

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
18 February 2010 (18.02.2010)

PCT

(10) International Publication Number
WO 2010/019262 A2

(51) International Patent Classification:
C07K 14/155 (2006.01) *A61K 39/21* (2006.01)
C12N 15/49 (2006.01)

(21) International Application Number:
PCT/US2009/004664

(22) International Filing Date:
14 August 2009 (14.08.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
12/192,015 14 August 2008 (14.08.2008) US

(71) Applicants (for all designated States except US): **LOS ALAMOS NATIONAL SECURITY, LLC** [US/US]; Los Alamos National Laboratory, LC/IP, Los Alamos, NM 87545 (US). **BETH ISRAEL DEACONNESS MEDICAL CENTER** [US/US]; 330 Brookline Avenue, Boston, MA 02215 (US). **DUKE UNIVERSITY** [US/US]; Box 90083, Durham, NC 27708-0083 (US). **THE UNIVERSITY OF ALABAMA AT BIRMINGHAM RESEARCH FOUNDATION** [US/US]; 1530 3rd Avenue South, Birmingham, AL 35294-0107 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **KORBER, Bette, T.** [US/US]; C/o Los Alamos National Security, LLC, Los Alamos National Laboratory, LC/IP, Los Alamos, NM 87545 (US). **FISCHER, William** [US/US]; C/o Los Alamos National Security, LLC, Los Alamos National Laboratory, LC/IP, Los Alamos, NM 87545 (US). **LETVIN, Norman** [US/US]; C/o Beth Israel Deaconness Medical Center, 330 Brookline Avenue, Boston, MA 02215 (US). **LIAO, Hua-xin** [US/US]; C/o Duke University, Box 90083, Durham, NC 27708-0083 (US). **HAYNES, Barton, F.** [US/US]; C/o Duke University,

Box 90083, Durham, NC 27708-0083 (US). **HAHN, Beatrice, H.** [US/US]; C/o The University of Alabama at Birmingham, Research Foundation, AB 1120g, 1530 3rd Avenue South, Birmingham, AL 35294-0111 (US).

(74) Agent: **WILSON, Mary, J.**; Nixon & Vanderhye P.C., 901 North Glebe Road, 11th Floor, Arlington, VA 22203-1808 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

Published:

— without international search report and to be republished upon receipt of that report (Rule 48.2(g))

(54) Title: POLYVALENT VACCINE

(57) Abstract: The present invention relates, in general, to an immunogenic composition (e.g., a vaccine) and, in particular, to a polyvalent immunogenic composition, such as a polyvalent HIV vaccine, and to methods of using same. The invention further relates to methods that use a genetic algorithm to create sets of polyvalent antigens suitable for use, for example, in vaccination strategies.



WO 2010/019262 A2

POLYVALENT VACCINE

This application is a continuation-in-part of U.S. Application No. 11/990,222, filed February 8, 2008, which is the U.S. national phase of International Application No. PCT/US2006/032907, filed August 23, 2006, which designated the U.S. and claims priority from U.S. Provisional Application No. 60/710,154, filed August 23, 2005, and U.S. Provisional Application No. 60/739,413, filed November 25, 2005, the entire contents of which applications are incorporated herein by reference.

TECHNICAL FIELD

The present invention relates, in general, to an immunogenic composition (e.g., a vaccine) and, in particular, to a polyvalent immunogenic composition, such as a polyvalent HIV vaccine, and to methods of using same. The invention further relates to methods that use a genetic algorithm to create sets of polyvalent antigens suitable for use, for example, in vaccination strategies.

BACKGROUND

Designing an effective HIV vaccine is a many-faceted challenge. The vaccine preferably elicits an immune response capable of either preventing infection or, minimally, controlling viral replication if infection occurs, despite the failure of immune responses to natural infection to eliminate the virus (Nabel, Vaccine 20:1945-1947 (2002)) or to protect from superinfection (Altfeld et al, Nature 420:434-439 (2002)). Potent vaccines are needed, with optimized vectors, immunization protocols, and adjuvants (Nabel, Vaccine 20:1945-1947 (2002)), combined with antigens that can stimulate cross-reactive responses against the diverse spectrum of circulating viruses (Gaschen et al, Science 296:2354-2360 (2002), Korber et al, Br. Med. Bull.

58:19-42 (2001)). The problems that influenza vaccinologists have confronted for decades highlight the challenge posed by HIV-1: human influenza strains undergoing antigenic drift diverge from one another by around 1-2% per year, yet vaccine antigens often fail to elicit cross-reactive B-cell responses from one year to the next, requiring that contemporary strains be continuously monitored and vaccines be updated every few years (Korber et al, Br. Med. Bull. 58:19-42 (2001)). In contrast, co-circulating individual HIV-1 strains can differ from one another by 20% or more in relatively conserved proteins, and up to 35% in the Envelope protein (Gaschen et al, Science 296:2354-2360 (2002), Korber et al, Br. Med. Bull. 58:19-42 (2001)).

Different degrees of viral diversity in regional HIV-1 epidemics provide a potentially useful hierarchy for vaccine design strategies. Some geographic regions recapitulate global diversity, with a majority of known HIV-1 subtypes, or clades, co-circulating (*e.g.*, the Democratic Republic of the Congo (Mokili & Korber, J. Neurovirol 11(Suppl. 1):66-75 (2005)); others are dominated by two subtypes and their recombinants (*e.g.*, Uganda (Barugahare et al, J. Virol. 79:4132-4139 (2005)), and others by a single subtype (*e.g.*, South Africa (Williamson et al, AIDS Res. Hum. Retroviruses 19:133-144 (2003)). Even areas with predominantly single-subtype epidemics must address extensive within-clade diversity (Williamson et al, AIDS Res. Hum. Retroviruses 19:133-44 (2003)) but, since international travel can be expected to further blur geographic distinctions, all nations would benefit from a global vaccine.

Presented herein is the design of polyvalent vaccine antigen sets focusing on T lymphocyte responses, optimized for either the common B and C subtypes, or all HIV-1 variants in global circulation [the HIV-1 Main (M) group]. Cytotoxic T-lymphocytes (CTL) directly kill infected, virus-producing host cells, recognizing them via viral protein fragments (epitopes) presented on infected cell surfaces by human leukocyte antigen (HLA) molecules. Helper T-cell responses control varied aspects of the immune response

through the release of cytokines. Both are likely to be crucial for an HIV-1 vaccine: CTL responses have been implicated in slowing disease progression (Oxenius et al, *J. Infect. Dis.* 189:1199-208 (2004)); vaccine-elicited cellular immune responses in nonhuman primates help control pathogenic SIV or SHIV, reducing the likelihood of disease after challenge (Barouch et al, *Science* 290:486-92 (2000)); and experimental depletion of CD8⁺ T-cells results in increased viremia in SIV infected rhesus macaques Schmitz et al, *Science* 283:857-60 (1999)). Furthermore, CTL escape mutations are associated with disease progression (Barouch et al, *J. Virol.* 77:7367-75 (2003)), thus vaccine-stimulated memory responses that block potential escape routes may be valuable.

The highly variable Env protein is the primary target for neutralizing antibodies against HIV; since immune protection will likely require both B-cell and T-cell responses (Moore and Burton, *Nat. Med.* 10:769-71 (2004)), Env vaccine antigens will also need to be optimized separately to elicit antibody responses. T-cell-directed vaccine components, in contrast, can target the more conserved proteins, but even the most conserved HIV-1 proteins are diverse enough that variation is an issue. Artificial central-sequence vaccine approaches (e.g., consensus sequences, in which every amino acid is found in a plurality of sequences, or maximum likelihood reconstructions of ancestral sequences (Gaschen et al, *Science* 296:2354-60 (2002), Gao et al, *J. Virol.* 79:1154-63 (2005), Doria-Rose et al, *J. Virol.* 79:11214-24 (2005), Weaver et al, *J. Virol.*, in press)) are promising; nevertheless, even centralized strains provide limited coverage of HIV-1 variants, and consensus-based reagents fail to detect many autologous T-cell responses (Altfeld et al, *J. Virol.* 77:7330-40 (2003)).

Single amino acid changes can allow an epitope to escape T-cell surveillance; since many T-cell epitopes differ between HIV-1 strains at one or more positions, potential responses to any single vaccine antigen are limited. Whether a particular mutation results in escape depends upon the

specific epitope/T-cell combination, although some changes broadly affect between-subtype cross-reactivity (Norris et al, AIDS Res. Hum. Retroviruses 20:315-25 (2004)). Including multiple variants in a polyvalent vaccine could enable responses to a broader range of circulating variants, and could also
5 prime the immune system against common escape mutants (Jones et al, J. Exp. Med. 200:1243-56 (2004)). Escape from one T-cell receptor may create a variant that is susceptible to another (Allen et al, J. Virol. 79:12952-60 (2005), Feeney et al, J. Immunol. 174:7524-30 (2005)), so stimulating polyclonal responses to epitope variants may be beneficial (Killian et al, Aids 19:887-96
10 (2005)). Escape mutations that inhibit processing (Milicic et al, J. Immunol. 175:4618-26 (2005)) or HLA binding (Ammaranond et al, AIDS Res. Hum. Retroviruses 21:395-7 (2005)) cannot be directly countered by a T-cell with a different specificity, but responses to overlapping epitopes may block even some of these escape routes.

15 The present invention relates to a polyvalent vaccine comprising several "mosaic" proteins (or genes encoding these proteins). The candidate vaccine antigens can be cocktails of k composite proteins (k being the number of sequence variants in the cocktail), optimized to include the maximum number of potential T-cell epitopes in an input set of viral proteins. The
20 mosaics are generated from natural sequences: they resemble natural proteins and include the most common forms of potential epitopes. Since CD8+ epitopes are contiguous and typically nine amino-acids long, sets of mosaics can be scored by "coverage" of nonamers (9-mers) in the natural sequences (fragments of similar lengths are also well represented). 9-Mers not found at
25 least three times can be excluded. This strategy provides the level of diversity coverage achieved by a massively polyvalent multiple-peptide vaccine but with important advantages: it allows vaccine delivery as intact proteins or genes, excludes low-frequency or unnatural epitopes that are not relevant to circulating strains, and its intact protein antigens are more likely to be
30 processed as in a natural infection.

SUMMARY OF THE INVENTION

In general, the present invention relates to an immunogenic composition. More specifically, the invention relates to a polyvalent immunogenic composition (e.g., an HIV vaccine), and to methods of using same. The invention further relates to methods that involve the use of a genetic algorithm to design sets of polyvalent antigens suitable for use as vaccines.

Objects and advantages of the present invention will be clear from the description that follows.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-1F. The upper bound of potential epitope coverage of the HIV-1 M group. The upper bound for population coverage of 9-mers for increasing numbers of variants is shown, for $k = 1-8$ variants. A sliding window of length nine was applied across aligned sequences, moving down by one position. Different colors denote results for different numbers of sequences. At each window, the coverage given by the k most common 9-mers is plotted for Gag (Figs. 1A and 1B), Nef (Figs. 1C and 1D) and Env gp120 (Figs. 1E and 1F). Gaps inserted to maintain the alignment are treated as characters. The diminishing returns of adding more variants are evident, since, as k increases, increasingly rare forms are added. In Figs. 1A, 1C and 1E, the scores for each consecutive 9-mer are plotted in their natural order to show how diversity varies in different protein regions; both p24 in the center of Gag and the central region of Nef are particularly highly conserved. In Figs. 1B, 1D and 1F, the scores for each 9-mer are reordered by coverage (a strategy also used in Fig. 4), to provide a sense of the overall population coverage of a given protein. Coverage of gp120, even with 8 variant 9-mers, is particularly poor (Figs. 1E and 1F).

Figures 2A-2C. Mosaic initialization, scoring, and optimization.

Fig. 2A) A set of k populations is generated by random 2-point recombination of natural sequences (1-6 populations of 50-500 sequences each have been tested). One sequence from each population is chosen (initially at random) for the mosaic cocktail, which is subsequently optimized. The cocktail sequences are scored by computing *coverage* (defined as the mean fraction of natural-sequence 9-mers included in the cocktail, averaged over all natural sequences in the input data set). Any new sequence that covers more epitopes will increase the score of the whole cocktail. Fig. 2B) The fitness score of any individual sequence is the coverage of a cocktail containing that sequence plus the current representatives from other populations. Fig. 2C) Optimization: 1) two “parents” are chosen: the higher-scoring of a randomly chosen pair of recombined sequences, and either (with 50% probability) the higher-scoring sequence of a second random pair, or a randomly chosen natural sequence. 2) Two-point recombination between the two parents is used to generate a “child” sequence. If the child contains unnatural or rare 9-mers, it is immediately rejected, otherwise it is scored (Gaschen et al, Science 296:2354-2360 (2002)). If the score is higher than that of any of four randomly-selected population members, the child is inserted in the population in place of the weakest of the four, thus evolving an improved population; 4) if its score is a new high score, the new child replaces the current cocktail member from its population. Ten cycles of child generation are repeated for each population in turn, and the process iterates until improvement stalls.

Figure 3. Mosaic strain coverage for all HIV proteins. The level of 9-mer coverage achieved by sets of four mosaic proteins for each HIV protein is shown, with mosaics optimized using either the M group or the C subtype. The fraction of C subtype sequence 9-mers covered by mosaics optimized on the C subtype (within-clade optimization) is shown in gray. Coverage of 9-mers found in non-C subtype M-group sequences by subtype-C-optimized

mosaics (between-clade coverage) is shown in white. Coverage of subtype C sequences by M-group optimized mosaics is shown in black. B clade comparisons gave comparable results (data not shown).

5 Figures 4A-4F. Coverage of M group sequences by different vaccine candidates, nine-mer by nine-mer. Each plot presents site-by-site coverage (*i.e.*, for each nine-mer) of an M-group natural-sequence alignment by a single tri-valent vaccine candidate. Bars along the x-axis represent the proportion of sequences matched by the vaccine candidate for a given alignment position:
10 9/9 matches (in red), 8/9 (yellow), 7/9 (blue). Aligned 9-mers are sorted along the x-axis by exact-match coverage value. 656 positions include both the complete Gag and the central region of Nef. For each alignment position, the maximum possible matching value (*i.e.* the proportion of aligned sequences without gaps in that nine-mer) is shown in gray. Fig. 4A) Non-optimal natural
15 sequences selected from among strains being used in vaccine studies (Kong et al, J. Virol. 77:12764-72 (2003)) including an individual clade A, B, and C viral sequences (Gag: GenBank accession numbers AF004885, K03455, and U52953; Nef core: AF069670, K02083, and U52953). Fig. 4B) Optimum set of natural sequences [isolates US2 (subtype B, USA), 70177 (subtype C,
20 India), and 99TH.R2399 (subtype CRF15_01B, Thailand); accession numbers AY173953, AF533131, and AF530576] selected by choosing the single sequence with maximum coverage, followed by the sequence that had the best coverage when combined with the first (*i.e.* the best complement), and so on, selected for M group coverage Fig. 4C) Consensus sequence cocktail (M
25 group, B- and C-subtypes). Fig. 4D) 3 mosaic sequences, Fig. 4E) 4 mosaic sequences, Fig. 4F) 6 mosaic sequences. Figs. 4D-4F were all optimized for M group coverage.

 Figures 5A and 5B. Overall coverage of vaccine candidates: coverage
30 of 9-mers in C clade sequences using different input data sets for mosaic

optimization, allowing different numbers of antigens, and comparing to different candidate vaccines. Exact (blue), 8/9 (one-off; red), and 7/9 (two-off; yellow) coverage was computed for mono- and polyvalent vaccine candidates for Gag (Fig. 5A) and Nef (core) (Fig. 5B) for four test situations:

5 within-clade (C-clade-optimized candidates scored for C-clade coverage), between-clade (B-clade-optimized candidates scored for C-clade coverage), global-against-single-subtype (M-group-optimized candidates scored for C-clade coverage), global-against-global (M-group-optimized candidates scored for global coverage). Within each set of results, vaccine candidates are

10 grouped by number of sequences in the cocktail (1-6); mosaic sequences are plotted with darker colors. "Non-opt" refers to one set of sequences moving into vaccine trials (Kong et al, J. Virol. 77:12764-72 (2003)); "mosaic" denotes sequences generated by the genetic algorithm; "opt. natural" denotes intact natural sequences selected for maximum 9-mer coverage; "MBC

15 consensus" denotes a cocktail of 3 consensus sequences, for M-group, B-subtype, and C-subtype. For ease of comparison, a dashed line marks the coverage of a 4-sequence set of M-group mosaics (73.7–75.6%). Over 150 combinations of mosaic-number, virus subset, protein region, and optimization and test sets were tested. The C clade/B clade/M group comparisons

20 illustrated in this figure are generally representative of within-clade, between-clade, and M group coverage. In particular, levels of mosaic coverage for B and C clade were very similar, despite there being many more C clade sequences in the Gag collection, and many more B clade sequences in the Nef collection (see Fig. 6 for a full B and C clade comparison). There were

25 relatively few A and G clade sequences in the alignments (24 Gag, 75 Nef), and while 9-mer coverage by M-group optimized mosaics was not as high as for subtypes for B and C clades (4-mosaic coverage for A and G subtypes was 63% for Gag, 74% for Nef), it was much better than a non-optimal cocktail (52% Gag, 52% for Nef).

30

Figures 6A and 6B. Overall coverage of vaccine candidates: coverage of 9-mers in B-clade, C-clade, and M-group sequences using different input data sets for mosaic optimization, allowing different numbers of antigens, and comparing to different candidate vaccines. Exact (blue), 8/9 (one-off; red), and 7/9 (two-off; yellow) coverage was computed for mono- and polyvalent vaccine candidates for Gag (Fig. 6A) and Nef (core) (Fig. 6B) for seven test situations: within-clade (B- or C-clade-optimized candidates scored against the same clade), between-clade (B- or C-clade-optimized candidates scored against the other clade), global vaccine against single subtype (M-group-optimized candidates scored against B- or C-clade), global vaccine against global viruses (M-group-optimized candidates scored against all M-group sequences). Within each set of results, vaccine candidates are grouped by number of sequences in the cocktail (1-6); mosaic sequences are plotted with darker colors. “Non-opt” refers to a particular set of natural sequences previously proposed for a vaccine (Kong, W.P. et al. J Virol 77, 12764-72 (2003)); “mosaic” denotes sequences generated by the genetic algorithm; “opt. natural” denotes intact natural sequences selected for maximum 9-mer coverage; “MBC consensus” denotes a cocktail of 3 consensus sequences, for M-group, B-subtype, and C-subtype. A dashed line is shown at the level of exact-match M-group coverage for a 4-valent mosaic set optimized on the M-group.

Figures 7A and 7B. The distribution of 9-mers by frequency of occurrence in natural, consensus, and mosaic sequences. Occurrence counts (y-axis) for different 9-mer frequencies (x-axis) for vaccine cocktails produced by several methods. Fig. 7A: frequencies from 0-60% (for 9-mer frequencies > 60%, the distributions are equivalent for all methods). Fig. 7B: Details of low-frequency 9-mers. Natural sequences have large numbers of rare or unique-to-isolate 9-mers (bottom right, Figs. 7A and 7B); these are unlikely to induce useful vaccine responses. Selecting optimal natural sequences does

select for more common 9-mers, but rare and unique 9-mers are still included (top right, Figs. 7A and 7B). Consensus cocktails, in contrast, under-represent uncommon 9-mers, especially below 20% frequency (bottom left, Figs. 7A and 7B). For mosaic sequences, the number of lower-frequency 9-mers
 5 monotonically increases with the number of sequences (top left, each panel), but unique-to-isolate 9-mers are completely excluded (top left of right panel: * marks the absence of 9-mers with frequencies < 0.005).

Figures 8A-8D. HLA binding potential of vaccine candidates.

10 Figs. 8A and 8B) HLA binding motif counts. Figs. 8C and 8D) number of unfavorable amino acids. In all graphs: natural sequences are marked with black circles (λ); consensus sequences with blue triangles (σ); inferred ancestral sequences with green squares (∇); and mosaic sequences with red diamonds ($\blacklozenge$). Left panel (Figs. 8A and 8C) shows HLA-binding-motif
 15 counts (Fig. 8A) and counts of unfavorable amino acids (Fig. 8C) calculated for individual sequences; Right panel (Figs. 8B and 8D) shows HLA binding motifs counts (Fig. 8B) and counts of unfavorable amino acids (Fig. 8D) calculated for sequence cocktails. The top portion of each graph (box-and-whiskers graph) shows the distribution of respective counts (motif counts or
 20 counts of unfavorable amino acids) based either on alignment of M group sequences (for individual sequences, Figs. 8A and 8C) or on 100 randomly composed cocktails of three sequences, one from each A, B and C subtypes (for sequence cocktails, Figs. 8B and 8D). The alignment was downloaded from the Los Alamos HIV database. The box extends from the 25 percentile
 25 to the 75 percentile, with the line at the median. The whiskers extending outside the box show the highest and lowest values. Amino acids that are very rarely found as C-terminal anchor residues are G, S, T, P, N, Q, D, E, and H, and tend to be small, polar, or negatively charged (Yusim et al, J. Virol. 76:8757-8768 (2002)). Results are shown for Gag, but the same qualitative
 30 results hold for Nef core and complete Nef. The same procedure was done for

supertype motifs with results qualitatively similar to the results for HLA binding motifs (data not shown).

Figure 9. Mosaic protein sets limited to 4 sequences ($k=4$), spanning
5 Gag and the central region of Nef, optimized for subtype B, subtype C, and the M group.

Figure 10. Mosaic sets for Env and Pol.

10 Figure 11. This plot is alignment independent, based on splintering all M group proteins, (database and CHAVI, one sequence per person) into all possible 9-mers, attending to their frequencies, and then looking for matches and near matches in each vaccine antigen or cocktail with the database.

15 Figure 12. Additional summaries of coverage.

Figure 13. 9-mer coverage by position (Mos.3 vaccine cocktail).

Figures 14A-14D. Plots resorted by frequency of 9-mer matches for
20 each vaccine proposed for use.

Figures 15A-15D. Plots mapping every amino acid in every sequence in the full database alignment.

25 Figure 16. 3 Mosaic, M group Optimizations.

Figure 17. Coverage of the HIV database plus CHAVI sequences (N = 2020).

30 Figure 18. Differences in acute infection patient sequences compared

to patient consensus.

Figure 19. The compromise and benefit in terms of coverage for Env M group versus subtype-specific design.

5

Figure 20. Proposed vaccine mosaic coverage of Gag and Env.

Figure 21. Gag, Nef and Env sequences.

10 Figure 22. Mosaic *gag* and *nef* genes and M consensus *gag* and *nef* genes.

DETAILED DESCRIPTION OF THE INVENTION

The present invention results from the realization that a polyvalent set of antigens comprising synthetic viral proteins, the sequences of which
15 provide maximum coverage of non-rare short stretches of circulating viral sequences, constitutes a good vaccine candidate. The invention provides a “genetic algorithm” strategy to create such sets of polyvalent antigens as mosaic blends of fragments of an arbitrary set of natural protein sequences provided as inputs. In the context of HIV, the proteins Gag and Nef are ideal
20 candidates for such antigens. To expand coverage, Pol and/or Env can also be used. The invention further provides optimized sets for these proteins.

The genetic algorithm strategy of the invention uses unaligned protein sequences from the general population as an input data set, and thus has the virtue of being “alignment independent”. It creates artificial mosaic proteins
25 that resemble proteins found in nature – the success of the consensus antigens in small animals models suggest this works well. 9 Mers are the focus of the studies described herein, however, different length peptides can be selected depending on the intended target. In accordance with the present approach, 9 mers (for example) that do not exist in nature or that are very rare can be

excluded – this is an improvement relative to consensus sequences since the latter can contain some 9 mers (for example) that have not been found in nature, and relative to natural strains that almost invariably contain some 9 mers (for example) that are unique to that strain. The definition of fitness used for the genetic algorithm is that the most “fit” polyvalent cocktail is the combination of mosaic strains that gives the best coverage (highest fraction of perfect matches) of all of the 9 mers in the population and is subject to the constraint that no 9 mer is absent or rare in the population.

The mosaics protein sets of the invention can be optimized with respect to different input data sets – this allows use of current data to assess virtues of a subtype or region specific vaccines from a T cell perspective. By way of example, options that have been compared include:

- 1) Optimal polyvalent mosaic sets based on M group, B clade and C clade. The question presented was how much better is intra-clade coverage than inter-clade or global.
- 2) Different numbers of antigens: 1, 3, 4, 6
- 3) Natural strains currently in use for vaccine protocols just to exemplify “typical” strains (Merck, VRC)
- 4) Natural strains selected to give the best coverage of 9-mers in a population
- 5) Sets of consensus: A+B+C..
- 6) Optimized cocktails that include one “given” strain in a polyvalent antigen, one ancestral + 3 mosaic strains, one consensus + 3 mosaic strains.
- 7) Coverage of 9 mers that were perfectly matched was compared with those that match 8/9, 7/9, and 6/9 or less.

This is a computationally difficult problem, as the best set to cover one 9-mer may not be the best set to cover overlapping 9-mers.

It will be appreciated from a reading of this disclosure that the approach described herein can be used to design peptide reagents to test HIV

immune responses, and be applied to other variable pathogens as well. For example, the present approach can be adapted to the highly variable virus Hepatitis C.

5 The proteins/polypeptides/peptides ("immunogens") of the invention can be formulated into compositions with a pharmaceutically acceptable carrier and/or adjuvant using techniques well known in the art. Suitable routes of administration include systemic (e.g. intramuscular or subcutaneous), oral, intravaginal, intrarectal and intranasal.

10 The immunogens of the invention can be chemically synthesized and purified using methods which are well known to the ordinarily skilled artisan. The immunogens can also be synthesized by well-known recombinant DNA techniques.

