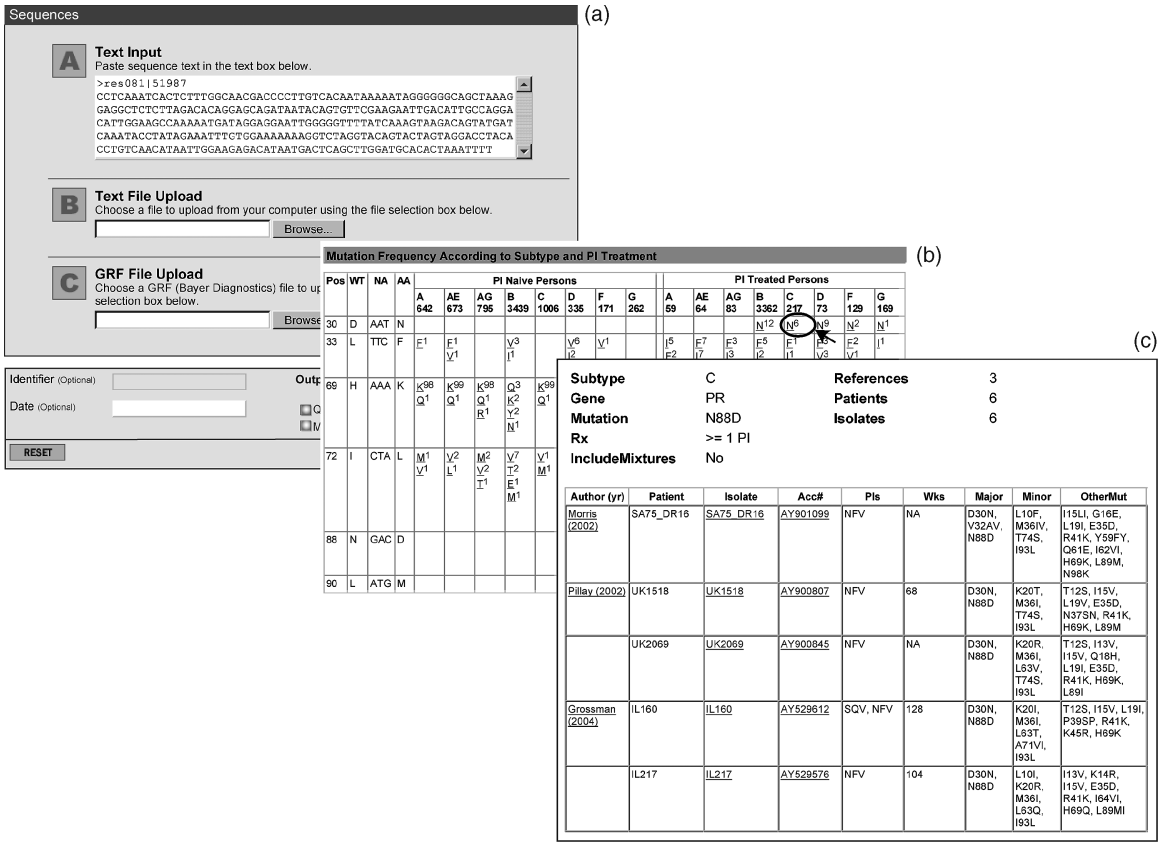

Fig. 1. HIVseq graphical user interface including the sequence upload form (a), program output showing the frequency of each mutation by subtype and treatment for those mutations in the submitted sequence (b), and the detailed isolate information page accessible through each mutation's hyperlink (c). The example shown here is a subtype C protease sequence. The circled text in (b) shows the frequency of N88D in protease inhibitor-treated individuals with subtype C viruses. (c) The top of a list of all published reports of N88D in protease inhibitor-treated persons with subtype C viruses.

contributed to the sequence quality score. Protease sequences with a sequence quality score of four or higher and RT sequences (positions 1–240) with a sequence quality score of six or higher were excluded from the dataset.

Based on the sequence quality control criteria, 47 subtype B (28 protease, 19 RT) and 79 non-B sequences (19 protease, 60 RT) were excluded. More than half of the excluded sequences had one or more stop codons,

Table 1. HIV-1 protease sequences according to subtype and treatment exposure: datasets used to provide mutation frequency data for the HIVseq program (September 2005).