Nucleic acids encoding the immunogens of the invention can be used as components of, for example, a DNA vaccine wherein the encoding
15 sequence is administered as naked DNA or, for example, a minigene encoding the immunogen can be present in a viral vector. The encoding sequences can be expressed, for example, in mycobacterium, in a recombinant chimeric adenovirus, or in a recombinant attenuated vesicular stomatitis virus. The encoding sequence can also be present, for example, in a replicating or non-
20 replicating adenoviral vector, an adeno-associated virus vector, an attenuated mycobacterium tuberculosis vector, a Bacillus Calmette Guerin (BCG) vector, a vaccinia or Modified Vaccinia Ankara (MVA) vector, another pox virus vector, recombinant polio and other enteric virus vector, Salmonella species bacterial vector, Shigella species bacterial vector, Venezuelan Equine
25 Encephalitis Virus (VEE) vector, a Semliki Forest Virus vector, or a Tobacco Mosaic Virus vector. The encoding sequence, can also be expressed as a DNA plasmid with, for example, an active promoter such as a CMV promoter. Other live vectors can also be used to express the sequences of the invention. Expression of the immunogen of the invention can be induced in a patient's
30 own cells, by introduction into those cells of nucleic acids that encode the

immunogen, preferably using codons and promoters that optimize expression in human cells. Examples of methods of making and using DNA vaccines are disclosed in U.S. Pat. Nos. 5,580,859, 5,589,466, and 5,703,055. Examples of methods of codon optimization are described in Haas et al, Current Biology
5 6:315-324 (1996) and in Andre et al, J. Virol. 72(2):1497-1503 (1998).

It will be appreciated that adjuvants can be included in the compositions of the invention (or otherwise administered to enhance the immunogenic effect). Examples of suitable adjuvants include TRL-9 agonists, TRL-4 agonists, and TRL-7, 8 and 9 agonist combinations (as well as alum).
10 Adjuvants can take the form of oil and water emulsions. Squalene adjuvants can also be used.

The composition of the invention comprises an immunologically effective amount of the immunogen of this invention, or nucleic acid sequence encoding same, in a pharmaceutically acceptable delivery system. The
15 compositions can be used for prevention and/or treatment of virus infection (e.g. HIV infection). As indicated above, the compositions of the invention can be formulated using adjuvants, emulsifiers, pharmaceutically-acceptable carriers or other ingredients routinely provided in vaccine compositions. Optimum formulations can be readily designed by one of ordinary skill in the
20 art and can include formulations for immediate release and/or for sustained release, and for induction of systemic immunity and/or induction of localized mucosal immunity (e.g, the formulation can be designed for intranasal, intravaginal or intrarectal administration). As noted above, the present compositions can be administered by any convenient route including
25 subcutaneous, intranasal, oral, intramuscular, or other parenteral or enteral route. The immunogens can be administered as a single dose or multiple doses. Optimum immunization schedules can be readily determined by the ordinarily skilled artisan and can vary with the patient, the composition and the effect sought.

The invention contemplates the direct use of both the immunogen of the invention and/or nucleic acids encoding same and/or the immunogen expressed as indicated above. For example, a minigene encoding the immunogen can be used as a prime and/or boost.

5 The invention includes any and all amino acid sequences disclosed herein, as well as nucleic acid sequences encoding same (and nucleic acids complementary to such encoding sequences).

Specifically disclosed herein are vaccine antigen sets optimized for single B or C subtypes, targeting regional epidemics, as well as for all HIV-1
10 variants in global circulation [the HIV-1 Main (M) group]. In the study described in Example 1 that follows, the focus is on designing polyvalent vaccines specifically for T-cell responses. HIV-1 specific T-cells are likely to be crucial to an HIV-1-specific vaccine response: CTL responses are correlated with slow disease progression in humans (Oxenius et al, J. Infect.
15 Dis. 189:1199-1208 (2004)), and the importance of CTL responses in non-human primate vaccination models is well-established. Vaccine elicited cellular immune responses help control pathogenic SIV or SHIV, and reduce the likelihood of disease after challenge with pathogenic virus (Barouch et al, Science 290:486-492 (2000)). Temporary depletion of CD8+ T cells results in
20 increased viremia in SIV-infected rhesus macaques (Schmitz et al, Science 283:857-860 (1999)). Furthermore, the evolution of escape mutations has been associated with disease progression, indicating that CTL responses help constrain viral replication *in vivo* (Barouch et al, J. Virol. 77:7367-7375 (2003)), and so vaccine-stimulated memory responses that could block
25 potential escape routes may be of value. While the highly variable Envelope (Env) is the primary target for neutralizing antibodies against HIV, and vaccine antigens will also need to be tailored to elicit these antibody responses (Moore & Burton, Nat. Med. 10:769-771 (2004)), T-cell vaccine components can target more conserved proteins to trigger responses that are more likely to
30 cross-react. But even the most conserved HIV-1 proteins are diverse enough

that variation will be an issue. Artificial central-sequence vaccine approaches, consensus and ancestral sequences (Gaschen et al, Science 296:2354-2360 (2002), Gao et al, J. Virol. 79:1154-1163 (2005), Doria-Rose et al, J. Virol. 79:11214-11224 (2005)), which essentially “split the differences” between
5 strains, show promise, stimulating responses with enhanced cross-reactivity compared to natural strain vaccines (Gao et al, J. Virol. 79:1154-1163 (2005)) (Liao *et al.* and Weaver *et al.*, submitted.) Nevertheless, even central strains cover the spectrum of HIV diversity to a very limited extent, and consensus-based peptide reagents fail to detect many autologous CD8+ T-cell responses
10 (Altfeld et al, J. Virol. 77:7330-7340 (2003)).

A single amino acid substitution can mediate T-cell escape, and as one or more amino acids in many T-cell epitopes differ between HIV-1 strains, the potential effectiveness of responses to any one vaccine antigen is limited. Whether a particular mutation will diminish T-cell cross-reactivity is epitope-
15 and T-cell-specific, although some changes can broadly affect between-clade cross-reactivity (Norris et al, AIDS Res. Hum. Retroviruses 20:315-325 (2004)). Including more variants in a polyvalent vaccine could enable responses to a broader range of circulating variants. It could also prime the immune system against common escape variants (Jones et al, J. Exp. Med.
20 200:1243-1256 (2004)); escape from one T-cell receptor might create a variant that is susceptible to another (Lee et al, J. Exp. Med. 200:1455-1466 (2004)), thus stimulating polyclonal responses to epitope variants may be beneficial (Killian et al, AIDS 19:887-896 (2005)). Immune escape involving avenues that inhibit processing (Milicic et al, J. Immunol. 175:4618-4626 (2005)) or
25 HLA binding (Ammaranond et al, AIDS Res. Hum. Retroviruses 21:395-397 (2005)) prevent epitope presentation, and in such cases the escape variant could not be countered by a T-cell with a different specificity. However, it is possible the presence of T-cells that recognize overlapping epitopes may in some cases block these even escape routes.

Certain aspects of the invention can be described in greater detail in the non-limiting Examples that follow.

EXAMPLE 1

Experimental Details

5 *HIV-1 sequence data.* The reference alignments from the 2005 HIV sequence database (<http://hiv.lanl.gov>), which contain one sequence per person, were used, supplemented by additional recently available C subtype Gag and Nef sequences from Durban, South Africa (GenBank accession numbers AY856956-AY857186) (Kiepiela et al, Nature 432:769-75 (2004)).
10 This set contained 551 Gag and 1,131 Nef M group sequences from throughout the globe; recombinant sequences were included as well as pure subtype sequences for exploring M group diversity. The subsets of these alignments that contained 18 A, 102 B, 228 C, and 6 G subtype (Gag), and 62 A, 454 B, 284 C, and 13 G subtype sequences (Nef) sequences were used for
15 within- and between-single-clade optimizations and comparisons.

The genetic algorithm. GAs are computational analogues of biological processes (evolution, populations, selection, recombination) used to find solutions to problems that are difficult to solve analytically (Holland, Adaptation in Natural and Artificial Systems: An Introductory Analysis with
20 Applicatins to Biology, Control, and Artificial Intelligence, (M.I.T. Press, Cambridge, MA (1992))). Solutions for a given input are “evolved” though a process of random modification and selection according to a “fitness” (optimality) criterion. GAs come in many flavors; a “steady-state co-evolutionary multi-population” GA was implemented. “Steady-state” refers to
25 generating one new candidate solution at a time, rather than a whole new population at once; and “co-evolutionary” refers to simultaneously evolving several distinct populations that work together to form a complete solution. The input is an unaligned set of natural sequences; a candidate solution is a set

of k pseudo-natural "mosaic" sequences, each of which is formed by concatenating sections of natural sequences. The fitness criterion is *population coverage*, defined as the proportion of all 9-amino-acid sequence fragments (potential epitopes) in the input sequences that are found in the cocktail.

To initialize the GA (Fig. 2), k populations of n initial candidate sequences are generated by 2-point recombination between randomly selected natural sequences. Because the input natural sequences are not aligned, "homologous" crossover is used: crossover points in each sequence are selected by searching for short matching strings in both sequences; strings of $c-1 = 8$, were used where a typical epitope length is $c = 9$. This ensures that the recombined sequences resemble natural proteins: the boundaries between sections of sequence derived from different strains are seamless, the local sequences spanning the boundaries are always found in nature, and the mosaics are prevented from acquiring large insertions/deletions or unnatural combinations of amino acids. Mosaic sequence lengths fall within the distribution of natural sequence lengths as a consequence of mosaic construction: recombination is only allowed at identical regions, reinforced by an explicit software prohibition against excessive lengths to prevent reduplication of repeat regions. (Such "in frame" insertion of reduplicated epitopes could provide another way of increasing coverage without generating unnatural 9-mers, but their inclusion would create "unnatural" proteins.) Initially, the cocktail contains one randomly chosen "winner" from each population. The fitness score for any individual sequence in a population is the coverage value for the cocktail consisting of that sequence plus the current winners from the other populations. The individual fitness of any sequence in a population therefore depends dynamically upon the best sequences found in the other populations.

Optimization proceeds one population at a time. For each iteration, two "parent" sequences are chosen. The first parent is chosen using "2-

tournament" selection: two sequences are picked at random from the current population, scored, and the better one is chosen. This selects parents with a probability inversely proportional to their fitness rank within the population, without the need to actually compute the fitness of all individuals. The second parent is chosen in the same way (50% of the time), or is selected at random from the set of natural sequences. 2-point homologous crossover between the parents is then used to generate a "child" sequence. Any child containing a 9-mer that was very rare in the natural population (found less than 3 times) is rejected immediately. Otherwise, the new sequence is scored, and its fitness is compared with the fitnesses of four randomly chosen sequences from the same population. If any of the four randomly chosen sequences has a score lower than that of the new sequence, it is replaced in the population by the new sequence. Whenever a sequence is encountered that yields a better score than the current population "winner", that sequence becomes the winner for the current population and so is subsequently used in the cocktail to evaluate sequences in other populations. A few such optimization cycles (typically 10) are applied to each population in turn, and this process continues cycling through the populations until evolution stalls (*i.e.*, no improvement has been made for a defined number of generations). At this point, the entire procedure is restarted using newly generated random starting populations, and the restarts are continued until no further improvement is seen. The GA was run on each data set with $n = 50$ or 500 ; each run was continued until no further improvement occurred for 12-24 hours on a 2 GHz Pentium processor. Cocktails were generated having $k = 1, 3, 4$, or 6 mosaic sequences.

The GA also enables optional inclusion of one or more fixed sequences of interest (for example, a consensus) in the cocktail and will evolve the other elements of the cocktail in order to optimally complement that fixed strain. As these solutions were suboptimal, they are not included here. An additional program selects from the input file the k best natural strains that in combination provide the best population coverage.

Comparison with other polyvalent vaccine candidates. Population coverage scores were computed for other potential mono- or polyvalent vaccines to make direct comparisons with the mosaic-sequence vaccines, tracking identities with population 9-mers, as well as similarities of 8/9 and 7/9 amino acids. Potential vaccine candidates based on natural strains include single strains (for example, a single C strain for a vaccine for southern Africa (Williamson et al, AIDS Res. Hum. Retroviruses 19:133-44 (2003))) or combinations of natural strains (for example, one each of subtype A, B, and C (Kong et al, J. Virol. 77:12764-72 (2003))). To date, natural-strain vaccine candidates have not been systematically selected to maximize potential T-cell epitope coverage; vaccine candidates were picked from the literature to be representative of what could be expected from unselected vaccine candidates. An upper bound for coverage was also determined using only intact natural strains: optimal natural-sequence cocktails were generated by selecting the single sequence with the best coverage of the dataset, and then successively adding the most complementary sequences up to a given k . The comparisons included optimal natural-sequence cocktails of various sizes, as well as consensus sequences, alone or in combination (Gaschen et al, Science 296:2354-60 (2002)), to represent the concept of central, synthetic vaccines. Finally, using the fixed-sequence option in the GA, consensus-plus-mosaic combinations in the comparisons; these scores were essentially equivalent to all-mosaic combinations were included for a given k (data not shown). The code used for performing these analyses are available at: <ftp://ftp-t10/pub/btk/mosaics>.

Results

Protein Variation. In conserved HIV-1 proteins, most positions are essentially invariant, and most variable positions have only two to three amino acids that occur at appreciable frequencies, and variable positions are generally well dispersed between conserved positions. Therefore, within the

boundaries of a CD8+ T-cell epitope (8-12 amino acids, typically nine), most of the population diversity can be covered with very few variants. Figure 1 shows an upper bound for population coverage of 9-mers (stretches of nine contiguous amino acids) comparing Gag, Nef, and Env for increasing numbers of variants, sequentially adding variants that provide the best coverage. In conserved regions, a high degree of population coverage is achieved with 2-4 variants. By contrast, in variable regions like Env, limited population coverage is possible even with eight variants. Since each new addition is rarer, the relative benefits of each addition diminish as the number of variants increases.

Vaccine design optimization strategies. Figure 1 shows an idealized level of 9-mer coverage. In reality, high-frequency 9-mers often conflict: because of local co-variation, the optimal amino acid for one 9-mer may differ from that for an overlapping 9-mer. To design mosaic protein sets that optimize population coverage, the relative benefits of each amino acid must be evaluated in combination with nearby variants. For example, Alanine (Ala) and Glutamate (Glu) might each frequently occur in adjacent positions, but if the Ala-Glu combination is never observed in nature, it should be excluded from the vaccine. Several optimization strategies were investigated: a greedy algorithm, a semi-automated compatible-9mer assembly strategy, an alignment-based genetic algorithm (GA), and an alignment-independent GA.

The alignment-independent GA generated mosaics with the best population coverage. This GA generates a user-specified number of mosaic sequences from a set of unaligned protein sequences, explicitly excluding rare or unnatural epitope-length fragments (potentially introduced at recombination breakpoints) that could induce non-protective vaccine-antigen-specific responses. These candidate vaccine sequences resemble natural proteins, but are assembled from frequency-weighted fragments of database sequences recombined at homologous breakpoints (Fig. 2); they approach maximal coverage of 9-mers for the input population.

Selecting HIV protein regions for an initial mosaic vaccine. The initial design focused on protein regions meeting specific criteria: i) relatively low variability, ii) high levels of recognition in natural infection, iii) a high density of known epitopes and iv) either early responses upon infection *or* CD8+ T-cell responses associated with good outcomes in infected patients. First, an assessment was made of the level of 9-mer coverage achieved by mosaics for different HIV proteins (Fig. 3). For each protein, a set of four mosaics was generated using either the M group or the B- and C-subtypes alone; coverage was scored on the C subtype. Several results are notable: i) within-subtype optimization provides the best within-subtype coverage, but substantially poorer between-subtype coverage — nevertheless, B-subtype-optimized mosaics provide better C-subtype coverage than a single natural B subtype protein (Kong et al, J. Virol. 77:12764-72 (2003)); ii) Pol and Gag have the most potential to elicit broadly cross-reactive responses, whereas Rev, Tat, and Vpu have even fewer conserved 9-mers than the highly variable Env protein, iii) within-subtype coverage of M-group-optimized mosaic sets approached coverage of within-subtype optimized sets, particularly for more conserved proteins.

Gag and the central region of Nef meet the four criteria listed above. Nef is the HIV protein most frequently recognized by T-cells (Frahm et al, J. Virol. 78:2187-200 (2004)) and the target for the earliest response in natural infection (Lichterfeld et al, Aids 18:1383-92 (2004)). While overall it is variable (Fig. 3), its central region is as conserved as Gag (Fig. 1). It is not yet clear what optimum proteins for inclusion in a vaccine might be, and mosaics could be designed to maximize the potential coverage of even the most variable proteins (Fig. 3), but the prospects for global coverage are better for conserved proteins. Improved vaccine protection in macaques has been demonstrated by adding Rev, Tat, and Nef to a vaccine containing Gag, Pol, and Env (Hel et al, J. Immunol. 176:85-96 (2006)), but this was in the context of homologous challenge, where variability was not an issue. The extreme

variability of regulatory proteins in circulating virus populations may preclude cross-reactive responses; in terms of conservation, Pol, Gag (particularly p24) and the central region of Nef (HXB2 positions 65-149) are promising potential immunogens (Figs. 1,3). Pol, however, is infrequently recognized during natural infection (Frahm et al, J. Virol. 78:2187-200 (2004)), so it was not included in the initial immunogen design. The conserved portion of Nef that were included contains the most highly recognized peptides in HIV-1 (Frahm et al, J. Virol. 78:2187-200 (2004)), but as a protein fragment, would not allow Nef's immune inhibitory functions (*e.g.* HLA class I down-regulation (Blagoveshchenskaya, Cell 111:853-66 (2002))). Both Gag and Nef are densely packed with overlapping well-characterized CD8+ and CD4+ T-cell epitopes, presented by many different HLA molecules (<http://www.hiv.lanl.gov/content/immunology/maps/maps.html>), and Gag-specific CD8+ (Masemola et al, J. Virol. 78:3233-43 (2004)) and CD4+ (Oxenius et al, J. Infect. Dis. 189:1199-208 (2004)) T-cell responses have been associated with low viral set points in infected individuals (Masemola et al, J. Virol. 78:3233-43 (2004)).

To examine the potential impact of geographic variation and input sample size, a limited test was done using published subtype C sequences. The subtype C Gag data were divided into three sets of comparable size – two South African sets (Kiepiela et al, Nature 432:769-75 (2004)), and one non-South-African subtype C set. Mosaics were optimized independently on each of the sets, and the resulting mosaics were tested against all three sets. The coverage of 9-mers was slightly better for identical training and test sets (77-79% 9/9 coverage), but essentially equivalent when the training and test sets were the two different South African data sets (73-75%), or either of the South African sets and the non-South African C subtype sequences (74-76%). Thus between- and within-country coverage approximated within-clade coverage, and in this case no advantage to a country-specific C subtype mosaic design was found.

Designing mosaics for Gag and Nef and comparing vaccine strategies.

To evaluate within- and between-subtype cross-reactivity for various vaccine design strategies, a calculation was made of the coverage they provided for natural M-Group sequences. The fraction of all 9-mers in the natural sequences that were perfectly matched by 9-mers in the vaccine antigens were computed, as well as those having 8/9 or 7/9 matching amino acids, since single (and sometimes double) substitutions within epitopes may retain cross-reactivity. Figure 4 shows M group coverage per 9-mer in Gag and the central region of Nef for cocktails designed by various strategies: a) three non-optimal natural strains from the A, B, and C subtypes that have been used as vaccine antigens (Kong et al, J. Virol. 77:12764-72 (2003)); b) three natural strains that were computationally selected to give the best M group coverage; c) M group, B subtype, and C subtype consensus sequences; and, d,e,f) three, four and six mosaic proteins. For cocktails of multiple strains, sets of $k=3$, $k=4$, and $k=6$, the mosaics clearly perform the best, and coverage approaches the upper bound for k strains. They are followed by optimally selected natural strains, the consensus protein cocktail, and finally, non-optimal natural strains. Allowing more antigens provides greater coverage, but gains for each addition are reduced as k increases (Figs. 1 and 4).

Figure 5 summarizes total coverage for the different vaccine design strategies, from single proteins through combinations of mosaic proteins, and compares within-subtype optimization to M group optimization. The performance of a single mosaic is comparable to the best single natural strain or a consensus sequence. Although a single consensus sequence out-performs a single best natural strain, the optimized natural-sequence cocktail does better than the consensus cocktail: the consensus sequences are more similar to each other than are natural strains, and are therefore somewhat redundant. Including even just two mosaic variants, however, markedly increases coverage, and four and six mosaic proteins give progressively better coverage than polyvalent cocktails of natural or consensus strains. Within-subtype

optimized mosaics perform best – with four mosaic antigens 80-85% of the 9-mers are perfectly matched – but between-subtype coverage of these sets falls off dramatically, to 50-60%. In contrast, mosaic proteins optimized using the full M group give coverage of approximately 75-80% for individual subtypes, comparable to the coverage of the M group as a whole (Figs. 5 and 6). If imperfect 8/9 matches are allowed, both M group optimized and within-subtype optimized mosaics approach 90% coverage.

Since coverage is increased by adding progressively rarer 9-mers, and rare epitopes may be problematic (*e.g.*, by inducing vaccine-specific immunodominant responses), an investigation was made of the frequency distribution of 9-mers in the vaccine constructs relative to the natural sequences from which they were generated. Most additional epitopes in a $k=6$ cocktail compared to a $k=4$ cocktail are low-frequency (<0.1 , Fig. 7). Despite enhancing coverage, these epitopes are relatively rare, and thus responses they induce might draw away from vaccine responses to more common, thus more useful, epitopes. Natural-sequence cocktails actually have fewer occurrences of moderately low-frequency epitopes than mosaics, which accrue some lower frequency 9-mers as coverage is optimized. On the other hand, the mosaics exclude unique or very rare 9-mers, while natural strains generally contain 9-mers present in no other sequence. For example, natural M group Gag sequences had a median of 35 (range 0-148) unique 9-mers per sequence. Retention of HLA-anchor motifs was also explored, and anchor motif frequencies were found to be comparable between four mosaics and three natural strains. Natural antigens did exhibit an increase in number of motifs per antigen, possibly due to inclusion of strain-specific motifs (Fig. 8).

The increase in ever-rarer epitopes with increasing k , coupled with concerns about vaccination-point dilution and reagent development costs, resulted in the initial production of mosaic protein sets limited to 4 sequences ($k=4$), spanning Gag and the central region of Nef, optimized for subtype B, subtype C, and the M group (these sequences are included in Fig. 9; mosaic

sets for Env and Pol are set forth in Fig. 10). Synthesis of various four-sequence Gag-Nef mosaics and initial antigenicity studies are underway. In the initial mosaic vaccine, targeted are just Gag and the center of the Nef protein, which are conserved enough to provide excellent global population coverage, and have the desirable properties described above in terms of natural responses (Bansal et al, Aids 19:241-50 (2005)). Additionally, including B subtype p24 variants in Elispot peptide mixtures to detect natural CTL responses to infection significantly enhanced both the number and the magnitude of responses detected supporting the idea that including variants of even the most conserved proteins will be useful. Finally, cocktails of proteins in a polyvalent HIV-1 vaccine given to rhesus macaques did not interfere with the development of robust responses to each antigen (Seaman et al, J. Virol. 79:2956-63 (2005)), and antigen cocktails did not produce antagonistic responses in murine models (Singh et al, J. Immunol. 169:6779-86 (2002)), indicating that antigenic mixtures are appropriate for T-cell vaccines.

Even with mosaics, variable proteins like Env have limited coverage of 9-mers, although mosaics improve coverage relative to natural strains. For example three M group natural proteins, one each selected from the A, B, and C clades, and currently under study for vaccine design (Seaman et al, J. Virol. 79:2956-63 (2005)) perfectly match only 39% of the 9-mers in M group proteins, and 65% have at least 8/9 matches. In contrast, three M group Env mosaics match 47% of 9-mers perfectly, and 70% have at least an 8/9 match. The code written to design polyvalent mosaic antigens is available, and could readily be applied to any input set of variable proteins, optimized for any desired number of antigens. The code also allows selection of optimal combinations of k natural strains, enabling rational selection of natural antigens for polyvalent vaccines. Included in Table 1 are the best natural strains for Gag and Nef population coverage of current database alignments.

TABLE 1

Natural sequence cocktails having the best available 9-mer coverage for different genes, subtype sets, and numbers of sequences

Gag, B-subtype, 1 natural sequence

B.US.86.AD87_AF004394

Gag, B-subtype, 3 natural sequences

B.US.86.AD87_AF004394
B.US.97.Ac_06_AY247251
B.US.88.WR27_AF286365

Gag, B-subtype, 4 natural sequences

B.US.86.AD87_AF004394
B.US.97.Ac_06_AY247251
B.US._R3_PDC1_AY206652
B.US.88.WR27_AF286365

Gag, B-subtype, 6 natural sequences

B.CN._CNHN24_AY180905
B.US.86.AD87_AF004394
B.US.97.Ac_06_AY247251
B.US._P2_AY206654
B.US._R3_PDC1_AY206652
B.US.88.WR27_AF286365

Gag, C-subtype, 1 natural sequence

C.IN._70177_AF533131

Gag, C-subtype, 3 natural sequences

C.ZA.97.97ZA012
C.ZA.x.04ZASK161B1
C.IN._70177_AF533131

Gag, C-subtype, 4 natural sequences

C.ZA.97.97ZA012
C.ZA.x.04ZASK142B1
C.ZA.x.04ZASK161B1
C.IN._70177_AF533131

Gag, C-subtype, 6 natural sequences

C.ZA.97.97ZA012
C.ZA.x.04ZASK142B1
C.ZA.x.04ZASK161B1
C.BW.99.99BWMC168_AF443087
C.IN._70177_AF533131
C.IN._MYA1_AF533139

Gag, M-group, 1 natural sequence

C.IN_.70177_AF533131

Gag, M-group, 3 natural sequences

B.US.90.US2_AY173953
C.IN_.70177_AF533131
15_01B.TH.99.99TH_R2399_AF530576

Gag, M-group, 4 natural sequences

B.US.90.US2_AY173953
C.IN_.70177_AF533131
C.IN.93.93IN999_AF067154
15_01B.TH.99.99TH_R2399_AF530576

Gag, M-group, 6 natural sequences

C.ZA.x.04ZASK138B1
B.US.90.US2_AY173953
B.US_.WT1_PDC1_AY206656
C.IN_.70177_AF533131
C.IN.93.93IN999_AF067154
15_01B.TH.99.99TH_R2399_AF530576

Nef (central region), B-subtype, 1 natural sequence

B.GB.94.028jh_94_1_NP_AF129346

Nef (central region), B-subtype, 3 natural sequences

B.GB.94.028jh_94_1_NP_AF129346
B.KR.96.96KCS4_AY121471
B.FR.83.HXB2_K03455

Nef (central region), B-subtype, 4 natural sequences

B.GB.94.028jh_94_1_NP_AF129346
B.KR.96.96KCS4_AY121471
B.US.90.E90NEF_U43108
B.FR.83.HXB2_K03455

Nef (central region), B-subtype, 6 natural sequences

B.GB.94.028jh_94_1_NP_AF129346
B.KR.02.02HYJ3_AY121454
B.KR.96.96KCS4_AY121471
B.CN_.RL42_U71182
B.US.90.E90NEF_U43108
B.FR.83.HXB2_K03455

Nef (central region), C-subtype, 1 natural sequence

C.ZA.04.04ZASK139B1

Nef (central region), C-subtype, 3 natural sequences

C.ZA.04.04ZASK180B1
C.ZA.04.04ZASK139B1
C.ZA._.ZASW15_AF397568

Nef (central region), C-subtype, 4 natural sequences

C.ZA.97.ZA97004_AF529682
C.ZA.04.04ZASK180B1
C.ZA.04.04ZASK139B1
C.ZA._.ZASW15_AF397568

Nef (central region), C-subtype, 6 natural sequences

C.ZA.97.ZA97004_AF529682
C.ZA.00.1192M3M
C.ZA.04.04ZASK180B1
C.ZA.04.04ZASK139B1
C.04ZASK184B1
C.ZA._.ZASW15_AF397568

Nef (central region), M-group, 1 natural sequence

B.GB.94.028jh_94_1_NP_AF129346

Nef (central region), M-group, 3 natural sequences

02_AG.CM._.98CM1390_AY265107
C.ZA.03.03ZASK020B2
B.GB.94.028jh_94_1_NP_AF129346

Nef (central region), M-group, 4 natural sequences

02_AG.CM._.98CM1390_AY265107
01A1.MM.99.mCSW105_AB097872
C.ZA.03.03ZASK020B2
B.GB.94.028jh_94_1_NP_AF129346

Nef (central region), M-group, 6 natural sequences

02_AG.CM._.98CM1390_AY265107
01A1.MM.99.mCSW105_AB097872
C.ZA.03.03ZASK020B2
C.03ZASK111B1
B.GB.94.028jh_94_1_NP_AF129346
B.KR.01.01CWS2_AF462757

Summarizing, the above-described study focuses on the design of T-cell vaccine components to counter HIV diversity at the moment of infection, and to block viral escape routes and thereby minimize disease progression in infected individuals. The polyvalent mosaic protein strategy developed here
5 for HIV-1 vaccine design could be applied to any variable protein, to other pathogens, and to other immunological problems. For example, incorporating a minimal number of variant peptides into T-cell response assays could markedly increase sensitivity without excessive cost: a set of k mosaic proteins provides the maximum coverage possible for k antigens.

10 A centralized (consensus or ancestral) gene and protein strategy has been proposed previously to address HIV diversity (Gaschen et al, Science 296:2354-2360 (2002)). Proof-of-concept for the use of artificial genes as immunogens has been demonstrated by the induction of both T and B cell responses to wild-type HIV-1 strains by group M consensus immunogens
15 (Gaschen et al, Science 296:2354-2360 (2002), Gao et al, J. Virol. 79:1154-63 (2005), Doria-Rose et al, J. Virol. 79:11214-24 (2005), Weaver et al, J. Virol., in press)). The mosaic protein design improves on consensus or natural immunogen design by co-optimizing reagents for a polyclonal vaccine, excluding rare CD8⁺ T-cell epitopes, and incorporating variants that, by virtue
20 of their frequency at the population level, are likely to be involved in escape pathways.

The mosaic antigens maximize the number of epitope-length variants that are present in a small, practical number of vaccine antigens. The decision was made to use multiple antigens that resemble native proteins, rather than
25 linking sets of concatenated epitopes in a poly-epitope pseudo-protein (Hanke et al, Vaccine 16:426-35 (1998)), reasoning that *in vivo* processing of native-like vaccine antigens will more closely resemble processing in natural infection, and will also allow expanded coverage of overlapping epitopes. T-cell mosaic antigens would be best employed in the context of a strong

polyvalent immune response; improvements in other areas of vaccine design and a combination of the best strategies, incorporating mosaic antigens to cover diversity, may ultimately enable an effective cross-reactive vaccine-induced immune response against HIV-1.

5

EXAMPLE 2

Group M consensus envelope and trivalent mosaic envelopes (both of which were designed by *in silico* modeling and are predicted to be superior than wildtype envelopes) will be compared to a monovalent wild-type envelope and trivalent wild-type transmitted envelopes in a 4 arm
10 immunogenicity clinical trial. The mosaic antigens have been designed based on the current Los Alamos database, a set that includes more full length envelopes sampled globally from more than 2000 individuals with a large set of sequences of transmitted viruses primarily from the CHAVI database.

The selection of the natural strains to be used for the comparison is
15 based on the following criteria: For the monovalent natural antigen, use will be made of the single transmitted virus that is the best choice in terms of providing coverage of potential T cell epitopes in the global database. The database is biased towards B clade envelopes, so the single best acute Env is a B clade representative. One A, one B and one C subtype transmitted virus
20 sequence is proposed for inclusion in the trivalent set, to compensate for the biases in sampling inherent in the global sequence collection, and to better reflect the circulating pandemic strains. The A and C natural sequences are those that optimally complement the best B clade sequence to provide potential epitope coverage of the database. Vaccine antigens have been
25 selected from among available SGA sequenced acute samples, each representing a transmitted virus. Therefore, this study, although primarily a T cell study, will also provide important additional data regarding the ability of transmitted envelope vaccines to elicit neutralizing antibodies.

For a mosaic/consensus human trial, the following 4 arm trial is proposed; 20 people per group, with a negative control:

- 5 1) Con S (a well studied consensus of the consensus of each clade, based on the 2002 database; Con S has been extensively tested in animal models, and has theoretical coverage roughly comparable to a single mosaic.)
- 10 2) A 3 mosaic M group antigen set designed to, in combination, provide optimal global coverage of 9 amino acid long stretches in the database. Such 9-mers represent potential epitope coverage of the database. Unnatural 9-mers are excluded in mosaics, and rare variants minimized.
- 15 3) The optimal single best natural protein selected from sequences sampled from acutely infected patients with SGA sequences available; these sequences should correspond to viable, transmitted sequences. As in (2), this sequence will be selected to be the one that provides optimal 9-mer coverage of the database. The B clade currently dominates sampling for the sequence database, so the sequence with the best database coverage will be a B clade sequence.
- 20 4) The best natural strains from acute infection SGA sequences that in combination provide the best global coverage. (Note: the B and C dominate the M group sampling hence the code naturally selects one of each as the two best. Thus, the third complementary sequence was forced to be selected from an acute SGA A clade set, to counter this bias and better reflect the global epidemic).
- 25 5) Negative control buffer/saline

The current M group alignment in the HIV database was combined with all of the newer CHAVI sequences – this includes a total of 2020 sequences:

30

728 B clade

599 C clade

693 that are all other clades, circulating recombinant forms, and unique recombinants. This was used for the M group vaccine design.

5

This sampling is obviously skewed toward the B and C clade. As will be shown subsequently, the coverage of “potential epitopes” (9-mers) in other clades is still excellent.

10 The sequences

M consensus

>ConS

MRVRGIQRNCQHLWRWGTLILGMLMICSAAENLWVTVYYGVVPWK
 EANTTLFCASDAKAYDTEVHNWVATHACVPTDPNPQEIVLENTENF
 15 NMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLNCTNVNVTNT
 TNNTTEKGGEIKNCSFNITTEIRDKKQKVYALFYRLDVVPIDNNNNSSN
 YRLINCNTSAITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFNGTGPCK
 NVSTVQCTHGIKPVVSTQLLNGSLAEEEEIIRSENITNNAKTIIVQLNES
 VEINCTRPNNNTRKSIRIGPGQAFYATGDIIGDIRQAHCNISGTKWNKT
 20 LQQVAKKLREHFNNKTIIFKPSSGGDLEITTHSFNCRGEFFYCNTSGLF
 NSTWIGNGTKNNTNDTITLPCRIKQIINMWQGVGQAMYAPPIEGKI
 TCKSNITGLLLTRDGGNNNTNETEIFRPGGGDMRDNRSELYKYKV
 KIEPLGVAPTKAKRRVVEREKRAVGIGAVFLGFLGAAGSTMGAASITL
 TVQARQLLSGIVQQQSNLLRAIEAQHLLQLTVWGIKQLQARVLAVE
 25 RYLKDQQLGIWGC SGKLICTTTVPWNSSWSNKSQDEIWDNMTWME
 WEREINNYTDIISLIEESQNQQEKNEQELLALDKWASLWNWFDITNW
 LWYIKIFIMIVGGLIGLRIVFAVLSIVNRVRQGYSPFSFQTLIPNPRGPDR
 PEGIEEGGEQDRDRSIRLVNGFLALAWDDLRLSLCLFSYHRLRDFILIA
 ARTVELLGRKGLRRGWEALKYLWNLLQYWGQELKNSAISLLDTTAIA
 30 VAEGTDRVIEVVQRACRAILNIPRRIRQGLERALL

3 mosaics

>M_mos_3_1

MRVKGIRKNYQHLWRWGTMLLGMLMICSAAEQLWVTVYYGVPVW
5 RDAETTLFCASDAKAYEREVHNWATHACVPTDPNPQEIVLENVTEE
FNMWKNNMVDQMHEDIISLWDESLKPCVKLTPLCVTLNCTDVNVTK
TNSTSWGMMEKGEIKNCSFNMTTEL RDKKQKVYALFYKLDIVPLEEN
DTISNSTYRLINCNTSAITQACPKVTFEPIPIHYCTPAGFAILKCNDKKFN
GTGPCKNVSTVQCTHGIRPVVTTQLLNGSLAEEEEIIRSENLTNNAKTI
10 IVQLNESVVINCTRPNNNTRKSIRIGPGQTFYATGDIIGNIRQAHCNISRE
KWINTTRDVRKKLQEHFNKTIIFNSSSGGDLEITTHSFNCRGEFFYCNTS
KLFNSVWGNSSNVTKVNGTKVKETITLPCIKQIINMWQEVGRAMYA
PPIAGNITCKSNITGLLLVRDGGNVTNNT EIFRPGGGNMKDNWRSELY
KYKVVEIKPLGIAPTKAKRRVVEREKRAVGLGAVFLGFLGAAGSTMG
15 AASMTLTVQARQLLSGIVQQQSNLLRAIEAQQHMLQLTVWGIKQLQA
RILAVEYLRDQQLGIWGC SGKLICTTNVPWNSSWSNKS LDEIWNN
MTWMQWEKEIDNYTSLIYTLIEESQNQQEKNEQDLLALDKWANLWN
WFDISNWLWYIRIFIMIVGGLIGLRIVFAVLSIVNRVRKGYSPLSFQTLT
PNPRGPDRLGRIEEEGGEQDKDRSIRLVNGFLALAWDDLRLNLCLFSYH
20 RLRDLLIVTRIVELLGRRGWEALKYLVNLLQYWIQELKNSAVSLLN
ATAIAVAEGTDRVIEVVQRACRAILHIPRRIRQGLERALL

>M_mos_3_2

MRVKETQMNWPNLWKWGTLLGLVIICSASDNLWVTVYYGVPVWKE
25 ATTLFCASDAKAYDTEVHNWATYACVPTDPNPQEVVLGNVTENF
NMWKNNMVEQMHEDIISLWDQSLKPCVRLTPLCVTLNCSNANTTNT
NSTEEIKNCSFNITTSIRDKVQKEYALFYKLDVVPIDNDNTSYRLISCNT
SVITQACPKVSFEPIPIHYCAPAGFAILKCKDKKFNGTGPCTNVSTVQC
THGIRPVVSTQLLNGSLAEEEVVIRSENFTNNAKTIIVHLNKSVEINCT
30 RPNNNTRKSIHIGPGRAFYATGEIIGDIRQAHCNISRAKWNNTLKQIVK

KLKEQFNKTIIFNQSSGGDPEITTHSFNCGGEFFYCNTSGLFNSTWNST
 ATQESNNTLNGNITLPCRIKQIVNMWQEVGKAMYAPPIRGQIRCSSNI
 TGLILTRDGGNNNSTNETFRPGGGDMRDNRSELYKYKVVKIEPLGV
 APTKAKRRVVQREKRAVGTIGAMFLGFLGAAGSTMGAASLTTLTVQA
 5 RLLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARVLAVERYLK
 DQQLLGIWGC SGKLICTTTVPWNTSWSNKSLEIWDNMTWMEWEREI
 DNYTGLIYTLLEESQNQQEKNEQELLELDKWASLWNWFDITKWLYWI
 KIFIMIVGGLVGLRIVFTVLSIVNRVRQGYSPLSFQTHLPAPRGPDRPEG
 IEEEGGERDRDRSGRLVDGFLAIWVDLRSCLFSYHQLRDFILIAARTV
 10 ELLGHSSLKGLRRGWEALKYWWNLLQYWSQELKNSAISLLNTTAIVV
 AEGTDRIIEVLQRAGRAILHIPTRIRQGLERLLL

>M_mos_3_3

MRVRGIQRNWPQWWIWGILGFWMLMICNVVGNLWVTVYYGVPVW
 15 KEAKTTLFCASDAKAYEKEVHNWATHACVPTDPSPQEVVLENTEN
 FNMWKNDMVDQMHEDEVISLWDQSLKPCVKLTHLCVTLNCTNATNT
 NYNNSTNVTSSMIGEMKNCSFNITTEIRDKSRKEYALFYRLDIVPLNEQ
 NSSEYRLINCNTSTITQACPKVSFDPIPIHYCAPAGYAILKCNNKTFNGT
 GPCNNVSTVQCTHGIKPVVSTQLLLNGSLAEGEIIIRSENLTDNAKTIIV
 20 HLNESVEIVCTRPNNNTRKSVRIGPGQAFYATGDIIGDIRQAHCNLSRT
 QWNNTLKQIVTKLREQFGNKTIVFNQSSGGDPEIVMHSFNCGGEFFYC
 NTTQLFNSTWENSNTITQPLTLNRTKGPNDTITLPCRIKQIINMWQGVGR
 AMYAPPIEGLIKCSSNITGLLLTRDGGNNSETKTTETFRPGGGNMRDN
 WRNELYKYKVQIEPLGVAPTRAKRRVVEREKRAVGIGAVFLGFLGT
 25 AGSTMGAASITLTVQARQVLSGIVQQQSNLLKAIEAQQHLLKLTWGI
 KQLQTRVLAIERYLKDQQLLGLWGC SGKLICTTAVPWNSSWSNKSQT
 DIWDNMTWMQWDREISNYTDTIYRLLED SQNQQEKNEKDLLALDSW
 KNLWNWFDITNWLWYIKIFIIIVGGLIGLRIIFAVLSIVNRCRQGYSPSL
 QTLIPNPRGPDRLGGIEEGGEQDRDRSIRLVSGFLALAWDDLRSCLF

SYHRLRDFILIVARAVELLGRSSLRGLQRGWEALKYLGSLVQYWGLEL
KKSAISLLDTIAIAVAEGTDRIIEVIQRICRAIRNIPRRIRQGFEAALL

Single optimal natural sequence selected from available acute SGA sequences:

5 >B.acute.Con.1059
MRVTEIRKNYLWRWGIMLLGMLMICSAAEQLWVTVYYGVPVWKEA
TTTLFCASDAKAYTAEAHNVWATHACVPTDPNPQEVVLENTENFN
MWKNNMVEQMHEIDIISLWDQSLKPCVKLTPLCVTLNCTDLANNTNL
ANNTNSSISSWEKMEKGEIKNCSFNITTVIKDKIQKNYALFNRLDIVPID
10 DDDTNVTNNASYRLISCNTSVITQACPKISFEPIPIHYCAPAGFAILKCN
DKKFNGTGPCTNVSTVQCTHGIKPVVSTQLLNGSLAEEEVVIRSENFT
DNVKTIIVQLNESVIINCTRPNNNTRKSITFGPGRAFYTGGDIIGDIRKAY
CNISSTQWNNTLRQIARRLREQFKDKTIVFNSSSGGDPEIVMHSFNCGG
EFFYCNTTQLFNSTWNGNDTGEFNNTGKNITYITLPCRIKQIINMWQEV
15 GKAMYAPPIAGQIRCSSNITGILLTRDGGNSSSEDKEIFRPEGGNMRDNW
RSELYKYKVVKIEPLGVAPTKAKRRVVQREKRAVGIGAVFLGFLGAA
GSTMGAASMTLTVQARLLLSGIVQQQNNLLRAIEAQQHLLQLTVWGI
KQLQARVLAVERYLKDQQLGIWGC SGKLICTTAVPWNASWSNRSLD
NIWNNMTWMEWDREINNYTNLIYNLIEESQNQQEKNEQELLELDKW
20 ASLWNWFDITKWLWYIKIFIMIVGGLVGLRIVFVILSIVNRVRQGYSPL
SFQTHLPTPRGLDRHEGTEEEGGERDRDRSGRLVDGFLTIIWIDLRSLC
LFSYHRLRDLLLIVTRIVELLGRRGWEILKYWWNLLQYWSQELKNSA
VSLLNATAIAVAEGTDRIIEIVQRIFRAILHIPTRIRQGLERALL

25 3 optimal natural selected from available acute samples, SGA sequences:

>B.acute.Con.1059
MRVTEIRKNYLWRWGIMLLGMLMICSAAEQLWVTVYYGVPVWKEA
TTTLFCASDAKAYTAEAHNVWATHACVPTDPNPQEVVLENTENFN
MWKNNMVEQMHEIDIISLWDQSLKPCVKLTPLCVTLNCTDLANNTNL
30 ANNTNSSISSWEKMEKGEIKNCSFNITTVIKDKIQKNYALFNRLDIVPID

DDDTNVTNNASYRLISCNTSVITQACPKISFEPIPIHYCAPAGFAILKCN
DKKFNGTGPCTNVSTVQCTHGIKPVVSTQLLLNGSLAEEEVVIRSENFT
DNVKTIIVQLNESVIINCTRPNNNTRKSITFGPGRAFYTGTGDIIGDIRKAY
CNISSTQWNNTLRQIARRLREQFKDKTIVFNSSSGGDPEIVMHSFNCGG
5 EFFYCNTTQLFNSTWNGNDTGEFNNTGKNITYITLPCRICKQIINMWQEV
GKAMYAPPIAGQIRCSSNITGILLTRDGGNSSSEDKEIFRPEGGNMRDNW
RSELYKYKVVKIEPLGVAPTKAKRRVVQREKRAVGIGAVFLGFLGAA
GSTMGAASMTLTVQARLLLSGIVQQQNNLLRAIEAQQHLLQLTVWGI
KQLQARVLAVERYLKDQQLGIWGC SGKLICTTAVPWNASWSNRSLD
10 NIWNNMTWMEWDREINNYTNLIYNLIEESQNQQEKNEQELLELDKW
ASLWNWFDITKWLWYIKIFIMIVGGLVGLRIVFVILSIVNRVRQGYSP
SFQTHLPTPRGLDRHEGTEEEGGERDRDRSGRLVDGFLTIIWIDLRSLC
LFSYHRLRDLIIIIVTRIVELLGRRGWEILKYWWNLLQYWSQELKNSA
VSLLNATAIAVAEGTDRIIEIVQRIFRAILHIPTRIRQGLERALL

15

>C.acute.Con.0393

MRVRGILRNYQQWWIWGILGFWMLMICSVGGNLWVTVYYGVPVWR
EAKTTLFCASDAKAYEREVHNWATHACVPTDPNPQELFLENTENF
NMWKNDMVDQMHEDIISLWDQSLKPCVKLTPLCVTLNCSNANITRNS
20 TDGNTTRNSTATPSDTINGEIKNCSFNITTELKDKKKKEYALFYRLDIV
PLNEENS NFNEYRLINCNTSAVTQACPKVSFDPIPIHYCAPAGYAILKC
NNKTFNGTGPCNNVSTVQCTHGIKPVVSTQLLLNGSLAEEIIIIRSENLT
NNAKTIIVHLKEPVEIVCTRPNNNTRKSMRIGPGQTFYATDIIGDIRQAS
CNIDEKTWNNTLNKVGEKLQEHFPNKTLNFAPSSGGDLEITTHSFNCR
25 GEFFYCNTSKLFYKTEFNSTTNSTITLQCRICKQIINMWQGVGRAMYAPP
IEGNITCKSNITGLLLTRDGGTNDMTETFRPGGGDMRDNRSELYKY
KVVEIKPLGVAPTEAKRRVVEREKRALTLGALFLGFLGTAGSTMGAA
SITLTVQARQLLSGIVQQQSNLLKAIEAQQHLLQLTVWGIKQLQTRVL
AIERYLQDQQLGLWGC SGKLICTTAVPWNSSWSNKSQGEIWGNMT
30 WMQWDREISNYTNTIYRLLED SIIQQEKNEKDLLALDSWKNLWSWFS

ITNWLWYIKIFIMIVGGLIGLRIIFAVLSIVNRVRQGYSPLPFQTLIPNPR
 GPDRLGRIEEEGGEQDRDRSIRLVNGFLAIAWDDLRLSLCLFSYHRLRDF
 ILIAARAAELLGRSSLRGLQRGWEALKYLGSLVQYWGLELKKSAISLL
 DTVAITVAEGTDRIIEVVQRICRAICNIPRRIRQGFEAALQ

5

Coverage comparison of the four vaccine antigens.

Mosaics and naturals are optimized for the first red bar on the left for each vaccine (the total). The “total” represents all sequences, database + CHAVI. The “B” is the subset that are B clade, “C” the subset that are C
 10 clade, and “N” the remaining M group sequences that are not B or C (all other clades and recombinants). As B is most common, the single best natural is of course a B, and B thus has the best coverage for Nat.1. Con S, as expected, provides much more even coverage for all clades, and provides better coverage for all the groups except B clade. (Note: in a Con S Macaque study,
 15 the natural B was not selected to be optimal, and Con S had better coverage *even within* B clade than the B vaccine strain that had been used; this was reflected in the number of detected responses to heterogeneous B’s. A difference here is that the natural B was selected to be the natural B clade sequence from acute infection that provides optimal coverage). Nat.3 gives
 20 good broad coverage, Mos.3 better. (See Fig. 11.)

The mosaics will minimize rare 9-mers but in Env they cannot be excluded or it is not possible to span certain really variable regions to make intact proteins. For all other HIV proteins tested, it was possible to exclude 9-mers that were found at 3 times or less. Still, the 3 best natural Envs contain
 25 more than twice the number of rare 9-mer variants relative to the 3 Env mosaics.

Figure 12 includes additional summaries of coverage; ConS gp160 contains quite a few conserved 9-mers that are missed in gp140DCFI, as one would expect. ConS provides slightly less coverage than a single mosaic, but
 30 it is already known that ConS works very well in macaques so serves as a

good positive control. 1, 2, and 3 mosaics give increasingly better coverage, and Nat.3 is not as good as Mos.3.

Figure 13 is alignment dependent, and based on the database alignment (the tow plots above this are alignment independent). Each position represents the 9-mer it initiates as one moves across the protein. The upper bound (black dashed lined) is the sum of the frequencies of the three most common 9-mers starting from each position; it represents the maximal limit that could be achieved for coverage with 3 proteins, and this is not quite achievable in practice because there can be conflicts in a given position for overlapping 9-mers, although the 3 mosaic combination very nearly achieves it. The reason the “total 9-mers” shown in grey varies is because of insertions and deletions in the alignment.

Only the Mos.3 vaccine cocktail is shown in Fig. 13. However, all four vaccines resorted by coverage is shown in Fig. 14, where those positions that start the 9-mers that are best covered by the vaccine are moved to the left. The exact match line is left in all four plots for a reference point. Not only does Mos.3 (red) approach the maximum, but the orange and yellow near-matches that have potential for cross-reactivity are also improved in this vaccine cocktail as compared to the others.

The plots shown in Fig. 15 map every amino acid in every sequence in the full database alignment. A row of pixels is a sequence, a column is an alignment position. White patches are insertions to maintain the alignment. All 9-mers that encompass an amino acid are considered. If every 9-mer that spans the amino acid has a perfect match in the vaccine cocktail, the pixel is yellow, *so yellow is good*. If one is off, light orange, two off, darker orange... through no spanning 9-mer matches represented by black. Note: lots of yellow for 3 mosaics, relative to the other vaccines. There is a big patch of the most yellow for the B clade in Nat.1 as the single best natural is a B clade. Note, all those dark bits: in these regions the sequences in the database are different than any 9-mer in the vaccine, so cross-reactivity would be several limited.

Optimization using 9-mers.