	PI naive[a]			PI experienced		
Subtype	Viruses[b]	Persons	Citations[c]	Viruses[b]	Persons	Citations[c]
Subtype B	4037	3369	124	5720	3354	104
Non-B subtypes						
A	637	621	52	71	59	10
AE	532	524	53	71	64	12
AG	794	791	55	101	83	16
C	993	973	83	276	216	20
D	343	311	55	89	73	23
F	178	173	41	139	129	18
G	261	259	31	185	169	10
Non-B total	3738	3652		932	793	

PI, protease inhibitor.
[a]Never received a PI, regardless of their reverse transcriptase inhibitor treatment history.
[b]Total number of virus isolates for which sequences exist in the Stanford HIV RT and Protease Sequence Database (according to subtype and treatment). The number of viruses is higher than the number of patients because some patients have had viruses assessed a multiple time points.
[c]Number of different published references (or unpublished studies associated with GenBank submissions) from which the sequence data were obtained.

Table 2. HIV-1 reverse transcriptase sequences according to subtype and treatment exposure: datasets used to provide mutation frequency data for the HIVseq Program (September 2005).

	RTI naive[a]			NRTI experienced but NNRTI naive[b]			NNRTI experienced[b]		
Subtype	Viruses[c]	Persons	Citations[d]	Viruses[c]	Persons	Citations[d]	Viruses[c]	Persons	Citations[d]
Subtype B	2080	2077	101	4245	3205	106	3334	1772	76
Non-B Subtypes									
A	454	453	50	60	54	16	251	222	11
AE	444	443	47	196	191	15	116	111	12
AG	441	441	34	63	56	15	76	66	8
C	888	863	67	249	198	23	246	222	14
D	135	135	30	75	69	15	206	160	13
F	123	123	32	76	69	19	76	67	12
G	127	127	26	122	118	4	204	194	7
Non-B total	2612	2585		841	755		1175	1042	

[a]RTI-naive persons are those who never received a nucleoside (NRTI) or non-nucleoside (NNRTI) reverse transcriptase inhibitor regardless of their protease inhibitor treatment history.
[b]Two separate categories for NRTI-experienced persons were created to enable the identification of mutations associated with NRTI therapy (by comparing viruses from RTI-naive persons with those from NRTI-experienced but NNRTI-naive persons) and those associated with NNRTI therapy (by comparing viruses from NNRTI-experienced persons with those from both categories of NNRTI-naive persons.).
[c]Total number of virus isolates for which sequences exist in the Stanford HIV RT and Protease Sequence Database (according to subtype and treatment). The number of viruses is higher than the number of patients because some patients have had viruses assessed from multiple time points.
[d]Number of different published references (or unpublished studies associated with GenBank submissions) from which the sequence data were obtained.

catalytic residue mutations (D25, T26, and G27 in protease and D110, D185, D186 in RT), or highly ambiguous nucleotides. In future versions of the program, users will be provided with the option of adjusting this sequence quality score cut-off.

Although most mutations occurring as part of a mixture usually reflect true mixed virus populations [7], these mutations are also statistically more likely to result from technical artifact or overinterpretation of sequence chromatograms than to be mutations that are present as the dominant variant in a sample [8]. Excluding such mutations is useful for some purposes – for example to facilitate the identification of non-polymorphic mutations (i.e., mutations observed only as part of mixtures are more likely to be artifactual). However, because most mutations occurring as part of a mixture are clinically and epidemiologically significant, these mutations should be counted for other purposes. Therefore, users are provided with the option to obtain results that include or do not include mutations present as part of an electrophoretic mixture.

HIVseq program

Users submit single sequences or sets of multiple sequences encompassing HIV-1 protease and/or RT by uploading a file containing nucleotide sequences in fasta

(a)

Pos	WT	NA	AA	PI Native persons								PI Treated persons							
				A 642	AE 673	AG 795	B 3439	C 1006	D 335	F 171	G 262	A 59	AE 64	AG 83	B 3362	C 217	D 73	F 129	G 169
82	V	---	A		I[3]	I[3]	I[1]	I[6]	I[1]	I[2]	I[96]	A[3] F[2] I[2]	A[7] I[5] F[2]	A[6] I[4] S[3] F[3] T[1]	A[23] T[3] F[2] I[2] S[1]	A[10] I[6] T[2] S[1] C[1]	A[13] I[6] L[3] E[1] T[1]	A[30] I[3] T[2] F[1] L[1]	I[72] T[13] S[4] A[2] M[2] F[1]

(b)