9-mers were selected because that is the most common size of an optimal CD8+ T cell epitope. They range from 8-12, and optimal CD4+ T cell epitopes can be even be larger or smaller. As it turns out, coverage of 9-mers is best when optimized for 9-mer coverage, but if optimization on a different size yields very little decrease in coverage for 9-mers. The same goes for all lengths, 8-12, the peak coverage is for the size selected but the coverage is excellent for other lengths, as the solutions are related. 9-versus 12-mers are shown in Fig. 16, 12 being the most extreme value one might reasonably consider. The coverage is nearly identical for 9-mers optimized for 9 or 12, or for 12-mers optimized for 9 or 12; it is 1-2% higher for the length selected for optimization. Naturally, 12-mers have fewer identities than 9-mers in general, because they are longer so it is harder to find a perfect match. A more comprehensive study was made of this for HIV proteins showing that the loss was consistently larger for 12-mers when optimized on 9 rather than *vice versa*, and that, in other proteins, this difference could be up to 4-5%. Thus, for Env the selection of 9-mers is less of a problem. Given all of the above, 9-mers were selected since this is the most common optimal CTL epitope length, and since optimal coverage of 9-mers provides approaching optimal coverage of other lengths.

Options for the 3 best natural strains: acute transmission cases, SGA sequences.

Use of all database sequences as a source for natural strains for vaccine cocktails was first explored, and then a comparison was made of that with selecting from a restricted group of just acute SGA sequences, essentially transmitted viruses. Essentially comparable coverage of the full database could be achieved by restricting to acute infection sequences. As these have other obvious advantages, they will be used for the natural sequences.

First, the exploration of coverage using the full database as a source for a natural cocktail. As noted above, the current M group Env one-seq-per-person data set is dominated by B clade infections, closely followed by C clade. Thus, the single best optimal natural selected by the vaccine design program to cover 9-mers in the (database + CHAVI) data set is a B. If one picks from among any sequence in the database, YU-2 comes up as the best single sequence. To get better representation of other clades, the best B was fixed, and then the next best sequence was added to complement YU-2, which is (logically) a C clade sequence, DU467. Those two were then fixed, and the third complement of the antigen was selected. (If the first two are not fixed, and the program is allowed to choose the third, it logically found a B/C recombinant, it has to be forced to select an A. It is believed that forcing the ABC set would improve global coverage, and partly counteract the B & C clade sampling bias among sequences.)

The optimal naturals from the database tend to harken back to older sequences; this is not surprising, as the older sequences tend to be more central in phylogenetic trees, and thus more similar other circulating strains. *For this study, however, it is preferred to use more contemporary Envelope proteins sampled during acute infection and sequenced using SGA, as these sequences accurately reflect the transmitted virus.* Given that constraint, it is still desired to optimize for 9-mer coverage, so that the cocktail of natural sequences is given the best chance for success in the comparison with mosaics. It turns out when this was done there was an extremely minor loss of coverage when comparing the trivalent cocktail selected from among acute SGA sequences to the trivalent antigen selected from the entire database, (in both cases optimizing for coverage the full database). Thus, by restricting the antigen cocktails to transmitted virus, coverage is not compromised. This alternative has several advantages. Most importantly, it enables a determination of the cross-reactive potential of antibodies generated from acute infection viruses used for the natural cocktail relative to consensus or mosaics as a secondary

endpoint of interest, without compromising the primary endpoint focusing on a comparison of T-cell response breadth of coverage. A large set of B (113) and C (40) clade acute samples sequenced from CHAVI study is available, giving a large dataset from which to select an optimum combination. For the selection of the complementary sequence from the A clade, to complete the B and C in the trivalent vaccine. Several acute sequences were available.

Analysis of gp160 was undertaken that included the 8 subtype A gp160s, and also a subregion analysis was done with all 15 in V1-V4, to get an indication of whether or not more sequencing was required. Fortunately, one of the available full length sequences made an excellent complement to the B and C acutes, essentially as good as any of the others. This comparison indicated there was no particular need to do more sequencing at this time. It is believed that this is appropriate since with such a limited A baseline to select from, because the A sequence only needs to complement the choice of B and C clade strains, and many Bs and Cs were available from which to choose. Two of the patients from which the Nat.3 cocktail is derived are below. Nat.1 is just the first one.

B patient 1059

Patient Sex=M

RiskFactor=PPD

Sample country=USA

Sample city=Long Beach, CA

Patient cohort=CA-UCSF

Patient health status=Acute

Viral Load=2,800,000

Infection country=USA

Sample date=03/26/98

C patient 0393

Fiebig Stage=4

Infection country=Malawi

Sample date=17-Jul-2003

5 Viral Load=12,048,485

Patient sex=F

CD4count=618 (measured 13 days after sequenced sample)

Patient age=23

STD=GUD,PID

10

Figs. 17 and 18 illustrate the minimal loss of coverage in selecting from acute SGA sequences, and a highlighter plot of each of the 3 patients env sequences, that shows that the consensus of each patient is equivalent to the most common strains, and thus an excellent estimate of the actual transmitted virus.

15

Why M group and not clade specific coverage?

It is believed that it is important to strive for a global HIV vaccine, if at all possible, with exploratory methods such as these since many nations have multiclade epidemics, and people travel. While intra-clade coverage can definitely be gained by a within-clade optimized vaccine, the result of such a strategy would be dramatic loss of inter-clade coverage. The hope is that a multivalent mosaic could provide enough breadth to counter viruses of virtually any clades or recombinants. The compromise and benefit in terms of coverage for Env M group versus subtype-specific design is shown in Fig. 19.

25

Why Env?

This proof of concept study is well positioned to see differences in breadth of responses using Env as the test antigen. This is partly because of the theoretical considerations described herein (ENV has twice many conserved 9

30

mers in the mosaics relative to the best natural strain, and only half as many rare variants) and partly because of the prior animal studies. Env studies with a consensus versus natural in macaques showed a highly significant increase in breadth of responses: 3-4 fold more epitopes per Env protein were
5 recognized (Santra et al, in press, PNAS). Env mosaics have shown an even more profound advantage in a mouse study (up to 10-fold over comparable numbers of natural antigens, manuscript in preparation in collaboration with the VRC). Based on this prior work, it makes sense to start with a small human trial testing the breadth of responses to Env. Ultimately, the hope is to
10 apply the proof of concept gained with Env to a more conserved protein like a Gag where it may be possible to confer broadest protection. Gag gives outstanding coverage of the full M group. Tests of Gag and Nef are ongoing in macaque, using a 4 mosaic vaccine cocktail approach (see Example 3). A coverage comparison of macaque 4 mosaic Gag vaccine and proposed human
15 Env 3 mosaic vaccine against the current database is in Fig. 20. There is more theoretical potential for cross-reactivity with the Gag vaccine, but more progress has been made with Env in the animal models to date, so Env has the best foundation to justify moving forward. The three mosaic Env sequences described above and the sequences used in Example 3 are shown in Fig. 21.

20

DNA

The DNAs to be used will be in the form of the full gp160 Env. The gp160 would be in the PCMVR plasmid (Gary Nabel) and will be the identical plasmid used in all VRC DNA immunization trials. Dose is anticipated to be 4
25 mg. The following DNA constructs will be used:

- DNA optimal Wildtype Env transmitted/founder env (WT Env)
- DNA group M consensus Env (ConS Env)
- DNA Trivalent optimal wildtype transmitted/founder Env (WT Tri Env)
- 30 • DNA Trivalent Mosaic Env

NYVAC

NYVAC (vP866) is a recombinant poxvirus vector which has an 18 gene deletion versus wild-type virus. The NYVAC vector will be licensed from
5 Sanofi-Pasteur and manufactured by a third party contractor and will be propagated on a CEF cell substrate. The Env construct expressed in NYVAC will be gp140C (entire Env with transmembrane and cytoplasmic domain deleted and gp41/gp120 cleavage site mutated) or will be a full gp160. The choice of construct design will depend on the ability to make the NYVAC
10 with gp160 forms vs gp140. The dose of NYVAC is anticipated to be $\sim 1 \times 10^7$ TCID₅₀. The following NYVAC constructs will be used:

- NYVAC WT Env
- NYVAC ConS Env
- NYVAC Trivalent Native Env
- 15 • NYVAC Trivalent Mosaic Env

Vaccinations will be given by intramuscular injection.

Table: Protocol Schema

Group	Number	Dose	Injection schedule in weeks			
			0	4	20	24
1	20		DNA WT	DNA WT	NYVAC WT	NYVAC WT
			Env	Env	EnvA	EnvA
	4		Placebo	Placebo	Placebo	Placebo
2	20		DNA ConS	DNA ConS		
			Env	Env	NYVAC ConS	NYVAC ConS
	4		Placebo	Placebo	Placebo	Placebo
3	20		DNA	DNA	NYVAC	NYVAC
			Trivalent	Trivalent	Trivalent	Trivalent
	4		Native Env	Native Env	Native Env	Native Env
4	20		Placebo	Placebo	Placebo	Placebo
			DNA	DNA	NYVAC	NYVAC
	4		Trivalent	Trivalent	Trivalent	Trivalent
Total	96 (80/16)		Mosaic Env	Mosaic Env	Mosaic Env	Mosaic Env
			Placebo	Placebo	Placebo	Placebo

Participants:

Healthy, HIV-1–uninfected volunteers aged 18 to 50 years:

- 5 80 vaccinees
 16 control recipients
 96 total participants

Design:

- 10 Randomized, placebo-controlled, double-blind trial

Duration per participant:

Approximately 12 months

- 15 **Estimated total study duration:**

Approximately 18 months

EXAMPLE 3

- 20 ***Construction of the plasmid DNA vaccines and recombinant vaccinia***

- (*rVV*). Mosaic *gag* and *nef* genes, group M consensus *gag* and *nef* genes were generated by converting amino acid sequences of said Gag and Nef, group M consensus Gag and Nef CON-S to nucleotide sequences using a strategy for optimal gene expression. For use as a DNA vaccine, mosaic *gag* and *nef* genes, group M consensus *gag* and *nef* genes were subcloned into WLV0001-
- 25

AM DNA vaccine vector. Endotoxin-free plasmid DNA preparation were produced by Puresyn, Inc. (Malvern, PA) for the immunization of rhesus monkeys. For boosting recombinant vaccinia viruses expressing the individual mosaic *gag* and *nef* genes, group M consensus *gag* and *nef* genes were generated. The methods used were as previously described (Liao et al, 5 Virology 353:268-282 (2006); Earl, BioTechniques 23:1094-1097 (1997)).

Experimental groups and vaccination schedule. Three groups of rhesus monkeys were immunized with either 10mg of the empty DNA vector plasmid (group 1, 6 monkeys), or 5mg each of group M *gag* and *nef* plasmid 10 DNA (group 2, 12 monkeys) or 1.25mg each of 4 mosaic *gag* and 4 *nef* plasmid DNA (group 3, 12 monkeys) intramuscularly at Day 0 and Day 30. The monkeys will be boosted with the corresponding rVV expressing the initial immunizing immunogen (10^9 pfu/monkey) 5 month post-15 immunization with the 2nd DNA immunization.

Myristoylation of Gag and Nef has a potential down regulation effect on immune responses and thus the myristoylation of Gag and Nef has been mutated in the sequences used in this study.

20

* * *

All documents and other information sources cited above are hereby incorporated in their entirety by reference.

WHAT IS CLAIMED IS:

1. A polypeptide or protein comprising at least one sequence of amino acids set forth in Fig. 21 or Fig. 22.
2. A nucleic acid encoding the polypeptide or protein according to claim 1.
3. A nucleic acid comprising at least one sequence of nucleotides set forth in Fig. 22.
4. A vector comprising the nucleic acid according to claim 2 or 3.
5. The vector according to claim 3 wherein said vector is a viral vector.
6. A composition comprising at least one polypeptide or protein according to claim 1 and a carrier.
7. A composition comprising at least one nucleic acid according to claim 2 or 3 and a carrier.
8. A method of inducing an immune response in a mammal comprising administering to said mammal an amount of at least one polypeptide or protein according to claim 1 sufficient to effect said induction.
9. A method of inducing an immune response in a mammal comprising administering to said mammal an amount of at least one nucleic acid according to claim 2 or 3 sufficient to effect said induction.

Fig. 1A

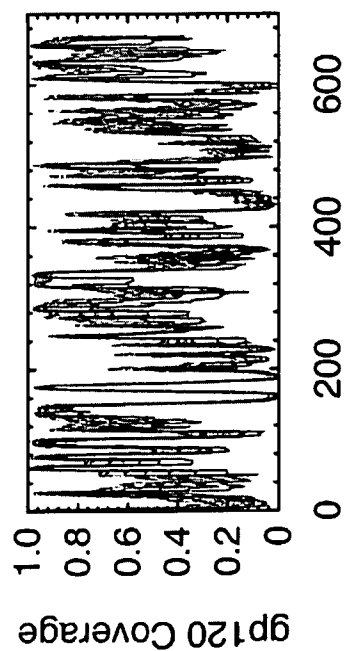


Fig. 1B

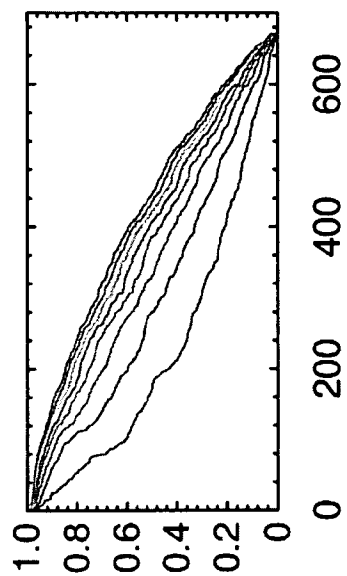


Fig. 1C

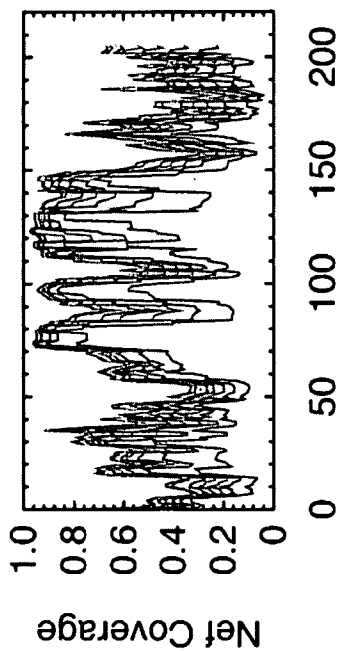


Fig. 1D

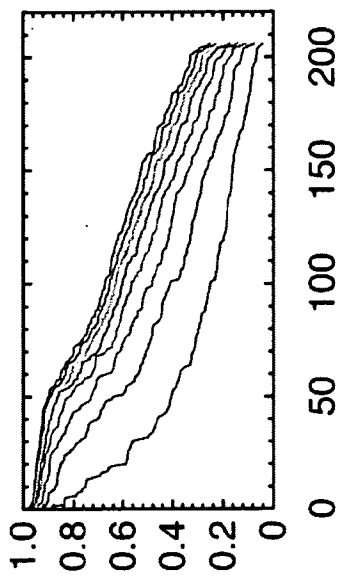


Fig. 1E

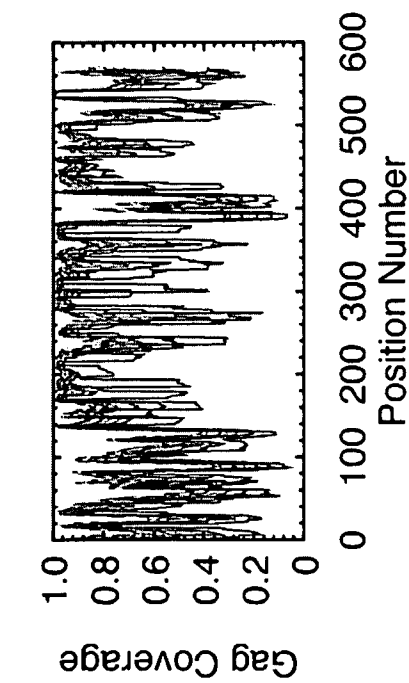


Fig. 1F

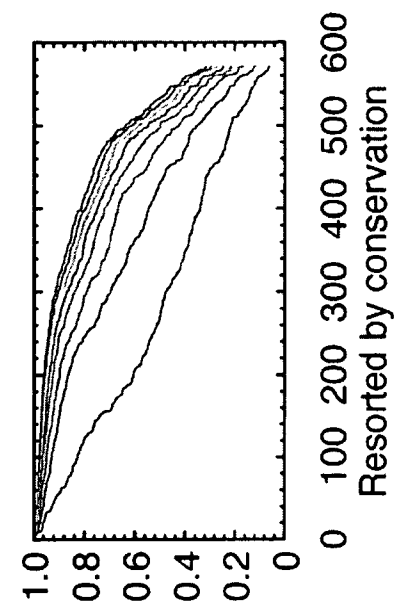


Fig. 2A

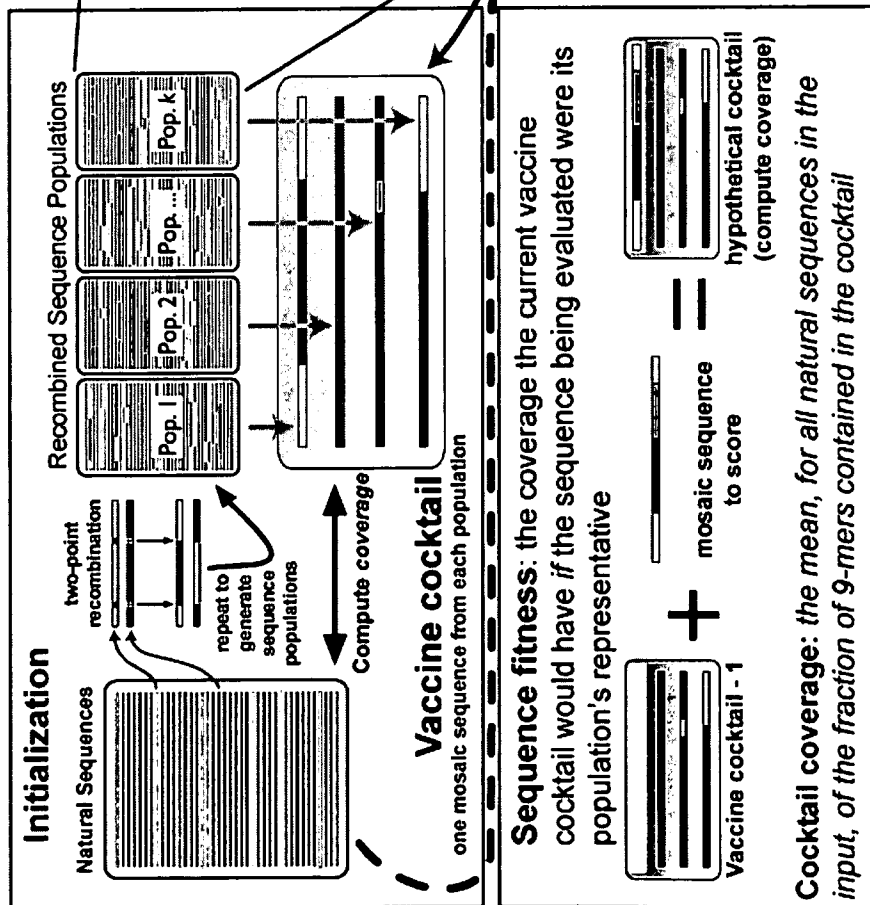


Fig. 2B

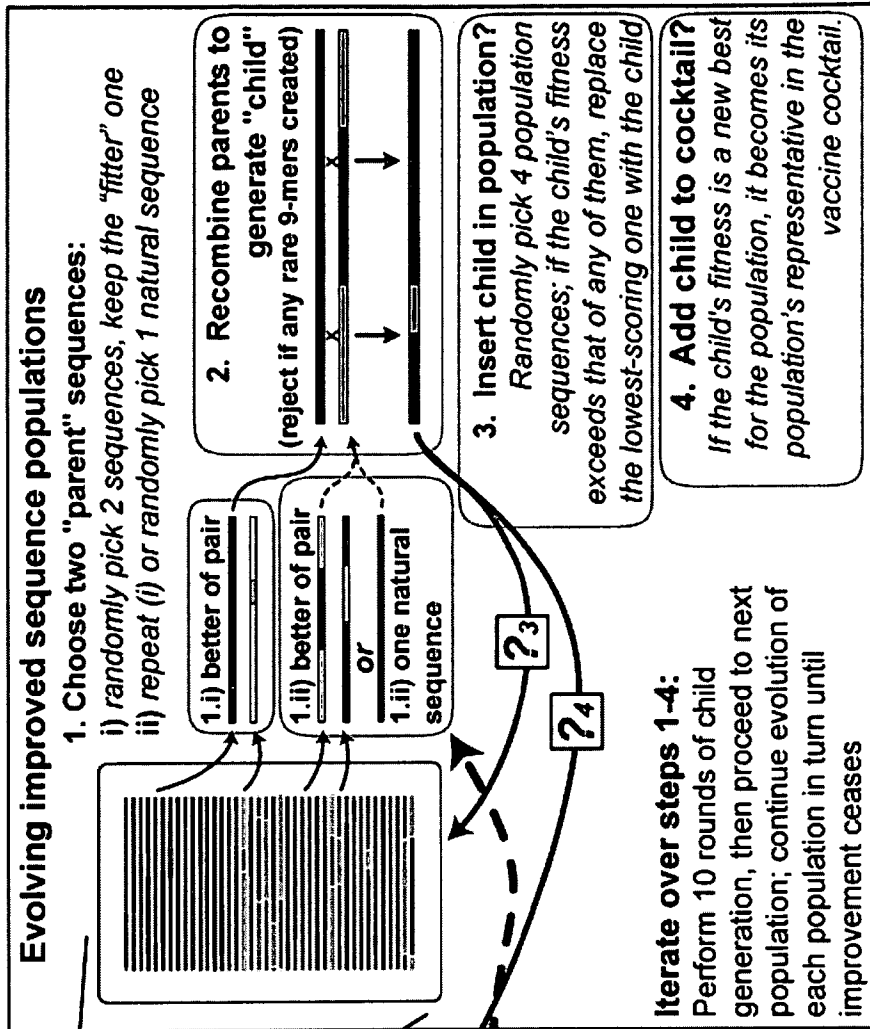
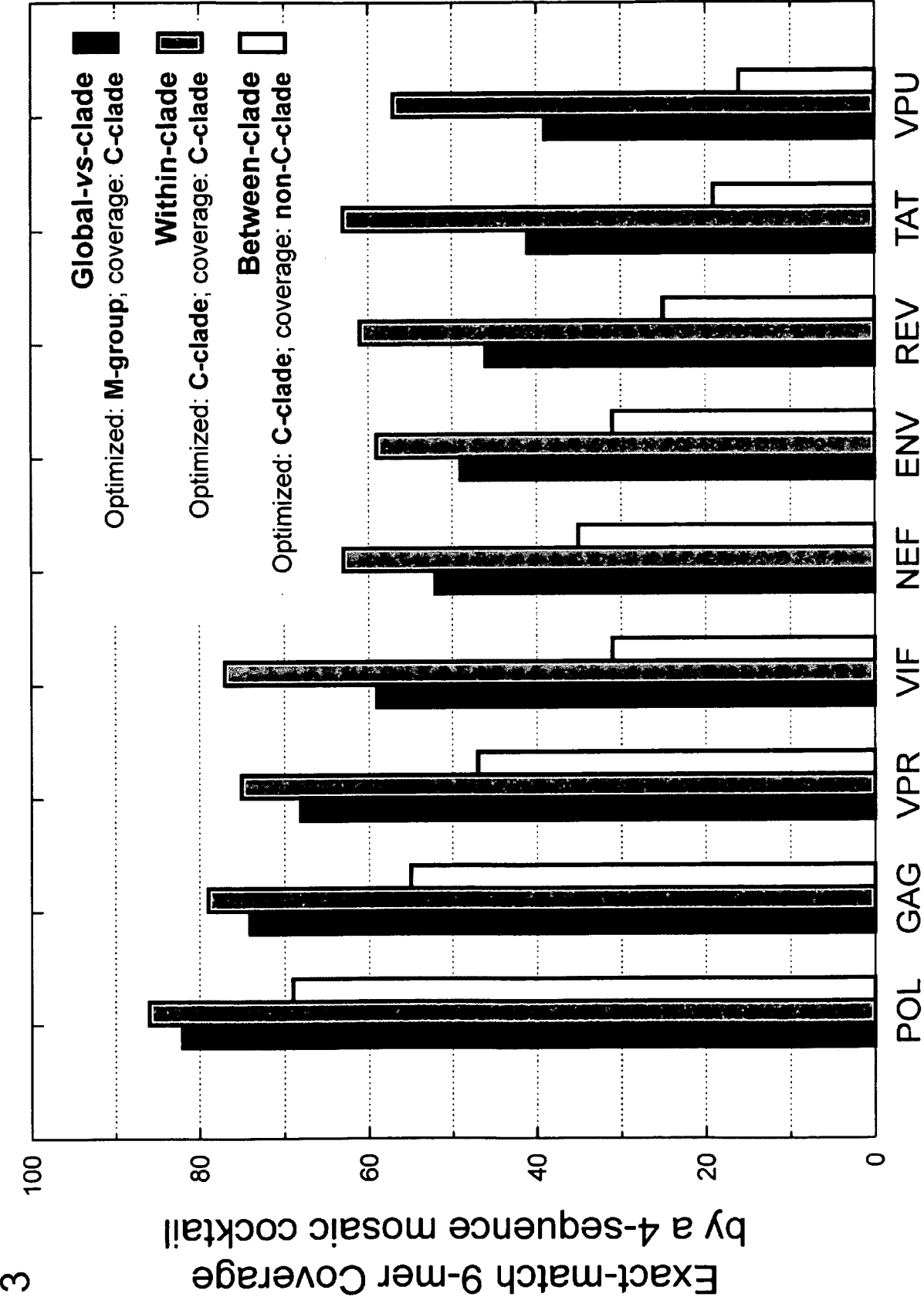