Pos	WT	NA	AA	NRTI (but no NNRTI) Treated persons								NNRTI Treated persons							
				A 50	AE 189	AG 54	B 3196	C 191	D 69	F 65	G 116	A 222	AE 111	AG 67	B 1776	C 221	D 160	F 67	G 194
106	V	---	M		I[3] L[1]	I[4]	I[1]		I[3]	I[2]		I[1]	I[4] L[3] M[2]	M[2]	A[2] I[2] M[1]	M[13]	M[1] I[1]	M[3] A[2]	I[3] A[1] M[1]

Fig. 2. Two examples of published inter-subtype differences in the genetic mechanisms of HIV-1 drug resistance as demonstrated by HIVseq output. (a) Association of the protease mutations V82T and V82M with protease inhibitor (PI) therapy in subtype G viruses [12]. V82I is the consensus amino acid sequence in subtype G. As a result, the mutations 82T and 82M require a change in only one nucleotide, in contrast to 82A (the most common PI-associated mutation in other subtypes), which requires changes in two nucleotides in subtype G. (b) Association of the reverse transcriptase mutation V106M in subtype C viruses with therapy with non-nucleoside reverse transcriptase inhibitors [10,11].

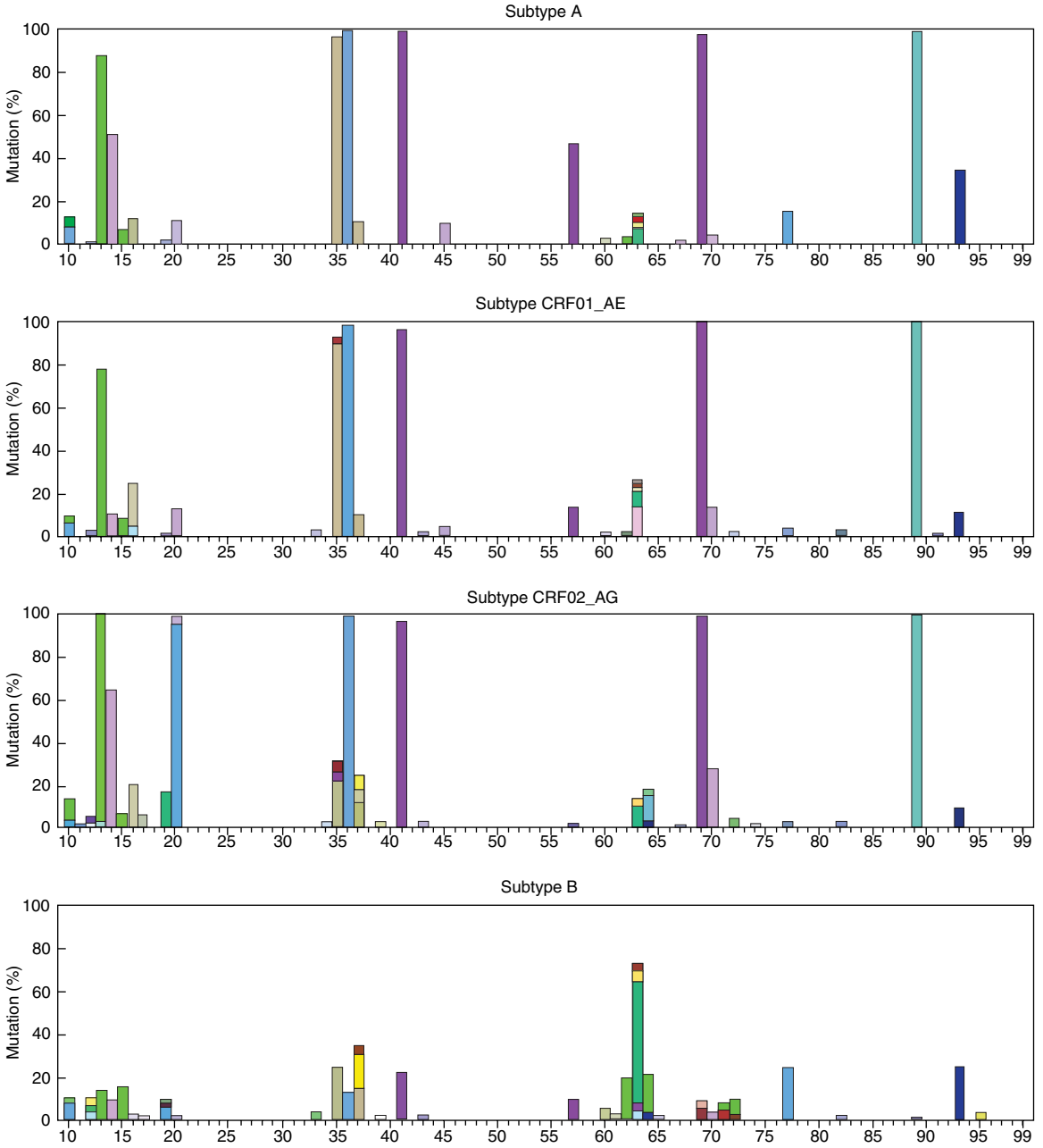

Fig. 3(a). Frequency of HIV-1 protease (positions 10–99) (a) and reverse transcriptase (positions 40–240) (b) according to subtype in previously untreated person. Alanine, cysteine, aspartate, glutamate, phenylalanine, glycine, histidine, isoleucine, lysine, leucine, methionine, asparganine, proline, glutamine, arginine, serine, threonine, valine, tryptophan, tyrosine, insertion, deletion.

format, pasting fasta nucleotide sequences into a text box, typing a list of mutations into a text box, or selecting mutations from a dropdown menu. Submitted sequences are aligned, translated, and compared with the subtype B consensus reference sequence to derive a list of mutations that are used to interrogate the HIV RT and Protease Sequence Database. Submitted mutations are used directly to interrogate the HIV RT and Protease Sequence Database (Fig. 1).

Program output consists of three tables: a protease table providing protease mutation frequency for HIV-1 isolates from PI-naive and PI-experienced persons, an NRTI table providing RT mutation frequency for isolates from RTI-naive and NRTI-treated (but NNRTI-naive) persons, and an NNRTI table providing RT mutation frequency for isolates from NNRTI-naive and NNRTI-treated persons. Each table contains one row for each mutation and 20 columns (Fig. 2). Columns 1 to 4 list the