Fig. 2C



4/109

Fig. 4A

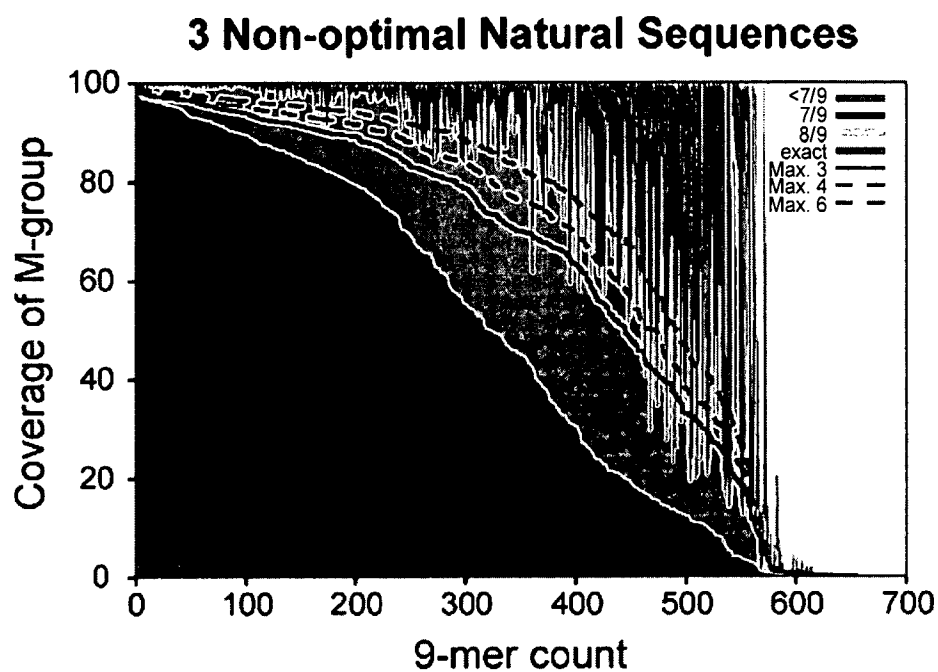
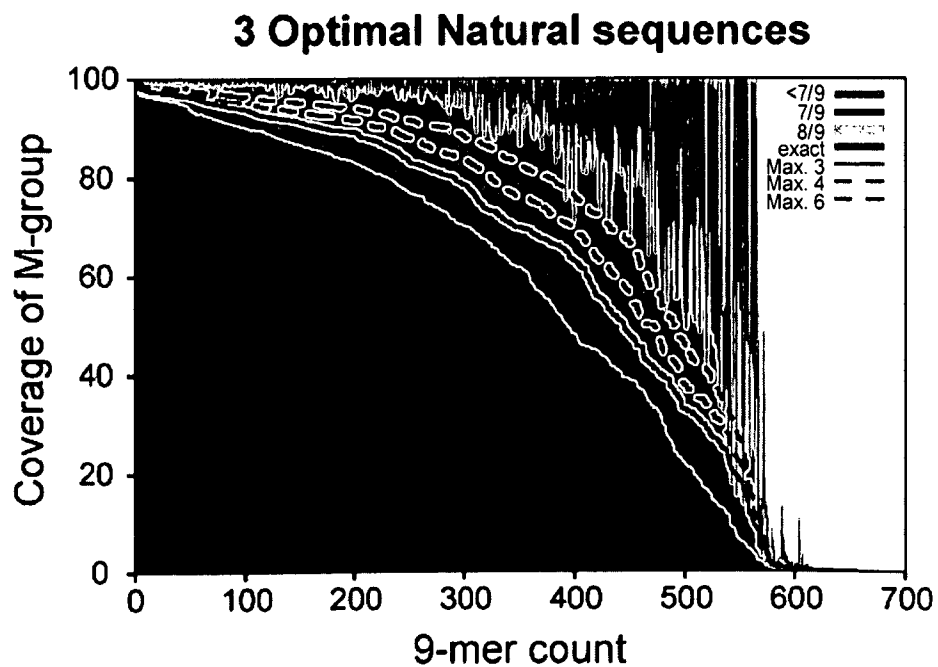


Fig. 4B



5/109

Fig. 4C

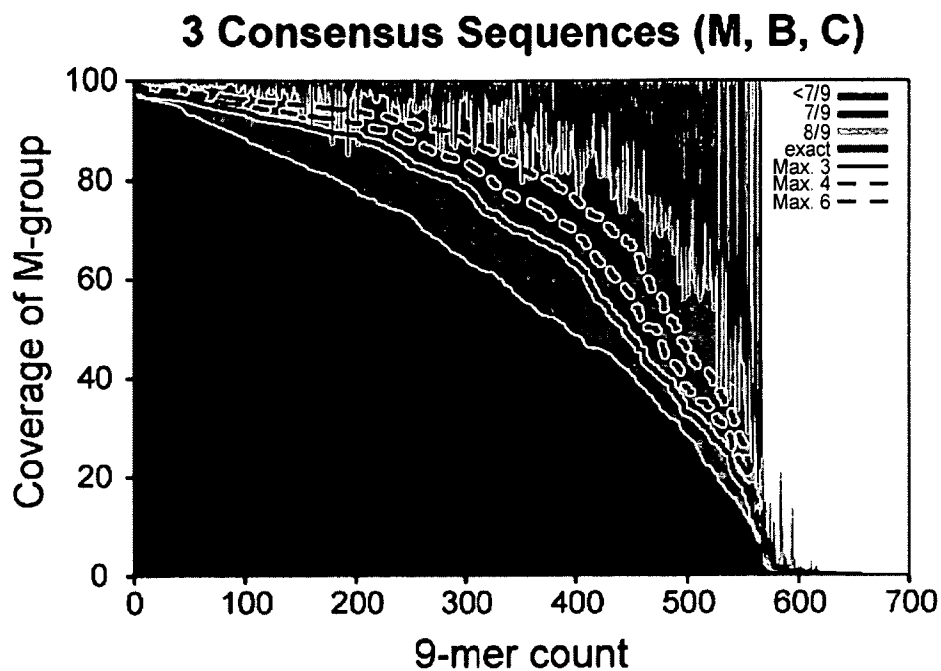
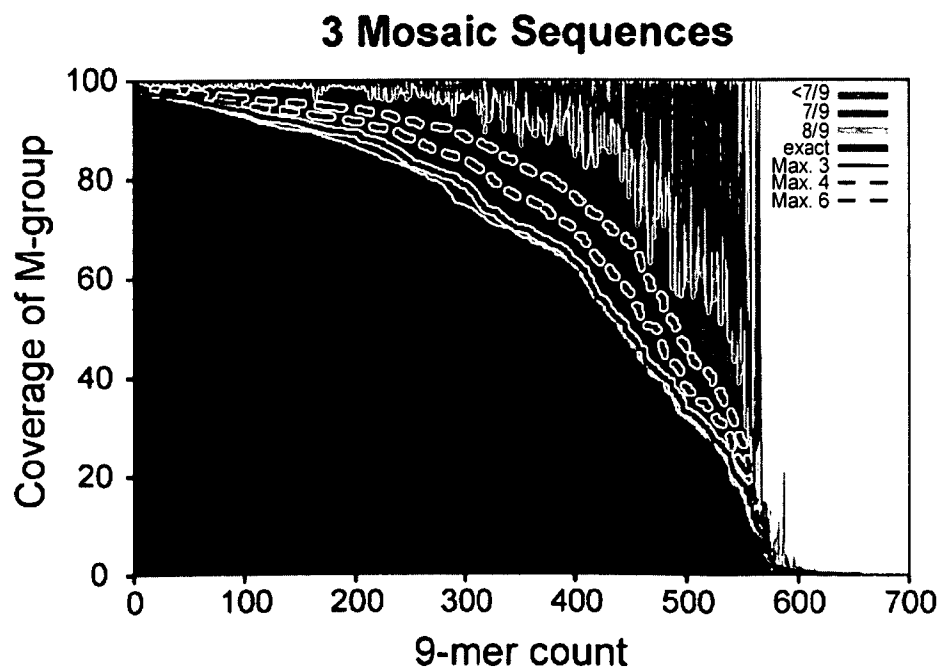


Fig. 4D



6/109

Fig. 4E

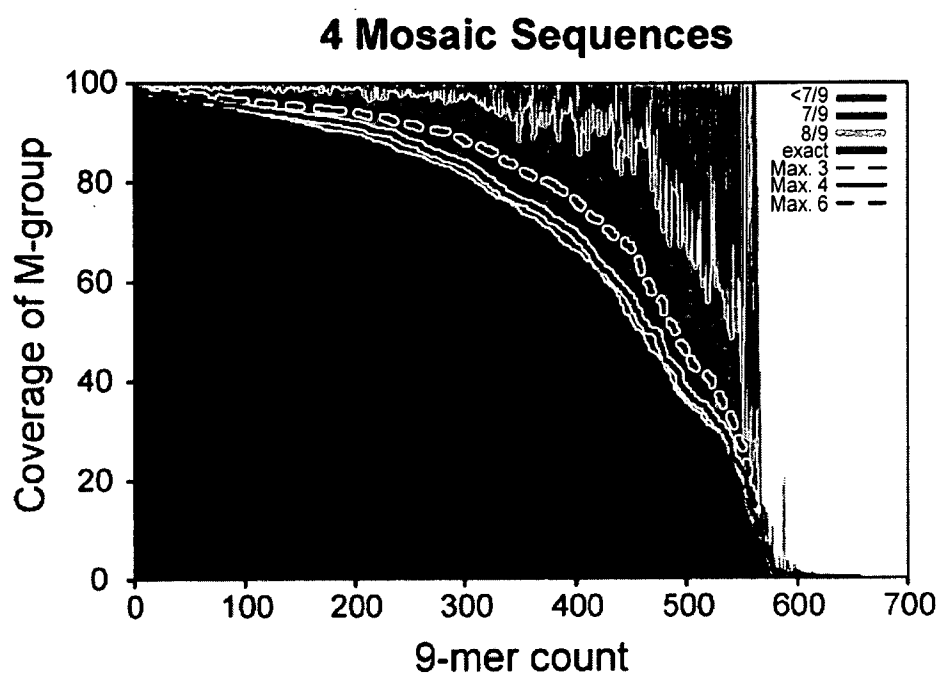
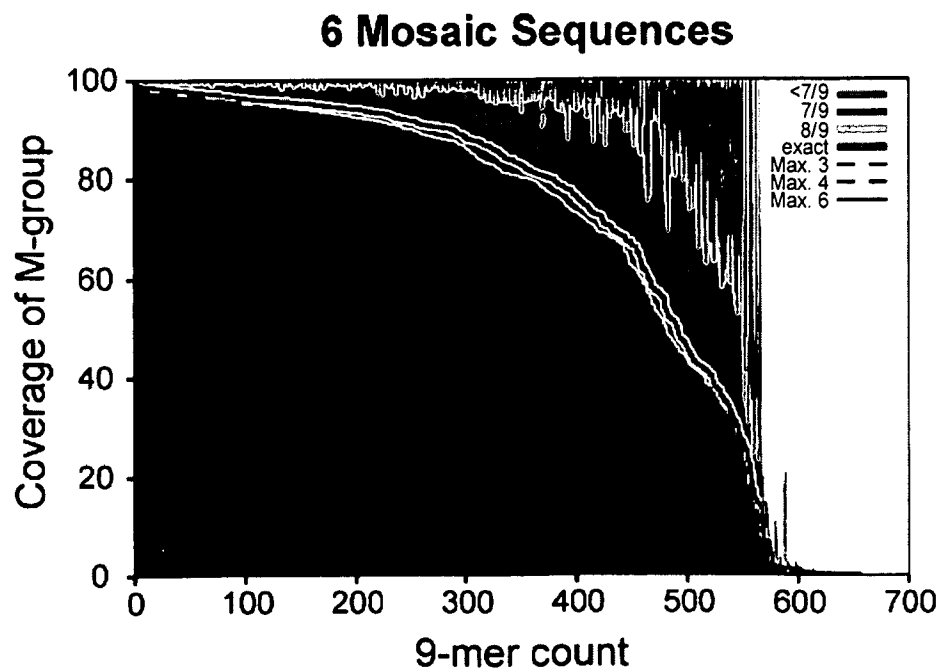
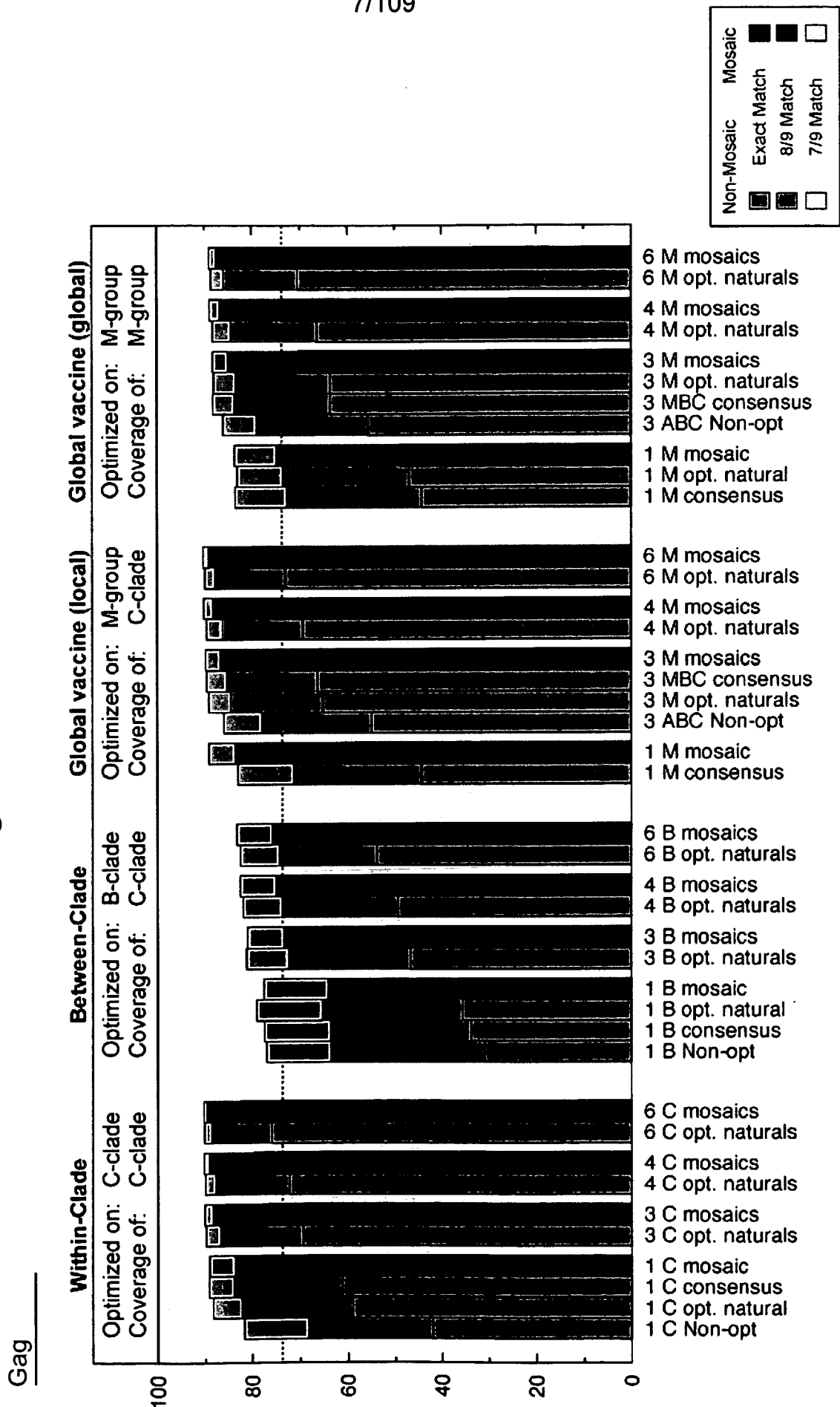


Fig. 4F



7/109

Fig. 5A



8/109

Fig. 5B

Nef

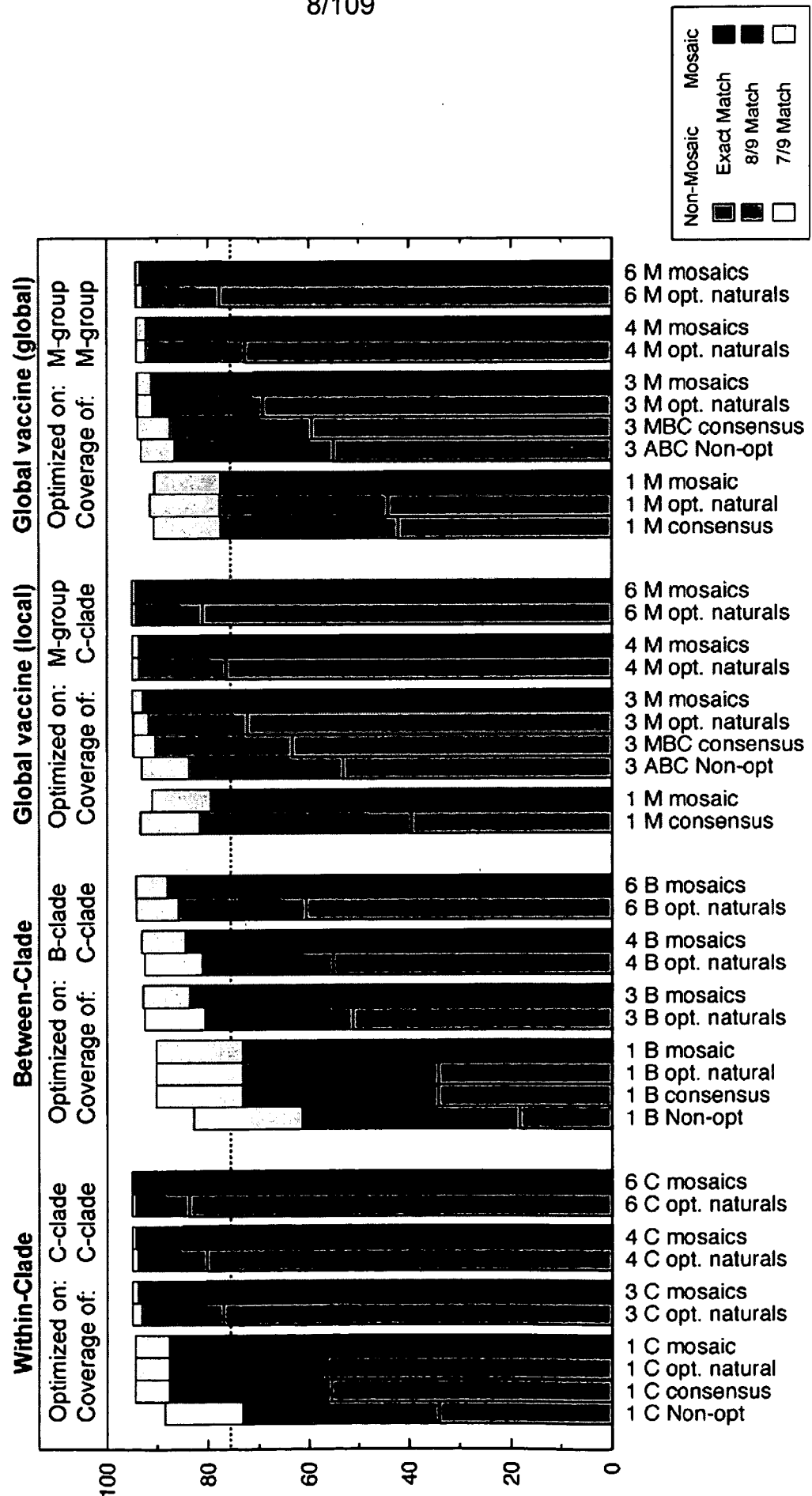
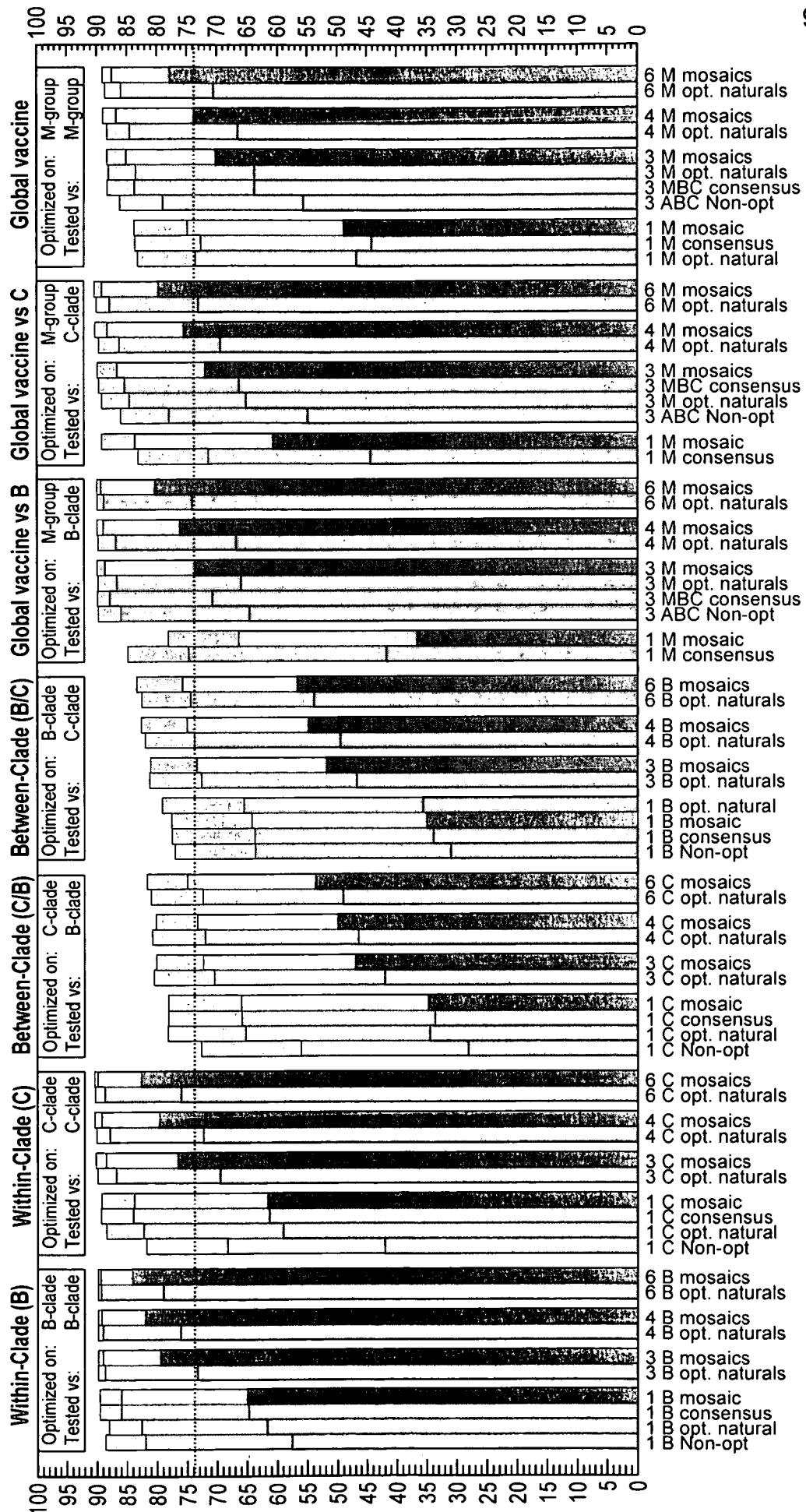


Fig. 6A

Gag coverage



10/10

Global vaccine

Optimized on: M-group
Tested vs: M-group

6 M mosaics
6 M opt. naturals
4 M mosaics
4 M opt. naturals
3 M mosaics
3 M opt. naturals
3 MBC consensus
3 ABC Non-opt
1 M mosaic
1 M consensus
1 M opt. natural

Global vaccine vs C

Optimized on: M-group
Tested vs: C-clade

6 M mosaics
6 M opt. naturals
4 M mosaics
4 M opt. naturals
3 M mosaics
3 M opt. naturals
3 MBC consensus
3 ABC Non-opt
1 M mosaic
1 M consensus

Global vaccine vs B

Optimized on: M-group
Tested vs: B-clade

6 M mosaics
6 M opt. naturals
4 M mosaics
4 M opt. naturals
3 M mosaics
3 M opt. naturals
3 MBC consensus
3 ABC Non-opt
1 M mosaic
1 M consensus

Between-clade (B/C)

Optimized on: B-clade
Tested vs: C-clade

6 B mosaics
6 B opt. naturals
4 B mosaics
4 B opt. naturals
3 B mosaics
3 B opt. naturals
1 B opt. natural
1 B mosaic
1 B consensus
1 B Non-opt

Between-clade (C/B)

Optimized on: C-clade
Tested vs: B-clade

6 C mosaics
6 C opt. naturals
4 C mosaics
4 C opt. naturals
3 C mosaics
3 C opt. naturals
1 C mosaic
1 C consensus
1 C opt. natural
1 C Non-opt

Within-clade (C)

Optimized on: C-clade
Tested vs: C-clade

6 C mosaics
6 C opt. naturals
4 C mosaics
4 C opt. naturals
3 C mosaics
3 C opt. naturals
1 C opt. natural
1 C mosaic
1 C consensus
1 C Non-opt

Within-clade (B)

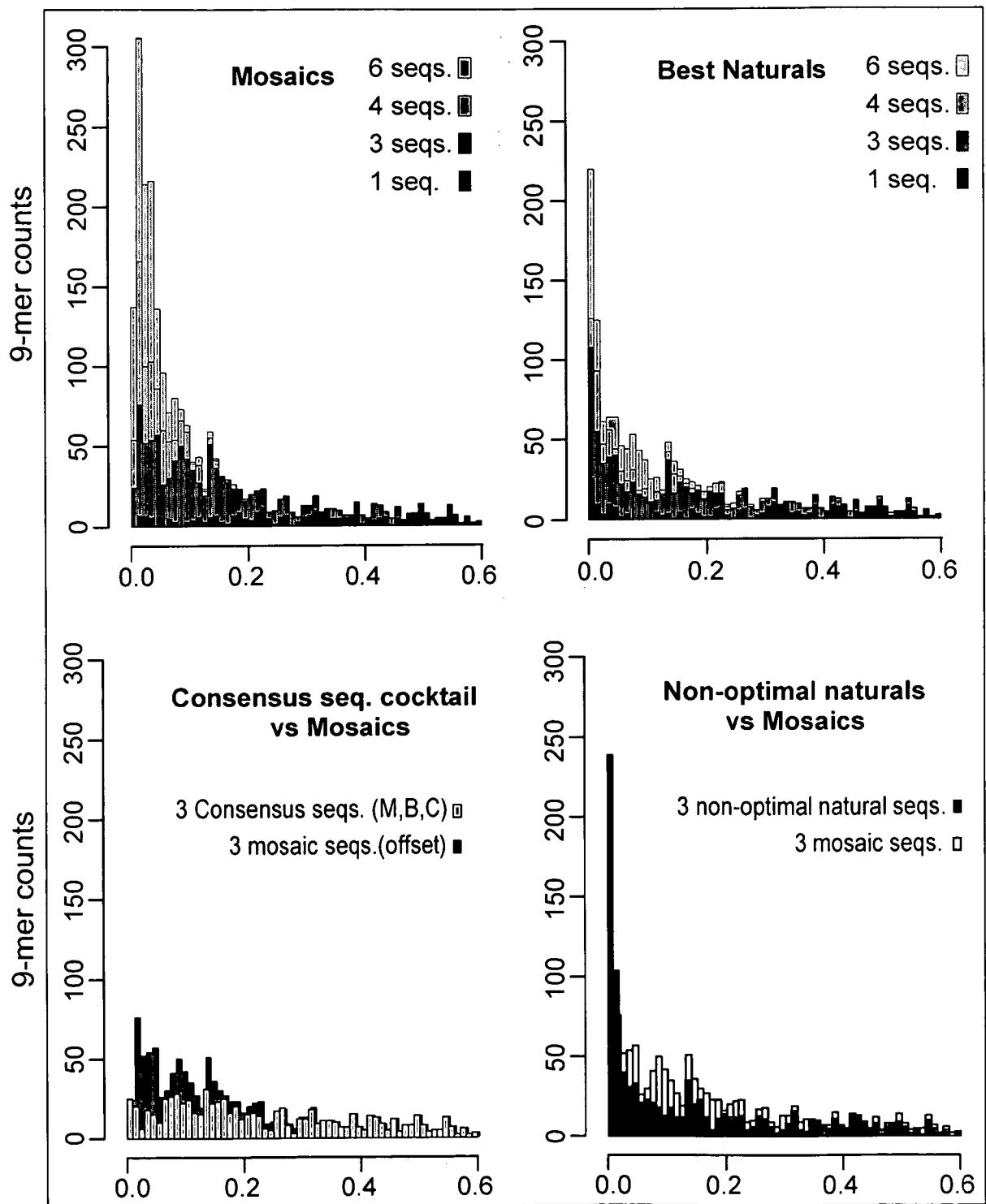
Optimized on: B-clade
Tested vs: B-clade

6 B mosaics
6 B opt. naturals
4 B mosaics
4 B opt. naturals
3 B mosaics
3 B opt. naturals
1 B mosaic
1 B consensus
1 B opt. natural
1 B Non-opt

11/109

Fig. 7A

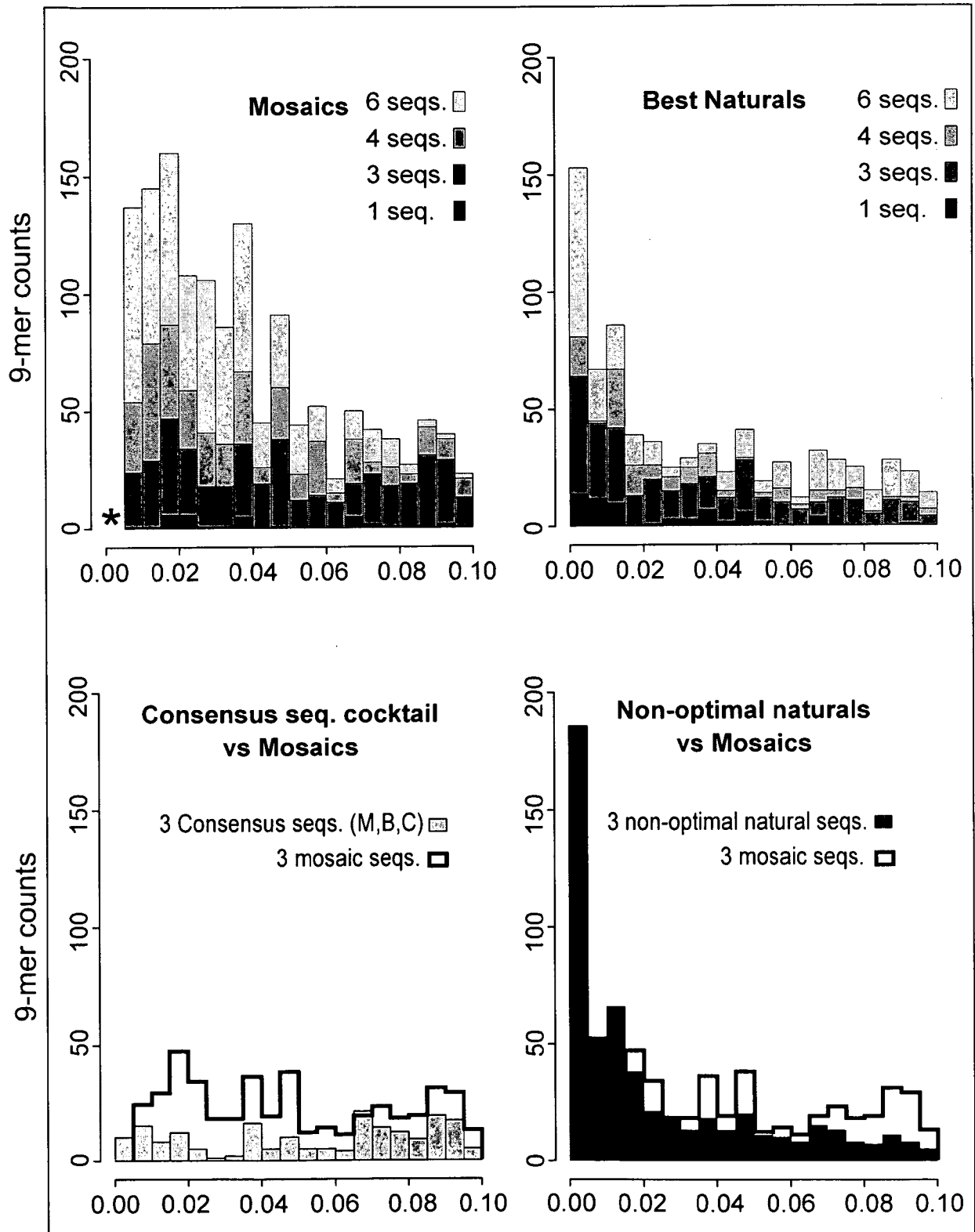
9-mer Frequencies (0–60%)

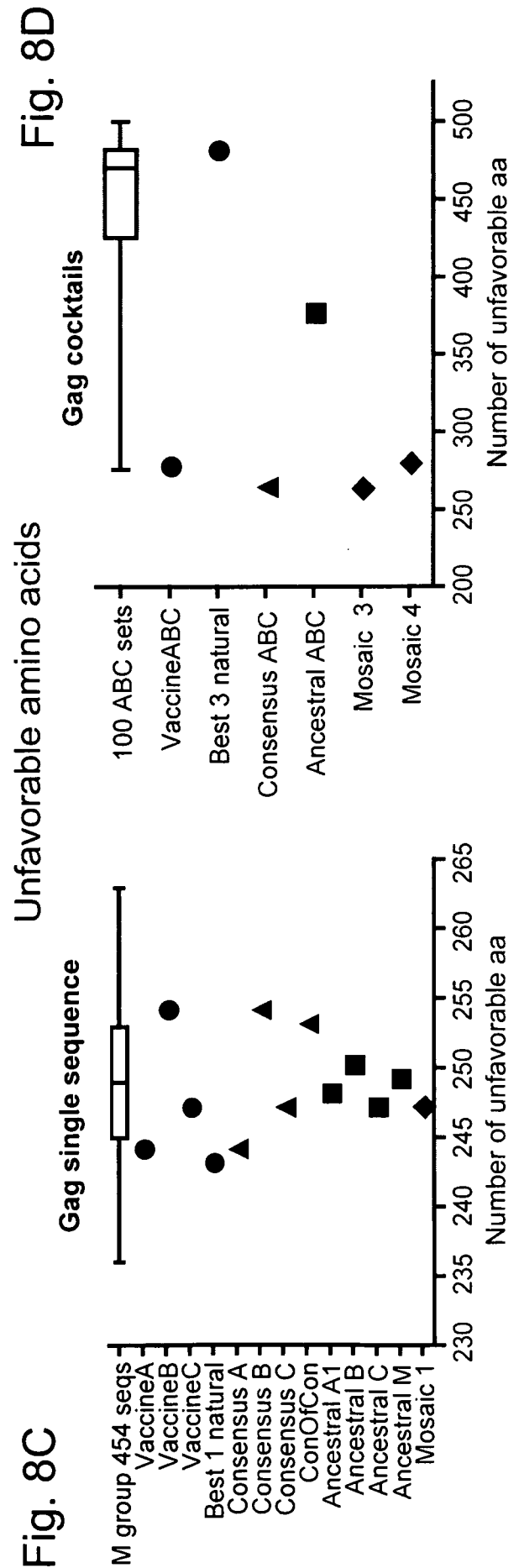
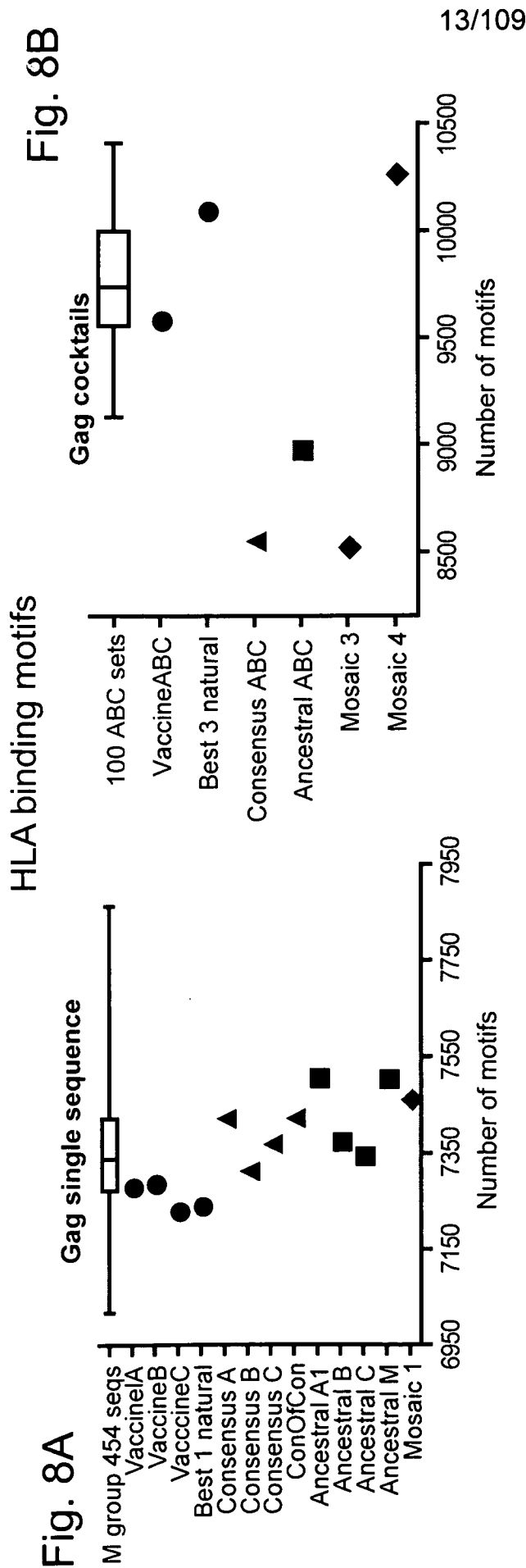


12/109

Fig. 7B

9-mer Frequencies (0–10%)





14/109

Fig. 9

```
>nef_coreB.syn1.1
EVGFVPVPQVPLRPMYKALDLSHFLKEKGGLEGLIYSQKRQDILDLWVYHTQGYFPDW
QNYTPGPGIRYPLTFGWCFKLVPVE

>nef_coreB.syn3.1
EVGFVPVPQVPLRPMYKAAVDLSHFLKEKGGLEGLVYSQKRQDILDLWIYHTQGYFPDW
QNYTPGPGTRFPLTFGWCFKLVPVD
>nef_coreB.syn3.2
EVGFVPVPQVPLRPMYKALDLSHFLREKGGLEGLIYSQKRQEILDLWVYHTQGYFPDW
HNYTPGPGVRYPLTFGWCFKLVPVE
>nef_coreB.syn3.3
EVGFVPVPQVPLRPMYKAAVDLSHFLKEKGGLEGLIHSQRRQDILDLWVYHTQGFFPDW
QNYTPGPGIRYPLTFGWCYKLVPVE

>nef_coreB.syn4.1
EVGFVPVPQVPLRPMYKAAVDLSHFLREKGGLEGLIHSQKRQDILDLWIYHTQGYFPDW
QNYTPGPGIRYPLTFGWCYKLVPVE
>nef_coreB.syn4.2
DVGFVPVPQIPLRPMYKAAVDLSHFLREKGGLEGLVYSQKRQDILDLWVYHTQGFFPDW
QNYTPGPGTRFPLTFGWCFKLVPVD
>nef_coreB.syn4.3
EVGFVPVPQVPLRPMYKAVDLSHFLKEKGGLEGLIYSQKRQEILDLWVYHTQGYFPDW
GNYTPGPGTRYPLTFGWCLKLVPVE
>nef_coreB.syn4.4
EVGFVPVPQVPLRPMYKALDLSHFLKEKGGLDGLIYSQRRQDILDLWVYNTQGYFPDW
HNYTPGPGVRYPLTFGWCFKLVPVE

>nef_coreB.syn6.1
EVGFPIRPQVPLRPMTFKALDLSHFLKEKGGLDGLIYSQKRQEILDLWVYHTQGYFPDW
HNYTPGPGIRYPLCFGWCFKLVPVE
>nef_coreB.syn6.2
EVGFVPVPQVPLRPMYKAVDLSHFLKEQGGLEGLIYSQRRQDILDLWIYNTQGYFPDW
QCYTPGPGVRFPLTFGWCFKLEPVD
>nef_coreB.syn6.3
EVGFVPVPQVPLRPMYKAAVDLSHFLKEKGGLEGLIYSQKRQDILDLWIYHTQGYFPDW
QNYTPGPGIRYPLTFGWCLKLVPVE
>nef_coreB.syn6.4
DVGFVPVPQVPLRPMYKAAVDLSHFLKEEGGLEGLIYSQRRQDILDLWVYHTQGFFPDW
QNYTPGPGTRYPLTFGWCYKLVPVE
>nef_coreB.syn6.5
EVGFVPVPQIPLRPMYKALDLSHFLREKGGLEGLVYSQKRQDILDLWVHHTQGYFPDW
GNYTPGPGTRFPLTFGWCFKLVPVD
>nef_coreB.syn6.6
EVGFVPVPQVPLRPMYKGAFDLSHFLKDKGGLEGLIHSQKRQDILDLWVYNTQGYFPDW
QNYTPGPGVRYPLTFGWCFKLVPLE

>nef_coreC.syn1.1
EVGFVPVPQVPLRPMYKAAFDLSFFLKEKGGLEGLIYSKKRQEILDLWVYHTQGFFPDW
QNYTPGPGVRYPLTFGWCFKLVPVD
```

15/109

Fig. 9 cont'd-1

>nef_coreC.syn3.1
EVGFVPVTPQVPLRPMTFKGAFDLGFFLKEKGGLDGLIYSKKRQDILDLWVYHTQGYFPDW
QNYTPGPGVRYPLTFGWICYKLVVD
>nef_coreC.syn3.2
EVGFVPKQVPLRPMTYKAAFDLSFFLKDKGGLEGLIWSKKRQEILDLWVYHTQGFFPDW
QNYTPGPGIRYPLTFGWPFKLVVD
>nef_coreC.syn3.3
EVGFVVRPQVPLRPMTYKGAFDLSFFLKEKGGLEGLIYSKKRQEILDLWVYNTQGFFPDW
HNYTPGPGVRFPLTFGWCFKLVVD

>nef_coreC.syn4.1
EVGFVVRPQVPLRPMTYKGAFDLSFFLKEKGGLEGLIWSKKRQEILDLWVYNTQGYFPDW
QCYTPGPGVRFPLTFGWCFKLVVD
>nef_coreC.syn4.2
DVGFVVRPQVPLRPMTYKAAFDLSFFLKDKGGLEGLIHSKKRQDILDLWVYNTQGFFPDW
HNYTPGPGIRYPLTFGWCFKLVVD
>nef_coreC.syn4.3
EVGFVPKQVPLRPMTYKAAVDLSFFLKEKGGLEGLIYSKKRQEILDLWVYHTQGFFPDW
QNYTPGPGTRYPLTFGWPFKLVVD
>nef_coreC.syn4.4
EVGFVPVTPQVPLRPMTFKGAFDLGFFLKEKGGLDGLIYSKKRQDILDLWVYHTQGYFPDW
QNYTPGPGVRYPLTFGWICYKLVVD

>nef_coreC.syn6.1
DVGFVVRPQVPLRPMTYKGAFDLSFFLKEKGGLDGLIYSKKRQEILDLWVYNTQGYFPDW
QCYTPGPGVRYPLTFGWICYKLVVD
>nef_coreC.syn6.2
EVGFVPKQVPLRPMTYKAAFDLSFFLKDKGGLEGLIWSKKRQEILDLWVYHTQGFFPDW
QNYTPGPGIRYPLTFGWCFKLVVD
>nef_coreC.syn6.3
EVGFVPKQVPLRPMTFKGAFDLGFFLKEKGGLEGLIYSKQRQDILDLWVYHTQGFFPDW
HNYTPGPGVRLPLTFGWCFKLVVD
>nef_coreC.syn6.4
GVGFVVRPQVPLRPMTYKAAFDLSFFLKDKGGLEGLIYSKKRQDILDLWVYNTQGFFPDW
QNYTPGPGVRFPLTFGWCFKLVVD
>nef_coreC.syn6.5
EVGFVPVTPQVPLRPMTYKAAVDLSWFLKEKGGLDGLIYSRKRQEILDLWVHHTQGFFPDW
QNYTPGPGTRFPLTFGWCFKLVVD
>nef_coreC.syn6.6
EVGFVVRPQVPLRPMTYKAVDLSFFLKEKGGLEGLIHSKKRQDILDLWVYHTQGYFPDW
QNYTPGPGTRYPLTFGWPFKLVVD

>nef_coreM.syn1.1
EVGFVVRPQVPLRPMTYKALDLSHFLKEKGGLEGLIYSKKRQEILDLWVYHTQGYFPDW
QNYTPGPGVRYPLTFGWCFKLVVD

16/109

Fig. 9 cont'd-2

```
>nef_coreM.syn3.1
DVGFPVRPQVPLRPMTYKAAVDLSHFLKEKGGLEGLVYSQKRQDILDLWVYHTQGFFPDW
QNYTPGPGVRYPLTFGW CYKLPVD
>nef_coreM.syn3.2
EVGFPVRPQVPRPMTYKGAFDLSFFLKEKGGLEGLIYSKKRQEILDLWVYHTQGYFPDW
HNYTPGPGTRFPLTFGWCFKLPVD
>nef_coreM.syn3.3
EVGFPVKPQVPLRPMTYKALDLSHFLKEKGGDGLIYSKKRQDILDLWVYNTQGYFPDW
QNYTPGPGIRYPLTFGWCFKLPVE

>nef_coreM.syn4.1
EVGFPVTPQVPLRPMTFKGAFDLGFFLKEKGGLEGLIYSKKRQEILDLWVYHTQGFFPDW
QNYTPGPGTRYPLCFGWCFKLPVE
>nef_coreM.syn4.2
EVGFPVKPQVPLRPMTYKAAVDLSHFLKEKGGLEGLVYSQKRQDILDLWVYHTQGYFPDW
QNYTPGPGIRYPLTFGW CYKLPVD
>nef_coreM.syn4.3
DVGFPVRPQVPLRPMTYKALDLSHFLKEEGGLEGLIYSQKRQEILDLWVYNTQGYFPDW
QNYTPGPGVRYPLTFGWCFKLPVD
>nef_coreM.syn4.4
EVGFPVRPQVPRPMTYKGAFDLSFFLKEKGGDGLIYSKKRQDILDLWVYNTQGFFPDW
HNYTPGPGTRFPLTFGWCFELVPVD

>nef_coreM.syn6.1
EVGFPVRPQVPTRPMTYKGAVDLSHFLKEKGGLEGLVYSQKRQDILDLWVHHTQGFFPDW
QNYTPGPGTRYPLTFGWPFKLPVD
>nef_coreM.syn6.2
DVGFPVRPQVPRPMTYKAAFDLSFFLREKGGDGLIYSKKRQDILDLWVYNTQGYFPDW
QNYTPGPGVRFPLTFGWCFELVPVD
>nef_coreM.syn6.3
NVGFPVRPQVPLRPMTFKGAFDLGFFLKEKGGLEGLIYSKKRQEILDLWVYHTQGYFPDW
HNYTPGPGTRFPLTFGWCFKLPVE
>nef_coreM.syn6.4
EVGFPVTPQVPLRPMTYKGAFDLSFFLKEKGGDGLIYSRKRQEILDLWVYNTQGFFPDW
QNYTPGPGIRYPLTFGWCFKLPMD
>nef_coreM.syn6.5
EVGFPVKPQVPLRPMTYKAAVDLSHFLREKGGLEGLIHSQRRQDILDLWIYHTQGYFPDW
QCYTPGPGVRYPLTFGW CYKLPVD
>nef_coreM.syn6.6
GVGFPVRPQIPLRPMTYKALDLSHFLKEEGGLEGLIYSQKRQDILDLWVYHTQGFFPDW
HNYTPGPGIRYPLCFGWCFKLPVD
```

17/109

Fig. 9 cont'd-3

>gagB.syn1.1

MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQI
 LGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEVKDTKEALDKIEEEQNKSKKKAQQAAA
 DTGNSSQVSQNYPIVQNLQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIMFSALESEGAT
 PQDLNMTMLNTVGGHQAAMQMLKETINEEAAEWDRLHPVHAGPIAPGQMREPRGSDIAGTT
 STLQEQIGWMTNNPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRF
 YKTLRAEQASQEVKNWMTETLLVQANANPDCKTILKALGPAATLEEMMTACQGVGGPGHKA
 RVLAEAMSQVTNSATIMMQRGNFRNQRKTVKCFNCGKEGHIKNCRAPRKKGCWKCGKEG
 HQMKDC-TERQANFLGKIWPSHKG-RPGNFLQSRP-----EPTAP
 PEESFRFGEETTTTPSQKQEPIDKELYPLASLRSFLGNDPSSQ

>gagB.syn3.1

MGARASILSGGELDRWEKIRLRPGGKKQYRLKHIVWASRELERFAINPGLLETSDGCRQI
 LGQLQPALQTGSEELKSLFNTVATLYCVHQRIDVKDTKEALDKIEEEQNKSKKKAQQAAA
 DTGNSSQVSQNYPIVQNIQGQMVHQAISPRTLNAWVKVIEEKAFSPEVIMFSALESEGAT
 PQDLNMTMLNTVGGHQAAMQMLKETINEEAAEDWDRLHPVHAGPIAPGQMREPRGSDIAGTT
 STLQEQIGWMTNNPPIPVGEIYKRWIIMGLNKIVRMYSPTSILDIKQGPKEPFRDYVDRF
 YKVLRAEQASQEVKNWMTETLLVQNSNPDCCTILKALGPGATLEEMMTACQGVGGPGHKA
 RVLAEAMSQMTNSATIMMQKGNFRNQRKPVKCFNCGKEGHIKNCRAPRKKGCWKCGREG
 HQMKDC-TERQANFLGKIWPSYKG-RPGNFLQNRP-----EPTAP
 PAESFRFGEETTTTPSQKQETIDKELYPLASLRSFLGSDPSSQ

>gagB.syn3.2

MGARASVLSGGQLDRWEKIRLRPGGKKKYRLKHLVWASRELERFAVNPGLLETAEGCRQI
 LEQLQPSLQTGSEELRSLFNTVAVLYCVHQRIEVKDTKEALDKVEEEQNKSKKKAQQTAA
 DTGNNSQVSQNYPIVQNMQGQMVHQAISPRTLNAWVKVIEEKAFSPEVIMFTALSEGAT
 PQDLNMTMLNTVGGHQAAMQMLKDTINEEAAEWDRLHPVQAGPVAPGQIREPRGSDIAGST
 STLQEQIGWMTSNPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRF
 YKTLRAEQATQEVKNWMTETLLVQANANPDCKTILKALGPGATLEEMMTACQGVGGPSHKA
 RVLAEAMSQVTNSATIMMQRGNFRNQRKIVKCFNCGKEGHIARNCRAPRKRGCWKCGKEG
 HQMKDC-NERQANFLGKIWPSHKG-RPGNFLQSRP-----EPSAP
 PEESFRFGEETTTTPSQKQEPIDKDLPLASLSLFGNDPSLQ