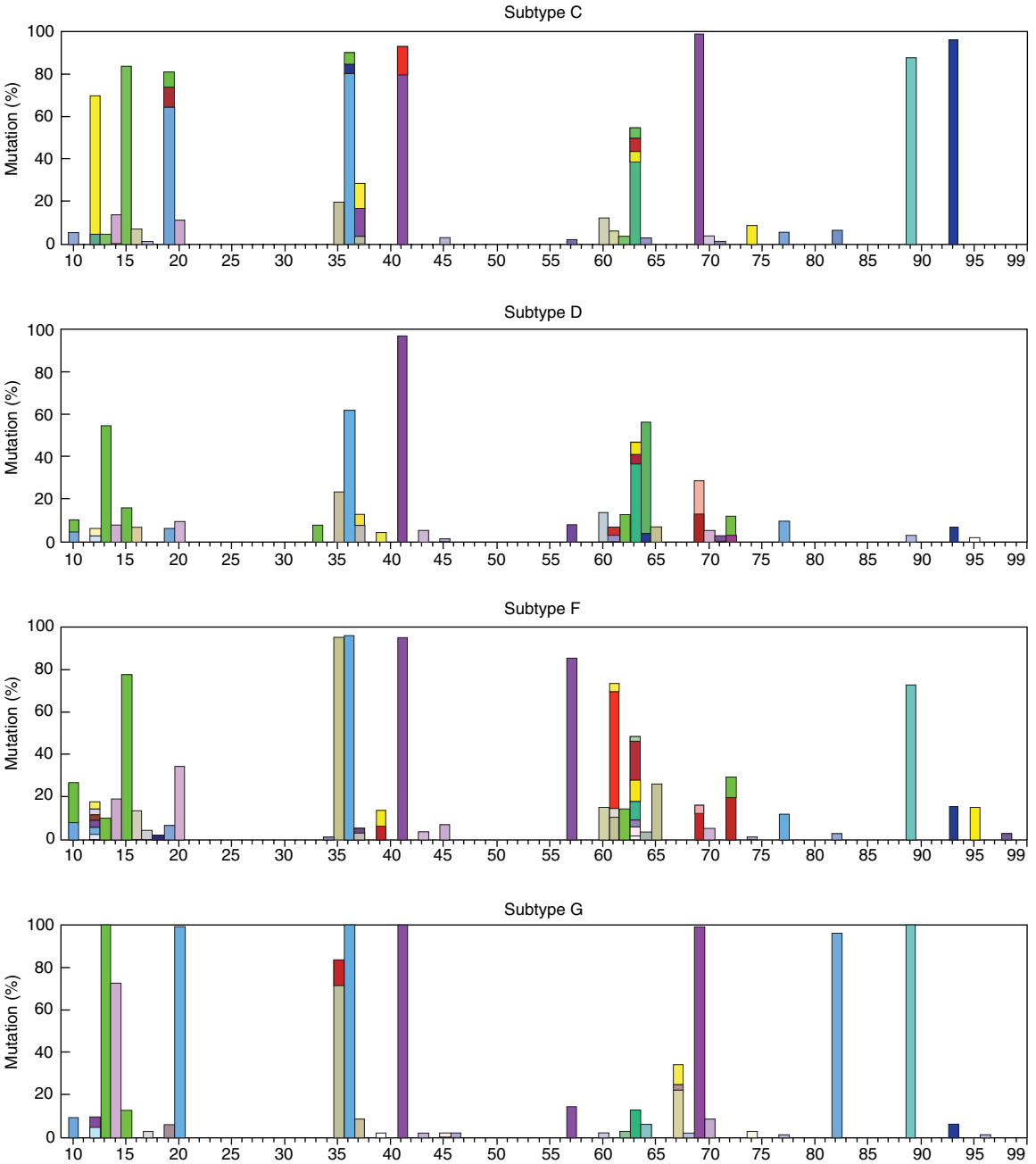

Fig. 3(a). (*Continued*).

position, the position's consensus amino acid, the submitted nucleotide triplet, and the submitted amino acid. Columns 5 to 12 list the frequency of each mutation in subtypes A, AE, AG, B, C, D, F, and G in drug-class-naive persons. Columns 13 to 20 list the frequency of each mutation in subtypes A, AE, AG, B, C, D, F, and G in drug-class-experienced persons.

Mutations that occur in at least two persons and have a frequency of at least 0.5% within a subtype are used to populate the appropriate columns of the table. A superscripted number following the mutation indicates the proportion of isolates belonging to the subtype and treatment category (drug-class naive or experienced) containing the mutation. Each mutation is also a hyperlink to a separate web page with detailed information on each isolate, including literature references with Medline abstracts, the GenBank accession number, and complete sequence and treatment records. Figure 1 summarizes the flow of information to the user showing the sequence upload form (Fig. 1a), the mutation frequency output (Fig. 1b), and the detailed isolate information