>gagB.syn3.3

MGARASVLSGGELDKWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQI
 LGQLQPSLQTGSEELRSLYNTVATLYCVHQKIEVKDTKEALEKIEEEQNCKKKKAQQAAA
 GTGNSSQVSQNYPIVQNLQGQMVHQPISPRTLNAWVKVVEEKAFSPEVIMFSALAEAGAT
 PQDLNMTMLNTVGGHQAAMQMLKETINEEAAEWDRLHPVHAGPVAPGQMRDPRGSDIAGTT
 SNLQEQIGWMTNHPPIPVGDIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRF
 YKTLRAEQASQDVKNWMTETLLVQANANPDCRTILKALGPAATLEEMMTACQGVGGPSHKA
 RILAEAMSQVTNPATIMMQRGNFRNQRKTVKCFNCGKEGHLARNCRAPRKKGCWKCGKEG
 HQMKEC-TERQANFLGKIWPSHKG-RPGNFLQSRP-----EPTAP
 PEESFRFGEETATPPQKQEPIDKELYPLTSLRSFLGNDPSSQ

>gagB.syn4.1

MGARASVLSGGQLDRWEKIRLRPGGKKKYRLKHLVWASRELERFAVNPGLLETAEGCRQI
 LEQLQPSLQTGSEELRSLFNTVATLYCVHQRIDVKDTKEALDKIEEEQNCKKKKAQQAAA
 DTGNSSQVSQNYPIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIMFTALSEGAT
 PQDLNMTMLNTVGGHQAAMQMLKETINEEAAEWDRLHPVQAGPVAPGQMREPRGSDIAGST
 STLQEQIGWMTNNPPIPVGDIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRF
 YKVLRAEQASQEVKNWMTETLLVQANANPDCRTILKALGPAATLEEMMTACQGVGGPGHKA

18/109

Fig. 9 cont'd-4

RVLAEAMSQMTNSATIMMQRGNFRNQKRTVKCFNCGKEGHIKNCRAPRKKGCWKCGKEG
 HQMKEC-TERQANFLGKIWPSYKG-RPGNFLQSRP-----EPSAP
 PAESFRFGEETTTTPSQKQETIDKELYPLTSLRSLFGNDPSSLQ

>gagB.syn4.2

MGARASVLSGGELDKWEKIRLRPGGKKKYKLKHIVWASRELERFAINPGLLETSEGCRQI
 LGQLQPALQTGSEELRSLYNTVATLYCVHQKIEVKDTKEALEKVEEEQNKSKQKAQQAAA
 DTGNNSQVSQNYPIVQNLQGQMVHQAI SPRTLNAWVKVIEEKA FSPEVIMFAALSEGAT
 PQDLNMTLNTVGGHQAAMQMLKETINEEAADWDRLHPVHAGPIAPGQIREPRGSDIAGTT
 SNLQEQIAWMTNPNPIPVGEIYKRWIIMGLNKIVRMYSPTSILDIKQGPKEPFRDYVDRF
 YRTLRAEQASQDVKNWMTETLLVQANANPD CRTILKALGPGATLEEMMTACQGVGGPSHKA
 RVLAEAMSQVTNPATIMMQRGNFRNQKRPVKCFNCGKEGHLAKNCRAPRKRGCWKCGKEG
 HQMKDC-NERQANFLGKIWPSHKG-RPGNFLQNRP-----EPSAP
 PEESFRFGEETATPSQKQEPIDKELYPLASLRSLFGSDPSSQ

>gagB.syn4.3

MGARASILSGGELDRWEKIRLRPGGKKQYRLKHIVWASRELERFAVNPGLLETSEGCKQI
 LEQLQPALQTGSEELKSLYNTVAVLYCVHQRIEIKDTKEALEKIEEEQNKSKKKAQQQTAA
 DTGNNSQVSQNYPIVQNLQGQMVHQPI SPRTLNAWVKVIEEKA FSPEVIMFSALAEAT
 PQDLNMTLNTVGGHQAAMQMLKETINEEAAEWDRLHPVHAGPVAPGQMRDPRGSDIAGTT
 SNLQEQIGWMTSNPNPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRF
 YKTLRAEQATQEVKNWMTETLLVQANANPDCKTILKALGPGATLEEMMTACQGVGGPSHKA
 RILAEAMSQVTNSATVMMQRGNFRNQKRTIKCFNCGKEGHLARNCRAPRKKGCWKCGREG
 HQMKDC-TERQANFLGKIWPSYKG-RPGNFLQNRP-----EPTAP
 PAESFRFGEETTTTPQKQEPIDKELYPLASLSLFGNDPSSQ

>gagB.syn4.4

MGARASVLSGGKLDKWEKIRLRPGGKKKYQLKHIVWASRELERFALNPGLLETSDGCRQI
 LGQLQPSLQTGSEELKSLFNTVAVLYCVHQRIEIKDTKEALDKVEEEQNKSKKKAQQQTAA
 GTGNSSQVSQNYPIVQNMQGMVHQAL SPRTLNAWVKVIEEKA FSPEVIMFSALAEAT
 PQDLNMTLNTVGGHQAAMQMLKDTINEEAAEWDRVHPVHAGPIAPGQMRDPRGSDIAGTT
 STLQEQIGWMTNPNPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRF
 YKTLRAEQASQEVKNWMTETLLVQNSNPDC KTILKALGPAATLEEMMTACQGVGGPSHKA
 RVLAEAMSQVTNSATIMMQGNFRNQKRTIKCFNCGKEGHIARNCRAPRKKGCWRCGKEG
 HQMKDC-TERQVNF LGKIWPSYKG-RPGNFLQSRP-----EPTAP
 PEESFRFGEETTTTPQKQEPIDKELYPLASLSLFGNDPLSQ

>gagB.syn6.1

MGARASILSGGELDRWEKIRLRPGGSKKYRLKHIVWASRELERFAVNPGLLETAEGCRQI
 LGQLQPSLQTGSEELRSLYNTIATLYCVHQRIEIKDTKEALEKIEEEQNKSKKKAQQQTAA
 DTGNNSQVSQNYPIVQNLQGQMVHQPI SPRTLNAWVKVVEEKA FSPEVIMFSALAEAT
 PQDLNMTLNTVGGHQAAMQMLKETINEEAAEWDRLHPVHAGPVAPGQIREPRGSDIAGTT
 SNLQEQIGWMTSNPNPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRF
 YRTLRAEQASQDVKNWMTETLLVQANANPDCKTILKALGPGATLEEMMTACQGVGGPGHKA
 RVLAEAMSQVTNSATVMMQRGNFRNQKRTVKCFNCGKEGHIARNCRAPRKKGCWKCGQEG
 HQMKDC-TERQANFLGKIWPSYKG-RPGNFLQSRP-----EPTAP
 PAESFRFGEETTTTPQKQEPIDKELYPLTSLKSLFGNDPSSQ

>gagB.syn6.2

MGARASVLSGGKLDKWEKIRLRPGGKKKYRLKHVVWASRELERFAVNPGLLESSEGCRQI
 LEQLQPSLQTGSEELRSLFNTVATLYCVHQRIDVKDTKEALDKIEEEQNKCKKKAQQQTAA
 DTGNNSQVSQNYPIVQNLQGQMVHQPI SPRTLNAWVKVVEEKA FSPEVIMFAALSEGAT
 PQDLNMTLNTVGGHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQMRDPRGSDIAGST
 STLQEQIGWMTNPNPIPVGDIYKRWIILGLNKIVRMYSPTASILDIRQGPKEPFRDYVDRF
 YKTLRAEQATQEVKNWMTETLLVQANANPD CRTILKALGPGATLEEMMTACQGVGGPGHKA
 RVLAEAMSQMTNSATIMMQGNFRNQKRTIKCFNCGKEGHIARNCKAPRKKGCWKCGREG
 HQMKDC-IERQANFLGKIWPSHKG-RPGNFLQNRP-----EPTAP
 PEESFRFGEETATPPQKQEPIDKELYPLASLSLFGSDPSSQ

19/109

Fig. 9 cont'd-5

>gagB.syn6.3

MGARASVLSGGELDRWEKIRLRPGGKKKYRLKHLVWASRELERFAVNPGLLETSDGCRQI
 LGQLQPALQTGSEELKSLYNTVATLYCVHQKIDVRDTKEALDKIEEEQNKSQKQAQAAAA
 DTGNSSQVSQNYPIVQNIQGQMVHQAI SPRTLNAWVKVVEEKAFSPEVIMFTALSEGAT
 PQDLNMTLNTVGGHQAAMQMLKETINEEAAEWDRLHPVQAGPVA PGQMREPRGSDIAGTT
 STLQEIQIGWMTHNPPIPVGEIYKRWIIMGLNKIVRMYS PSSILDIRQGPKEFRDYVDRF
 YKVLRAEQASQEVKNWMTETLLVQANANPDCKTILKALGPGATLEEMMTACQGVGGPSHKA
 RVLAEAMSQVTNSATIMMQRGNFRNQRKTVKCFNCGKEGHI AKNCRAPRKRGCWKCGKEG
 HQMKEC-TERQANFLGKIWPSHKG-RPGNFLQSRP-----EPTAP
 PEESFRFGEEKTPSQKQETIDKELYPLASLRSLFGNDPSSQ

>gagB.syn6.4

MGARASVLSGGELDKWEKIRLRPGGKKKYQLKHIVWASRELERFAVNPGLLETSEGCKQI
 LEQLQPALQTGSEELRSLYNTIAVLYCVHQKIEIKDTKEALDKVEEEQNKS KKAQAAAA
 GTGNSSQVSQNYPIVQNLQGQMVHQPLSPRTLNAWVKVVEEKAFSPEVIMFTALSEGAT
 PQDLNMTLNTVGGHQAAMQMLKETINEEAAEWDRLHPVHAGPIAPGQMREPRGSDIAGST
 STLQEIQIGWMTHNPPIPVGDIYKRWIILGLNKIVRMYSPTSILDIKQGPKEFRDYVDRF
 YKTLRAEQATQEVKNWMTETLLVQANANPDCKTILKALGPAATLEEMMTACQGVGGPSHKA
 RILAEAMSQVTNPATIMMQRGNFRNQRKPVKCFNCGKEGHI ARNCRAPRKRGCWKCGKEG
 HQMKDC-TERQVNFLGKIWPSHKG-RPGNFLQNRP-----EPSAP
 PAESFRFGREETTPSQKQEPIDKEMYPLASLRSLFGSDPSSQ

>gagB.syn6.5

MGARASVLSGGQLDRWEKIRLRPGGKKQYRLKHIVWASRELERFAINPGLLETSEGCRQI
 LGQLQPSLQTGSEELKSLFNTVAVLYCVHQRIEVKDTKEALEKVEEEQNKS KKAQAAAA
 DTGNSSQVSQNYPIVQNLQGQMVHQALSPRTLNAWVKVVEEKAFSPEVIMFSALSEGAT
 PQDLNMTLNTVGGHQAAMQMLKDTINEEAAEWDRLHPVHAGPVA PGQMRDPRGSDIAGTT
 STLQEIQIAWMTNPNPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEFRDYVDRF
 YKVLRAEQASQDVKNWMTETLLVQANANPDCKTILKALGPAATLEEMMTACQGVGGPGHKA
 RVLAEAMSQVTNSTTIMMQRGNFRNQRKIVKCFNCGKEGHLAKNCRAPRKKGCWKCGKEG
 HQMKEC-TERQANFLGKIWPSHKG-RPGNFLQSRP-----EPSAP
 PEESFRFGREETTPSQKQEPIDKELYPLASLSLFGNDPLSQ

>gagB.syn6.6

MGARASVLSGGKLDKWEKIRLRPGGKKKYKLKHIVWASRELERFALNPGLLETSEGCRQI
 LRQLQPSLQTGSEELRSLYNTVATLYCVHQKIEVKDTKEALEKIEEEQNKS KKAQQTAA
 DTGNSSQVSQNYPIVQNMQGQMVHQALSPRTLNAWVKVVEEKAFSPEVIMFSALSEGAT
 PQDLNMTLNTVGGHQAAMQMLKDTINEEAAEWDRLHPVHAGPIAPGQIREPRGSDIAGTT
 SNLQEIQIAWMTNPNPPIPVGDIYKRWIILGLNKIVRMYS PVSILDIRQGPKEFRDYVDRF
 YKTLRAEQASQEVKNWMTETLLVQNSNPDKTILKALGPGATLEEMMTACQGVGGPGHKA
 RVLAEAMSQVTNPATIMMQGNFKNQRKTVKCFNCGKEGHLARNCRAPRKKGCWRCGKEG
 HQMKDC-NERQANFLGKIWPSHKG-RPGNFLQSRP-----EPTAP
 PEESFRFGREETTPAQKQEPIDKELYPLTSLRSLFGNDPSLQ

>gagC.syn1.1

MGARASILRGGKLDKWEKIRLRPGGKKHYMLKHLVWASRELERFALNPGLLETSEGCKQI
 IKQLQPALQTGTEELRSLYNTVATLYCVHEKIEVRDTKEALDKIEEEQNKS QKQTQQA
 ADG---KVSQNYPIVQNLQGQMVHQAI SPRTLNAWVKVVEEKAFSPEVIMFTALSEGAT
 PQDLNMTLNTVGGHQAAMQMLKDTINEEAAEWDRLHPVHAGPIAPGQMREPRGSDIAGTT
 STLQEIQIAWMTSNPPIPVGDIYKRWIILGLNKIVRMYS PVSILDIKQGPKEFRDYVDRF
 FKTLRAEQATQDVKNWMTDTLLVQANANPDCKTILRALGPGATLEEMMTACQGVGGPSHKA
 RVLAEAMSQ-ANNTNIMMQR SNFKGSKRIVKCFNCGKEGHI ARNCRAPRKKGCWKCGKEG
 HQMKDC-TERQANFLGKIWPSHKG-RPGNFLQSRPE-----PTAPPA-----EPTAP
 PAESFRFEE--TTPAPKQEPKDRE--PLTSLKSLFGSDPLSQ

>gagC.syn3.1

MGARASVLRGEKLDKWERIRLRPGGKKHYMLKHLVWASRELERFALNPGLLETSEGCKQI
 IQQLQPALKTGTEELRSLFNTVATLYCVHAGIEVRDTKEALDKIEEEQNKC QKQTQQA
 ADG---KVSQNYPIVQNLQGQMVHQALSPRTLNAWVKVVEEKAFSPEVIMFTALSEGAT

20/109

Fig. 9 cont'd-6

PSDLNTMLNTVGGHQAAMQMLKDTINEEAAEWDRLHPVQAGPVAPGQMREPRGSDIAGTT
 SNLQEQIAWMTGNPPVPVGDYIKRWIILGLNKIVRMYSVPSILDIRQGPKEPFRDYVDRF
 FKTLRAEQSTQEVKNWMTETLLVQANANPDCKTILRALGPAASLEEMMTACQGVGGPSHKA
 RVLAEAMSQ-ANSTNIMMQRGNFKGPKRIKCFNCGKEGHLARNCRAPRKKGCWKCGKEG
 HQMKDC-IERQANFLGKIWPSNKG-RPGNFLQNRPE-----PTAPPVEPTAPPAEPTAP
 PAESFKFEE--TTPAPKQEPKDRE--PLTSLRSLFGSDPLSQ

>gagC.syn3.2

MGARASILRGEKLDWEKIRLRPGGRKHYMLKHIVWASRELEKFALNPGLLETSEGCKQI
 MKQLQPALQTGTEELKSLYNTVATLYCVHEKIEVRDTKEAVDKIEEEQNKSQOKTQQA
 ADE---KVSQNYPIVQNIQGMVHQPISPRTLNAWVKVIEEKGFNPEVIMFTALSEGAT
 PQDLNTMLNTVGGHQAAMQMLKDTINDEAAEWDRLHPVHAGPIAPGQMREPRGSDIAGST
 STLQEQITWMTSNPPVPVGEIYKRWIILGLNKIVRMYSVPSILDIKQGPKEPFRDYVDRF
 FKVLRAEQATQDVKNWMTDTLLIQANANPDCKTILKALGPGASLEEMMTACQGVGGPGHKA
 RVLAEAMSQ-ANNTNIMMQRSNFKGPKRIVKCFNCGKEGHIKNCRAPRKKGCWKCGQEG
 HQMKDC-TERQANFLGKIWPSQKG-RPGNFLQNRP-----EPSAP
 PAESFRFGE--TTPAPKQELKDRE--PLTSLKSLFGSDPLSQ

>gagC.syn3.3

MGARASILRGGKLDKWEKIRLRPGGKKRYMLKHLIWASRELERFALNPGLLETAEGCKQI
 IKQLQPALQTGTEELRSLYNTVATLYCVHKRIDVRDTKEALDKIEEEQNKIQQKTQQA
 ADG---KVSQNYPIVQNAQGMVHQAI SPRTLNAWVKVVEEKAFSPEIIPMFTALSEGAT
 PQDLNSMLNAVGGHQAAMQMLKDTINEEAAEWDRVHPVHAGPVAPGQIREPRGSDIAGTT
 STLQEQIAWMTSNPPIPVGDYIKRWIIMGLNKIVRMYSVPSILDIKQGPKEPFRDYVDRF
 FKTLRAEQATQEVKNWMTDTLLVQANANPDCKTILRALGPGATLEEMMTACQGVGGPSHKA
 RVLAEAMSQ-TNSA-ILMQRSNFKGSKRIVKCFNCGKEGHIARNCRAPRKRGCWKCGKEG
 HQMKEC-TERQANFLGKIWPSHKG-RPGNFLQSRP-----EPTAP
 PAESFRFEE--TTPAPKQESKDRE--PLISLKSFLGNDPLSQ

>gagC.syn4.1

MGARASVLRGEKLDWERIKLRPGGKKHYMIKHLVWASRELEKFALNPGLLETSEGCKQI
 IQQLQPALKTGTEELKSLYNTVATLYCVHERIDVRDTKEALDKIEEEQNKIQQKTQQA
 ADG---KVSQNYPIVQNAQGMVHQALSPRTLNAWVKVVEEKAFSPEIIPMFTALSEGAT
 PSDLNTMLNTIGGHQAAMQMLKDTINDEAAEWDRLHPVHAGPIAPGQIREPRGSDIAGTT
 STLQEQITWMTSNPPVPVGEIYKRWIILGLNKIVRMYSVPSILDIKQGPKEPFRDYVDRF
 FKTLRAEQSTQEVKNWMTETLLVQANANPDCKTILKALGPGATLEEMMTACQGVGGPGHKA
 RVLAEAMSQ-TNSA-ILMQRSNFKGSKRIVKCFNCGREGHIARNCKAPRKKGCWKCGKEG
 HQMKDC-IERQANFLGKIWPSHKG-RPGNFLQNRPE-----PTAPP-----EPTAP
 PAESFRFGE--TTPAPKQEQKDRE--PLISLKSFLGSDPLLQ

>gagC.syn4.2

MGARASILRGEKLDKWERIKLRPGGKKHYMLKHLVWASRELDRLFALNPGLLETSDGCKQI
 IKQLHPALQTGTEELRSLFNTVATLYCVHAGIEVRDTKEAVDKIEEEQNKSQOKTQQA
 ADE---KVSQNYPIVQNIQGMVHQPI SPRTLNAWVKVIEEKAFSPEVIMFTALSDGAT
 PQDLNTMLNTVGGHQAAMQMLKDTINEEAAEWDRHPVHAGPVAPGQMRDPRGSDIAGST
 STLQEQIAWMTNNPPVPVGDYIKRWIIMGLNKIVRMYSVPSILDIRQGPKEPFRDYVDRF
 FRTLRAEQATQDVKNWMTDTLLIQANANPDCKTILRALGPGASLEEMMTACQGVGGPGHKA
 RVLAEAMSQ-ANNTNIMMQSNFKGPKRTVKCFNCGKEGHIARNCRAPRKRGCWKCGKEG
 HQMKEC-TERQANFLGKIWPSQKG-RPGNFLQSRP-----EPSAP
 PAESFRFEE--TTPAPKQESKDRE--PLTSLKSLFGSDPSSQ

>gagC.syn4.3

MGARASILRGGKLDWEKIRLRPGGKKRYMLKHLIWASRELERFALNPGLLETSEGCKQI
 MKQLQPALQTGTEELRSLYNTVATLYCVHKGIVQDTKEALDKIEEEQKKSQOKTQQA
 ADK--GKVSQNYPIVQNLQGMVHQPLSPRTLNAWVKVIERKAFSPEVIMFALSALSEGAT
 PQDLNSMLNAVGGHQAAMQMLKDTINEEAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTT
 SNLQEQIAWMTSNPPIPVGDYIKRWIVLGLNKIVRMYSVPSILDIKQGPKEPFRDYVDRF
 FKTLRAEQATQEVKNWMTDTLLVQANANPDCKTILRALGPAASLEEMMTACQGVGGPGHKA

21/109

Fig. 9 cont'd-7

RVLAEAMSQ-ANS-NIMMQRGNFKGPKRI IKCFNCGKEGHLARNCRAPRKKGCWKCGREG
 HQMKDC-TERQANFLGKIWPSNKG-RPGNFLQSRP-----EPTAP
 PAESFGFGE--TTPAPKQELKDRE--PLTSLKSLFGNDPLSQ

>gagC.syn4.4

MGARASILRGGKLDKWEKIRLRPGGRKHYMLKHIVWASRELERFALNPGLLETAEGCKQI
 IKQLQPALQTGTEELKSLFNTVATLYCVHEKIEVRDTKEALDKIEEEQNKCQKQTQQA
 ADG---KVSQNYPIVQNLQGMVHQAI SPRTLNAWVKVIEEKGFNPEVIMFTALSEGAA
 PQDLNMTLNTIGGHQAAMQMLKDTINEEAAEWDR LHPVQAGPVAPGQMREPRGSDIAGTT
 SSLQE QIAWMTGNPPVPVPGDIYKRWIILGLNKIVRMYS PVSILDIRQGPKEPFRDYVDRF
 FKVLRAEQATQDVKNWMTETLLVQANANPDCKTILRALGPGATLEEMMTACQGVGGPSHKA
 RVLAEAMSQ-ANNANIMMQR SNFKGPKRIVKCFNCGKEGHI AKNCRAPRKKGCWKCGQEG
 HQMKDC-NERQANFLGRIWPSHKG-RPGNFIQSRPEPTAPLEPTAPPA-----EPTAP
 PAESFKFEE--TTPAPKQEPKDRE--PLTSLRSLFGSDPLSQ

>gagC.syn6.1

MGARASVLRGEKLDTWERIKLRPGGKKHYMIKHLVWASRELEKFALNPGLLETAEGCKQI
 IRQLQPALQTGTEELRSLYNTVATLYCVHKRIDVRDTKEALDKIEEEQNKSQKQAQQA
 ADG---KVSQNYPIVQNLQGMVHQSLSPRTLNAWVKVIERKA FSPEVIMFSA LSEGAT
 PQDLNMTLNTVGGHQAAMQMLKDTINEEAAEWDR IHPVHAGPIAPGQMREPRGSDIAGTT
 STLQE QITWMTSNPPVPVGEIYKRWIILGLNKIVRMYS PVSILDIRQGPKEPFRDYVDRF
 FKTLRAEQSTQEVKNWMTDTLLAQANANPDCKIILRGLGPGATLEEMMTACQGVGGPGHKA
 RVLAEAMSQ-ANS-NILMQR SNFKGPRRTIKCFNCGKEGHL AKNCRAPRKKGCWKCGKEG
 HQMKEC-TERQANFLGKIWPSQKG-RPGNFLQSRP-----EPSAP
 PAESFRFEE--TTPALKQEPKDRE--PLTSLRSLFGSDPLSQ

>gagC.syn6.2

MGASASILRGEKLDKWEKIRLRPGGKKCYMLKHI IWASKELERFALNPGLLETSEGCKQI
 MKQLQPALQTGTEELKSLYNTVATLYCVHEKIEVRDTKEALDKIEEEQKKSQKQTQQA
 ADK---GKVSQNYPIVQNAQGMVHQALSPRTLNAWVKVVEEKAFSPEIIPMFTALSEGAA
 PQDLNMTLNTVGGHQAAMQMLKDTINEEAAEWDR VHPVHAGPIAPGQIREPRGSDIAGTT
 STLQE QVAWMTSNPPVPVPGDIYKRWIVLGLNKIVRMYS PVSILDIKQGPKEPFRDYVDRF
 FKTLRAEQSSQEVKNWMTDTLLVQANANPDCKTILRALGPAASLEEMMTACQGVGGPSHKA
 RVLAEAMSQ-ANSTNIMMQRGNFKGPKRIVKCFNCGREGHI ARNCRAPRKKGCWKCGQEG
 HQMKDC-IERQANFLGKIWPSHKG-RPGNFIQSRPE-----PTAPP-----EPTAP
 PAESFRFGE--TTPAPKQESKDRE--PLTSLKSLFGNDPLSQ

>gagC.syn6.3

MGARASVLKGEKLDKWERIRLRPGGKKQYRLKHLVWASRELERFALNPSSLLETSEGCRQI
 IKQLQPALKTGTEELRSLYNTIATLYCVHKG IYVQDTKEALDKVEEEQNKSQKQTQQA
 ADE---KVSQNYPIVQNLQGMVHQPLSPRTLNAWVKVIEEKGFNPEVIMFTALSEGAT
 PQDLNSMLNAVGGHQAAMQMLKDTINEEAAEWDR LHPVQAGPVAPGQIREPRGSDIAGTT
 SSLQE QIAWMTGNPPVPVPGDIYKRWIILGLNKIVRMYS PVSILDIKQGPKEPFRDYVDRF
 FKTLRAEQATQDVKNWMTETLLVQANANPDCKTILRALGPGATLEEMMTACQGVGGPSHKA
 RVLAEAMSQ-ANNANIMMQR SNFKGPKRTVKCFNCGKEGHI AKNCRAPRKKGCWKCGREG
 HQMKDC-IERQANFLGKIWPSNKG-RPGNFLQNRTE-----PTAPPA-----EPTAP
 PAESFRFEE--TTPAPKQELKDRE--PLTSLKSLFGSDPLLQ