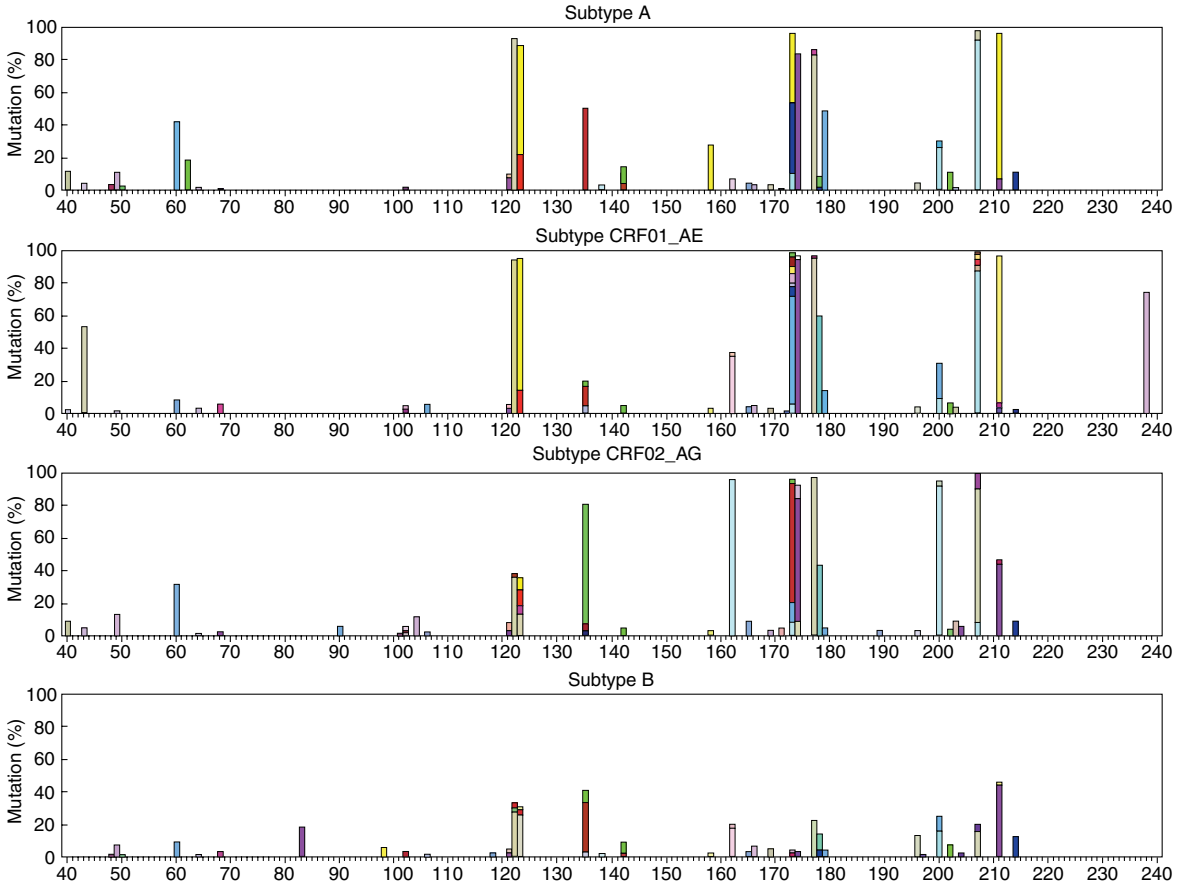

Fig. 3(b). (*Continued*).

page accessible through each mutation's hyperlink (Fig. 1c).

Complete dataset view

In contrast to the HIVseq program, which provides mutation frequency data only for those mutations present in a user-submitted sequence, the complete dataset view (http://hivdb.stanford.edu/cgi-bin/MutPrevBySubtypeRx.cgi) allows users to examine the frequency of all mutations according to subtype and treatment. Those viewing the complete dataset can adjust three parameters: (i) whether the mutation count is restricted to a single isolate per individual (i.e., the most recently obtained isolate while on antiretroviral therapy) or whether it may include more than one isolate from an individual if and only if different isolates from a person have different mutations at the same position; (ii) whether mutations present as part of electrophoretic mixtures contribute to mutation frequency; and (iii) the minimum frequency required for a displayed mutation. Three mutation frequency thresholds are available: all mutations, mutations with a reported frequency of at least 0.5%, and mutations with a reported frequency of at least 1% and occurring in at least two persons. The last option is necessary because fewer than 100 isolates from treated persons are currently available for some subtypes.

Tables 1 and 2 summarize the number of isolates in the dataset according to subtype and treatment classification. There are a mean number of 534 protease sequences per non-B subtype from PI-naive persons and 133 from PI-treated persons, representing 13.2% and 2.3%, respectively, of the data available for subtype B. There are a mean of 373 RT sequences per non-B subtype from RTI-naive persons and 288 from RTI-treated persons, representing 17.9% and 3.8%, respectively, of the data available for subtype B.

Figure 3(a) and 3(b) provides graphical summaries of the reported frequency of each mutation by subtype at protease positions 10–99 and RT positions 40–240. The complete dataset view on the web also provides the option of viewing the Shannon entropy [9] at each position for each subtype: a measure of variability at each position that is independent of the fact that subtype B consensus sequence was used as the reference sequence for all subtypes.

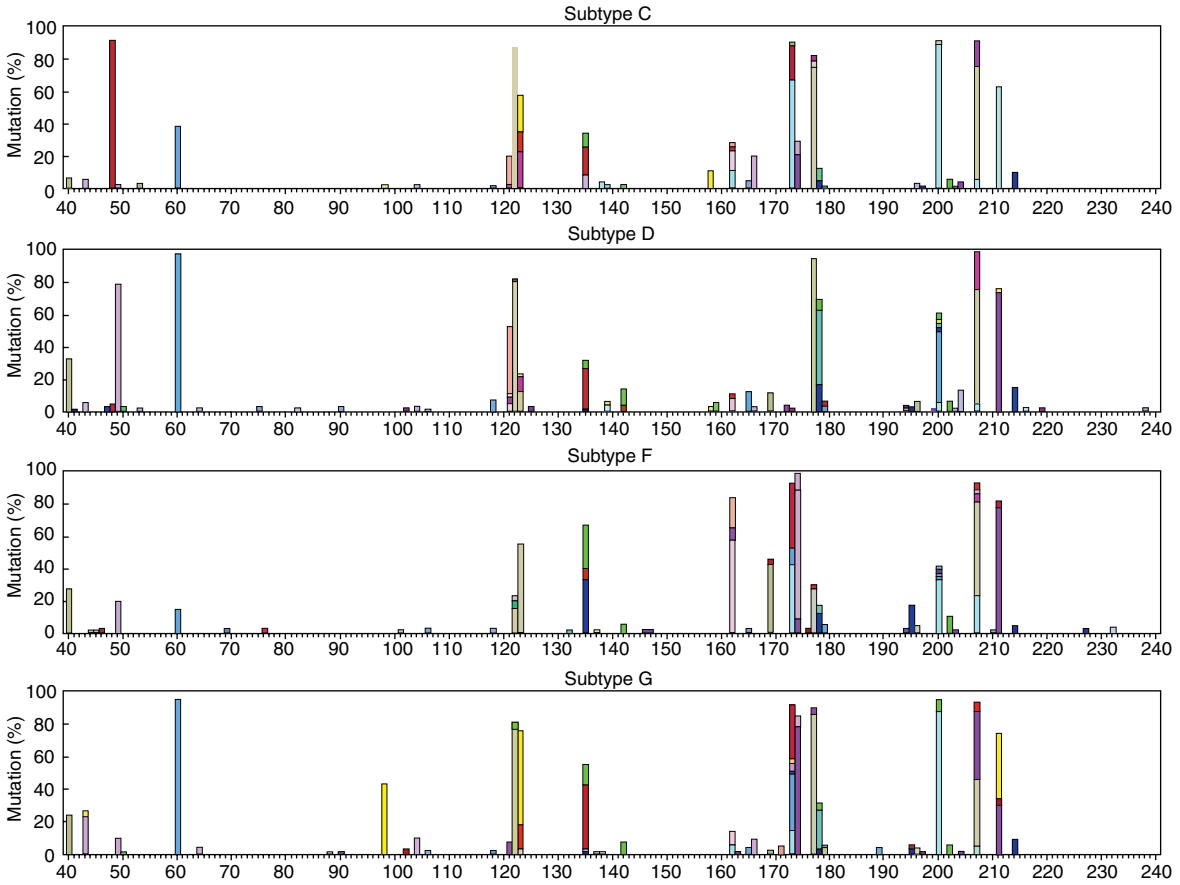

Fig. 3(b). *(Continued)*.

Discussion

The HIVseq program demonstrates that the body of published sequence data on a gene can be made available in real time to laboratories and researchers sequencing new isolates of that gene. Because most published HIV-1 drug-resistance data were originally based on subtype B sequences, the initial version of HIVseq was designed primarily for this subtype. However, subtype B represents only about 10% of the global HIV-1 epidemic. Although more non-B protease and RT genes are being sequenced as access to antiretroviral treatment increases worldwide [1], the mean number of sequences per subtype from treated subjects is only approximately 2–4% of the number of sequences from treated subjects with subtype B.

We have recently collaborated to assess the impact of HIV-1 subtype and antiretroviral treatment on the distribution of mutations in protease and RT [4]. The analysis showed that mutations at each of 55 established subtype B drug-resistance positions occurred in one or more non-B subtypes, and that mutation at 80% of these positions were significantly associated with treatment in non-B isolates. However, to increase the statistical power of the analysis, no distinction was made between different substitutions at the same position; all differences from consensus B were considered mutations. Additional data and further analyses, therefore, are required to identify differences in the spectrum of mutations at individual positions. Indeed, two such differences have been reported and Fig. 2 shows HIVseq output examples for these [10–12].

There is a potential for biased results whenever data are derived from retrospective studies rather than from prospective epidemiological studies designed to obtain representative samples from the population in question (HIV-1-infected persons worldwide). However, because epidemiologically sound global drug-resistance data do not yet exist, the data we are describing take on added importance. Nonetheless, this study and the HIVseq program provide frequency data on mutations from published studies rather than unbiased estimates of the prevalence of these mutations.

A sustainable mechanism to expand the data for HIVseq is required. Although it is customary for authors to submit nucleotide sequences to GenBank at the time of their description in a published manuscript, there is no mechanism compelling authors to submit important correlated data such as date of virus isolation, geographic origin, and, in particular, treatment history. In addition,

obtaining accurate treatment histories is becoming progressively more difficult as the number of antiretroviral drugs and the duration of therapy with these drugs increases. Mechanisms for making certain of the accuracy of a person's treatment history are required to optimize the reliability of published data and the usefulness of applications, such as HIVseq, that are based on such data.

Additional considerations are necessary before the data provided with HIVseq can be used to compare the reported frequency of mutations between different subtypes. In creating the mutation datasets, no effort was made to distinguish mutations developing in multiple individuals from those that developed in a smaller number of founder viruses. For example, if a large number of closely related viruses from one geographic region are published, the mutations present in these viruses would be overrepresented in the dataset with respect to their actual prevalence in all viruses of the same subtype. Although we have developed an approach for dealing with this problem of population stratification [4], it is not yet implemented through a strictly computational approach. However, in future versions of the program, viewers of the complete dataset will be provided with the option of adjusting output for population stratification.

Sponsorship: SYR and RWS were supported in part by a grant from the National Institutes of Allergy and Infectious Diseases (AI46148–01); DK and RK received support from the Doris Duke Charitable Foundation.

References

1. Parkin NT, Schapiro JM. **Antiretroviral drug resistance in non-subtype B HIV-1, HIV-2 and SIV.** *Antivir Ther* 2004; **9**:3–12.
2. Shafer RW, Jung DR, Betts BJ. **Human immunodeficiency virus type 1 reverse transcriptase and protease mutation search engine for queries.** *Nat Med* 2000; **6**:1290–1292.
3. Rhee SY, Gonzales MJ, Kantor R, Betts BJ, Ravela J, Shafer RW. **Human immunodeficiency virus reverse transcriptase and protease sequence database.** *Nucl Acids Res* 2003; **31**:298–303.
4. Kantor R, Katzenstein DA, Efron B, Carvalho P, Wynhoven B, Cane P, *et al*. **Impact of HIV-1 subtype and antiretroviral therapy on protease and reverse transcriptase genotype: results of a global collaboration.** *PLoS Med* 2005; **2**:e112.
5. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, *et al*. **Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination.** *J Virol* 1999; **73**:152–160.
6. Leitner T, Foley B, Hahn B, Marx P, McCutchan F, Mellors J, *et al*. *HIV Sequence Compendium, 2003*. Los Alamos, NM: Los Alamos National Laboratory; 2003.
7. Shafer RW, Hertogs K, Zolopa AR, Warford A, Bloor S, Betts BJ, *et al*. **High degree of interlaboratory reproducibility of human immunodeficiency virus type 1 protease and reverse transcriptase sequencing of plasma samples from heavily treated patients.** *J Clin Microbiol* 2001; **39**:1522–1529.
8. Rhee SY, Fessel WJ, Zolopa AR, Hurley L, Liu T, Taylor J, *et al*. **HIV-1 protease and reverse-transcriptase mutations: correlations with antiretroviral therapy in subtype B isolates and implications for drug-resistance surveillance.** *J Infect Dis* 2005; **192**:456–465.
9. Durbin R, Eddy S, Krogh A, Mitchison G. *Biological Sequence Analysis*. Cambridge, UK: Cambridge University Press; 1998.
10. Brenner B, Turner D, Oliveira M, Moisi D, Detorio M, Carobene M, *et al*. **A V106M mutation in HIV-1 clade C viruses exposed to efavirenz confers cross-resistance to non-nucleoside reverse transcriptase inhibitors.** *AIDS* 2003; **17**:F1–F5.
11. Grossman Z, Istomin V, Averbuch D, Lorber M, Risenberg K, Levi I, *et al*. **Genetic variation at NNRTI resistance-associated positions in patients infected with HIV-1 subtype C.** *AIDS* 2004; **18**:909–915.
12. Camacho R, Godinho A, Gomes P, Abecasis A, Vandamme AM, Palma C, *et al*. **Different substitutions under drug pressure at protease codon 82 in HIV-1 subtype B compared to subtype B infected individuals including a novel I82M resistance mutations.** *Antivir Ther* 2005; **10**:S151.

Appendix

Other members of the International Non Subtype B HIV-1Working Group who contributed to this study included: Ana Patricia Carvalho (Hospital Egas Moniz, Lisbon, Portugal); Patricia Cane (Centre for Infections, Health Protection Agency, UK); Zehava Grossman, Hagit Rudich (Central Virology, PHL, Ministry of Health, Tel-Hashomer, Israel); Praphan Phanuphak (Chulalongkorn University, Bangkok, Thailand); Maria Belen Bouzas, Pedro Cahn (Fundación Huesped, Buenos Aires, Argentina); Africa Holguin, Vincent Soriano (Department of Infectious Diseases, Hospital Carlos III, Madrid, Spain); Rosangela Rodrigues (Instituto Adolfo Lutz, Sao Paulo, Brazil); Joke Snoeck, Annemie Vandamme (Rega Institute for Medical Research, Katholieke Universiteit, Leuven, Belgium); Amilcar Tanuri, Macelo Soares (Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil); John Weber (Wright Fleming Institute, Imperial College, St Mary's Hospital, London, UK); Deenan Pillay (University College of London, London, UK); Wataru Sugiura, Koya Ariyoshi (National Institute of Infectious Diseases, Tokyo, Japan); Terese Katzenstein (Danish HIV Database, Copenhagen, Denmark).