>gagC.syn6.4

MGARASILRGEKLDKWEKIRLRPGGRKHYMLKHIVWASRELEGFALNPGLLETSEGCKQI
 IKQLQPALQTGTEELRSLFNTVATLYCVHSGIEVRDTKEAVDKIEEEQNKIQKQTQQA
 ADG---KVSQNYPIVQNSQGMVHQAI SPRTLNAWVKVIEEKAFSPEVIMFTALSEGAT
 PSDLNMTLNTIGGHQAAMQMLKDTINDEAAEWDR LHPVHAGPIAPGQMRDPRGSDIAGST
 STLQE QIAWMTNNPPVPVPGDIYKRWIIMGLNKIVRMYS PVSILDIKQGPKEPFRDYVDRF
 FRTLRAEQATQEVKNWMTETLLVQANANPDCRTILKALGPGATLEEMMTACQGVGGPGHKA
 RVLAEAMSQ-ANNINIMMQRNNFKGPKRI IKCFNCGKEGHI ARNCKAPRKKGCWKCGKEG
 HQMKDC-NERQANFLGKIWPSHKG-RPGNFLQNRP-----EPTAP
 PAESFRFEE--TTPTPKQEPKDRE--PLTSLKSLFGSDPSSQ

22/109

Fig. 9 cont'd-8

>gagC.syn6.5

MGARASILRGGKLDKWEKIRLRPGGKKRYMLKHLI WASRELERFALNPGLLETSDGCKQI
 IQQLQPALKTGTEELKSLFNTVAVLYCVHKGIEVRDTKEAVDKIEEEQNKIQQKMQQQKV
 TDG---KVSQNYPIVQNIQGMVHQPI SPRTLNAWVKVIEEKGFSP EVIPMFTALSEGAT
 PSDLNTMLNTVGGHQAAMQMLKDTINEEAAEWDRTHPVHAGPVAPGQMREPRGSDIAGTT
 STLQEQIGWMTNNPPIPVGEIYKRWIILGLNKIVRMYS PVSILDIKQGPKEPFRDYVDRF
 FKALRAEQATQEVKGWMTDTLLVQANANPDCKTILKALGSGATLEEMMTACQGVGGPGHKA
 RVLAEAMSQ-ANNTNIMMQSNFKGPRRIVKCFNCGREGHIAKNCRAPRKKGCWKCGQEG
 HQMKDC-TERQANFLGRIWPSHKG-RPGNFLQSRPE-----PTAPL-----QPTAP
 PAESFKFEE--TTPAPKQEQKDRE--PLTSLRSLFGNDPLSQ

>gagC.syn6.6

MGARASILRGGKLDTWEKIRLRPGGKKHYMLKHLV WASRELD RFALNPGLLETADGCKQI
 IKQLHPALQTGTEEIKSLFNTVATLYCVHAGIEVRDTKEALDKIEEEQNKCQKQTQQAKE
 ADK---KVSQNYPIVQNLQGMVHQAI SPRTLNAWVKVIEEKAFNPEIIPMFTALSDGAT
 PQDLNTMLNTVGGHQAAMQMLKDTINEEAAWDRLHPVHAGPVAPGQLREPRGSDIAGTT
 SNLQEQIAWMTSNPPIPVGDIYKRWIILGLNKIVRMYS PVSILDIKQGPKEPFRDYVDRF
 FKVLRAEQATQDVKNWMTDTLLIQANANPDCKTILRALGPGASLEEMMTACQGVGGPSHKA
 RVLAEAMSQ-TNSA-ILMQRSNFKGSKRIVKCFNCGKEGHLARNCRAPRKRGCWKCGKEG
 HQMKDCTTERQANFLGKIWPSHKGGRPGNFLQNRPE-----PTAPL-----EPTAP
 PAESFGFGE--TTPAPKQEPKDRE--PLISLKS LFGSDPLSQ

>gagM.syn1.1

MGARASILRGGKLDKWEKIRLRPGGKKHYMLKHLV WASRELERFALNPGLLETSEGCKQI
 IKQLQPALQTGTEELRSLFNTVATLYCVHEKIEVRDTKEALDKIEEEQNKSQKQTQQAQA
 ADG---KVSQNYPIVQNLQGMVHQAI SPRTLNAWVKVIEEKAFSP EVIPMFTALSEGAT
 PQDLNTMLNTVGGHQAAMQMLKDTINEEAAEWDR LHPVHAGPIAPGQMREPRGSDIAGTT
 STLQEQIAWMTSNPPIPVGEIYKRWIILGLNKIVRMYS PVSILDIRQGPKEPFRDYVDRF
 FKTLRAEQATQDVKNWMTDTLLVQANANPDCKTILRALGPGATLEEMMTACQGVGGPSHKA
 RVLAEAMSQ-ANNTNIMMQR SNFKGSKRIVKCFNCGKEGHIARNCRAPRKKGCWKCGKEG
 HQMKDC-TERQANFLGKIWPSHKG-RPGNFLQSRPE-----PTAPPA-----EPTAP
 PAESFRFEE--TTPAPKQEPKDRE--PLTSLKS LFGSDPLSQ

>gagM.syn3.1

---RASVLSGGKLDWEKIRLRPGGKKKYKLKHIV WASRELERFAVNPG LLETSEGCRQI
 LGQLQPSLQTGSEELRSLYNTVAVLYCVHQRIE VKDTKEALEKIEEEQNKS KKAQAAAA
 DTGNSSQVSQNYPIVQNLQGMVHQAI SPRTLNAWVKVIEEKAFSP EVIPMFTALSEGAT
 PSDLNTMLNTIGGHQAAMQILKDTINEEAAEWDR LHPVHAGPIAPGQMREPRGSDIAGTT
 STLQEQIGWMTNNPPIPVGEIYKRWIILGLDKIVRMYS PPTSILDIRQGPKEPFRDYVDRF
 YKTLRAEQASQEVKNWMTETLLVQANANPDCKTILKALGPAATLEEMMTACQGVGGPSHKA
 RILAEAMSQVTNSATIMMQRGNFRNQRKT VKCFNCGKEGHI AKNCRAPRKKGCWKCGKEG
 HQMKDC-NERQANFLGKIWPSHKG-RPGNFLQSRP-----EPTAP
 PEESFRFGEETTPSQKQEPIDKELYPLASLKS LFGNDPLSQ

>gagM.syn3.2

MGARASILRGGKLDKWEKIRLRPGGKKHYMLKHLV WASRELERFALNPGLLETAEGCKQI
 IKQLQPALKTGTEELKSLYNTVATLYCVHEKIEVRDTKEALDKLEEEQNKSQKQTQAAAA
 GTGSSSKVSQNYPIVQNAQGQMVHQALSPRTLNAWVKVVEEKAFSP EVIPMFSALAEAT
 PQDLNTMLNTVGGHQAAMQMLKDTINEEAAWDRLHPVHAGPVAPGQMREPRGSDIAGST
 STLQEQIAWMTSNPPIPVGDIYKRWIILGLNKIVRMYS PVSILDIKQGPKEPFRDYVDRF
 FKTLRAEQATQDVKNWMTDTLLVQANANPDCKTILRALGPGASLEEMMTACQGVGGPGHKA
 RVLAEAMSQ-ANNANIMMQRGNFKGQKR- IKCFNCGKEGHLARNCRAPRKKGCWKCGREG
 HQMKDC-TERQANFLGRIWPSSKG-RPGNFPQSRP-----EPSAP
 PAESFGFGE--TTPAPKQELKDRE--PLTSLKS LFGSDPLSQ

>gagM.syn3.3

MGARASVLSGGELDRWEKIRLRPGGKKKYRLKHLV WASRELEKFALNPGLLETSEGCKQI
 MKQLQPALQTGTEELRSLFNTVATLYCVHQRIDVKDTKEALDKIEEEQNKIQQKTQQAQA
 ADG---KVSQNYPIVQNIQGMVHQPI SPRTLNAWVKVVEEKGFNPEVIPMFSALSEGAT

23/109

Fig. 9 cont'd-9

PQDLNMMMLNIVGGHQAAMQMLKETINEEAAEWDRVHPVHAGPIPPGQMREPRGSDIAGTT
 SNLQEQIGWMTSNPPVPVGDYKRWIVLGLNKIVRMYSVPSILDIRQGPKEPFRDYVDRF
 FKVLRAEQATQEVKNWMTDTLLIQNANPDCKSILRALGPGATLEEMMTACQGVGGPSHKA
 RVLAEAMSQ-TNSA-ILMQRSNFKGSKRIVKCFNCGKEGHIARNCRAPRKRGCWKCGQEG
 HQMKDC-TERQVNFLGKIWPSNKG-RPGNFLQNRPE-----PTAPPA-----EPTAP
 PAESFRFEE--TTPAPKQEPKDRE--PLTSLRSLFGNDPSSQ

>gagM.syn4.1

MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQI
 LGQLQPSLQTGSEELRSLYNTVAVLYCVHQRIDVKDTKEALEKIEEENQKSQKQTQQA
 ADG---KVSQNYPIVQNAQGMVHQAI SPRTLNAWVKVIEEKGFSPEVIMFSALEAGAT
 PQDLNMTLNTIGGHQAAMQMLKDTINDEAAEWDRHLHPVQAGPVAPGQIREPRGSDIAGTT
 SNLQEQIGWMTSNPPPIVGDYKRWIIMGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRF
 FRTLRAEQASQEVKNWMTETLLVQNSNPDKTILKALGPAATLEEMMTACQGVGGPGHKA
 RVLAEAMSQ-VQQPNIMMQRGNFKGQKR- IKCFNCGREGHLARNCRAPRKKGCWKCGREG
 HQMKDC-TESKANFLGKIWPSNKG-RPGNFLQSRP-----EPSAP
 PAESFGFEE-ITPSQKQEQDKELYPLASLKSFLGNDPLSQ

>gagM.syn4.2

MGARASVLSGGKLDWEKIRLRPGGKKKYRLKHLVWASRELDRLFALNPSLLETAEGCKQI
 MKQLQPALKTGTEELKSLYNTVATLYCVHEKIDVRDTKEALDKIEEENQNKIQKQTQQA
 ADG---KVSQNYPIVQNIQGMVHQPI SPRTLNAWVKVVEEKAFSPEVIMFTALSDGAT
 PQDLNMLNAVGGHQAAMQILKDTINEEAAEWDRHLHPVHAGPVAPGQMREPRGSDIAGTT
 STLQEQIGWMTNNPPPIVGEIYKRWIILGLNKIVRMYSVPSILDIKQGPKEPFRDYVDRF
 FKVLRAEQATQDVKNWMTDTLLIQNANPDCKSILRALGPGATLEEMMTACQGVGGPSHKA
 RILAEAMSQVNTSATIMMQRGNFRNQRKTVKCFNCGKEGHLARNCKAPRKRGCWKCGKEG
 HQMKEC-TERQANFLGKIWPSSKG-RPGNFPQSRP-----EPTAP
 PEESFRFGEETTTSPQKQEPIDKELYPLASLRSFLGNDPSSQ

>gagM.syn4.3

MGARASILRGGKLDKWEKIRLRPGGKKRYMLKHLI WASRELERFALNPGLLETAEGCQOI
 IEQLQSTLKTGSEELKSLFNTVATLYCVHQRIEVKDTKEALDKVEEENQNSKKKAQAAAA
 DTGNSSQVSQNYPIVQNLQGMVHQALSPRTLNAWVKVIEEKAFSPEIIPMFTALSEGAT
 PSDLNTMLNTVGGHQAAMQMLKDTINEEAAEWDRVHPVHAGPIPPGQMREPRGSDIAGTT
 SSLQEQIAWMTSNPPVPVGEIYKRWIVLGLNKIVRMYSVPSILDIRQGPKEPFRDYVDRF
 FKTLRAEQASQDVKNWMTETLLVQANPDCKTILRALGPGASLEEMMTACQGVGGPSHKA
 RVLAEAMSQ-TNSA-ILMQRSNFKGSKRIVKCFNCGKEGHIARNCRAPRKRGCWKCGQEG
 HQMKDC-NERQANFLGKIWPSSKG-RPGNFLQNRPE-----PTAPP-----EPTAP
 PAESFRFEE--TTPAPKQELKDRE--PLTSLKSLFGSDPLSQ

>gagM.syn4.4

MGARASVLRGEKLDKWERIRLRPGGKKHYMLKHLVWASRELEKFALNPGLLETSEGCKQI
 IKQLQPALQTGTEELRSLFNTVATLYCVHAGIEVRDTKEALDKIEEIQNKSQKQTQAAAA
 GTGSSSKVSQNYPIVQNLQGMVHQPLSPRTLNAWVKVVEEKGFNPEVIMFSALSEGAT
 PQDLNMMMLNIVGGHQAAMQMLKETINEEAAEWDRHLHPVHAGPIAPGQMREPRGSDIAGST
 STLQEQIAWMTGNPPVPVGDYKRWIILGLNKIVKMYSPSILDIKQGPKEPFRDYVDRF
 YKTLRAEQATQEVKNWMTDTLLVQANPDCKSILKALGTGATLEEMMSACQGVGGPAHKA
 RVLAEAMSQ-ANNTNIMMQRSNFKGPKRI IKCFNCGKEGHIANKRANRPRKKGCWKCGKEG
 HQMKDC-TERQANFLGRIWPSSKG-RPGNFLQSRPE-----PTAPPA-----EPTAP
 PAESFKFEE--TTPAPKQEPKDRE--PLTSLRSLFGSDPLLQ

>gagM.syn6.1

MGARASILSGGKLDWEKIRLRPGGRKHYMLKHIVWASRELERFALNPGLLETAEGCQOI
 IEQLQSTLKTGSEELKSLFNTVATLWCVHQRIEVKDTKEALDKLEENQNSQKQTQQA
 ADG---KVSQNYPIVQNLQGMVHQSI SPRTLNAWVKAIEEKAFSPEVIMFSALEAGAT
 PQDLNMTLNTIGGHQAAMQILKDTINEEAAEWDRHHPVHAGPVAPGQMRDPRGSDIAGTT
 SNLQEQIAWMTSNPPVPVGEIYKRWIILGLDKIVRMYSVPSILDIRQGPKEPFRDYVDRF
 FKTLRAEQATQEVKGWMTDTLLVQANPDCKTILKALGPGATLEEMMSACQGVGGPGHKA

24/109

Fig. 9 cont'd-10

RVLAEAMSQ-ANNTNIMMQKSNFKGPKRI IKCFNCGKEGHLARNCRAPRKKGCWKCQEG
 HQMKDC-TERQANFLGRIWPSHKG-RPGNFPQSRL-----EPTAP
 PAESFGFGEE-IAPSPKQEPKEKELYPLTSLKSLFGNDPLSQ

>gagM.syn6.2

MGARASILRGGLDKWEKIRLRPGGKKYKLKHIVWASRELEKFALNPGLLETSEGCRQI
 LGQLQPSLQGTSEELKSLYNTVATLYCVHQRIDVKDTKEALEKIEEEQNKSQQKTQAAAA
 DKG----VSQNYPIVQNLQGQMVHQAI SPRTLNAWVKVIEEKAFSPEIIPMFTALSEGAT
 PQDLTTMLNTVGGHQAAMQMLKETINDEAAEWDR LHPVHAGPVAPGQLREPRGSDIAGST
 STLQEQIAWMTGNPPVPVGDYKRWIVLGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRF
 YKTLRAEQASQDVKNWMTETLLVQANPDCRTILKALGPAATLEEMMTACQGVGGPAHKA
 RVLAEAMSQVTNPATIMMQRGNFRNQRKTVKCFNCGKEGHLAKNCRAPRKRGCWKCQEG
 HQMKDC-NERQANFLGKIWPSNKG-RPGNFLQNR-----EPTAP
 PAESFRFGEEKTPSQKQEPIDKELYPLASLRSLSFGNDPLSQ

>gagM.syn6.3

MGARASVLRGEKLDKWERIRLRPGGKKRYMLKHLI WASRELERFALNPSSLLETSEGCKQI
 IQQLQPALKTGTEELRSLYNTVATLYCVHEKIEVRDTKEAVDKIEEEQNKSCKKAQAAAA
 DTGNSSQVSQNYPIVQNIQGQMVHQALSPRTLNAWVKVVEEKGFNPEVIMFSALESEGAT
 PQDLNMLNIVGGHQAAMQMLKDTINEEAAEWDR LHPVQAGPVAPGQMPREPRGSDIAGTT
 STLQEQITWMTSNPPIPVGEIYKRWIIMGLNKIVRMYSVPSILDIKQGPKEPFRDYVDRF
 FRTLRAEQASQEVKNWMTETLLIQNANPDCKTILRALGPAASLEEMMTACQGVGGPGHKA
 RVLAEAMSQ-TNSA-ILMQRSNFKGSKRIVKCFNCGKEGHIARNCRAPRKKGCWRCGKEG
 HQMKDC-TESKANFLGKIWPSHKG-RPGNFLQNRPEPTAPPEPTAPPAEPTAPPAEPTAP
 PAESFKFEE--TTPAPKQELKDRE--PLISLKSLSFGSDPLLQ

>gagM.syn6.4

MGARASILRGEKLDWEKIRLRPGGKKQYRLKHIVWASRELD RFALNPSSLLETAEGCKQI
 IKQLHPALQTGTEELRSLFNTVATLYCVHAGIEVRDTKEALDKIEEEQNKIQQKTQAKA
 ADE---KVSQNYPIVQNMQGQMVHQPLSPRTLNAWVKVVEEKAFSPEVIMFAALSEGAT
 PSDLNTMLNTVGGHQAAMQMLKDTINDEAAEWDR LHPAQAGPIPPGQIREPRGSDIAGTT
 STPQEQIGWMTNNPPIPVGEIYKRWIVLGLNKIVRMYSPI SILDIRQGPKEPFRDYVDRF
 FKALRAEQATQEVKGWMTETLLVQNSNPDCCKTILRALGPGASLEEMMTACQGVGGPSHKA
 RILAEAMSQ-ANS-NIMMQRSNFKGPKRIVKCFNCGKEGHI AKNCRAPRKKGCWKCQREG
 HQMKDC-IERQANFLGKIWPSQKG-RPGNFLQSRP-----EPSAP
 PAESFRFGE--TTPAPKQEPKDRE--PLTSLRSLFGSDPLSQ

>gagM.syn6.5

MGARASVLSGGELDRWEKIRLRPGGKKYRLKHLVWASRELERFAINPGLLETSDGCKQI
 IKQLQPALQTGSEELRSLYNTIATLYCVHQKIEVKDTKEALDKIEEIQNKSKQKTQAAAA
 GTGSSSKVSQNYPIVQNAQGQMVHQSLSPRTLNAWVKVIEEKGFNPEVIMFTALSEGAT
 PHDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVHAGPIPPGQMPREPRGSDIAGST
 STLQEQIGWMTSNPPIPVGDIYKRWIILGLNKIVRMYS PSSILDIRQGPKEPFRDYVDRF
 FKCLRAEQATQEVKNWMTDTLLIQNANPDCKSILRALGPGATLEEMMTACQGVGGPSHKA
 RILAEAMSQ-VQQPNIMMQRGNFKGQKR- IKCFNCGREGHIARNCKAPRKKGCWKCQKEG
 HQMKDC-TERQVNFLGKIWPSYKG-RPGNFLQSRP-----EPTAP
 PEESFRFGEETTPSQKQETIDKELYPLASLKSLSFGNDPSSQ

>gagM.syn6.6

MGARASVLSGGKLDAWERIRLRPGGKKHYMLKHLVWASRELERFAVNPGLLETSEGCKQI
 MKQLQPALQTGTEELKSLYNTVAVLYCVHQRIEIKDTKEALDKIEEIQNKCCQKTQAAKE
 ADG---KVSQNYPIVQNLQGQMVHQPI SPRTLNAWVKVIEEKGFSPPEVIMFTALSDGAT
 PQDLNSMLNAVGGHQAAMQMLKDTINEEAAEWDR LHPVHAGPIAPGQMPREPRGSDIAGTT
 SSLQEQIAWMTNNPPVPVGEIYRRWIILGLNKIVKMYSPTSILDIKQGPKEPFRDYVDRF
 FKVLRAEQATQDVKNWMTDTLLVQANANPDCKSILKALGTGATLEEMMTACQGVGGPSHKA
 RVLAEAMSQVTNSATIMMQRGNFRNQRKIVKCFNCGREGHLARNCKAPRKRGCWKCQKEG
 HQMKEC-TERQANFLGKIWPSSKG-RPGNFPQSRP-----EPTAP
 PAESFRFEE--TTPAPKQESKDRE--PLTSLKSLFGSDPSSQ

25/109

Fig. 10

>ENV-B.syn1.1

MRVTGIRKNYQHLWRWGTMLLWRWGTMLLGLMICSATEKLWVTVYYGVPVWKEATTTLF
 CASDAKAYDTEVHNVWATHACVPTDPNPQEVVLENTENFNMWKNNMVEQMHEDIISLWD
 QSLKPCVKLTPLCVTLNCTDDVRNVT-NNATNTNSSW-GEPMEKGEIKNCSFNITTSIRD
 KVQKEYALFYKLDVVP-NDNSNNTN-----YRLISCNTSVITQACPKVSFEPIPIH
 YCAPAGFAILKCNDKKFNGTGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEEVVIRSEN
 FTNNAKTIMVQLNVSVEINCTRPNNNTRKSIHIGPGRFYTTGDIIGDIRQAHCNISRAQ
 WNNTLKHIVEKLKGQFGNNKTIVFNHSSGGDPEIVMHSFNCGGEFFYCNSTKLFNSTWTR
 N-NGTWTRN---DTERSSTE---EHITLPCRIKQIINMWQEVGKAMYAPPIRGQIRCSS
 NITGLLLTRDGGNDT-----SGTEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTAK
 RRVVQREKRAVG-IGAVFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQNNLLRAIEA
 QQHLLQLTVWGIKQLQARVLAVERYLRDQQLLGIWGC SGKLICTTAVPWNASWSNKSLE
 IWNNMTWMEWEKEIDNYTNLIYNLLEESQNQQEKNEQELLELDKWASLWNWFDISNWLWY
 IKIFIMIVGGLVGLRIVFAVLSIVNRVRQGYSPFSFQTRLPA PRGPDRPEGIEEEGGERD
 RDRSVRLVDGFLALIWDRLSLCLFSYHRLRDL LLIVTRIVELLG-----RGWEALK
 YWWNLLQYWSQELKNSAVSLLNATAIAVAEGTDRVIEALQRACRAILHIPRRIRQGLERA
 LL-

>ENV-B.syn3.1

MRVKETRKNYQHLWKWGTM-----LGMLMICSATEKLWVTVYYGVPVWRDANATLF
 CASDAKAYDTEAHNVWATHACVPTDPNPQEVLEKNVTENFNMWKNDMVEQMHEDIINLWD
 QSLKPCVELTPLCVTLNCTDYVKNIT-NNATSTNSSW-GKPMKGEIKFCSFNITTSIRN
 KVQKQYALFYKLDIVPI-DNDNTS-----YRLISCNTSTITQACPKVTFEPIPIH
 YCAPAGFAILKCNNKTFNGTGPCTNVSTVQCTHGIRPVISTQLLNGSLAEEEVVIRSEN
 FTNNAKTIMVQLNVSVEINCTRPNNNTRRSIPIGPGRVFTTEDIIGDIRQAHCNISRAQ
 WNNTLKHIVEKLKGQFGNNKTIVFNHSSGGDPEIVMHSFNCRGEFFYCKSTKLFNSTWTR
 N-NGTWTRN---DTERSSTE---EHITLPCRIKQIINMWQEVGKAMYAPPIKGQISCSS
 NITGLLLTRDGGNDT-----SGTEIFRPGGGDMRDNRSELYKYKVVRIEPLGVAPTEAK
 RRVVQREKRAVG-IGAVFLGFLGAAGSTMGAASMTLTVQARLLLSGIVQQSNLLRAIEA
 QQHMLQLTVWGIKQLQARLLAVERYLGDQQLLGLWGCSGKLICTTTVPWNWSNKSLE
 IWDNMTWMQWEREIDNYTGLIYNLLEKSQNQQEKNEQELLELDKWASLWNWFDITNWLWY
 IKIFIMIVGGLVGLRIVFTVLSIVNRVRKGYSPFSFQTRLPTPRGPDRPGGIEEEGGEQD
 RDRSGPLVNGFLALIWDRLSLFLFSYHRLRDL LLIVARIVELLG-----RGWEILK
 YWWNLLLYWSQELKNSAVSLLNATAIAVAEGTDRVIEVVQRAFRAILHIPRRIRQGFERA
 LL-

>ENV-B.syn3.2

MRVTGIRKNYQHLWRWGTMLLWRWGTMLLGLMICSAAAGKLWVTVYYGVPVWKEANTTLF
 CASDAKAYDTEVHNVWATHACVPIDPNPQEVVLGNVTENFNMWKNNMVEQMHEDIISLWD
 ESLKPCVKLTPICVTLNCTDDVRNVT-NNATNTNSSW-GEPMEKGEIKNCSFNITTSIRD
 KVQKQYALFYKLDVVP-NDNSNNTN-----YRLISCNTSVITQACPKISFEPIPIH
 FCAPAGFAILKCNDKKFNGTGPCKNVSTVQCTHGKIPVVSTQLLNGSLAEEEIVIRSEN
 FTDNAKTIIVQLNESVVINCTRPNNNTRKSIHIGPGRFYATGEIIGDIRQAHCNLSRAK
 WNDTLKQIVIKLREQFG-NKTIVFNQSSGGDPEIVMHSFNCGGEFFYCNTTQLFNSTWTW
 N-NSTW--N---NTRKSDTE---EIITLPCRIKQIVNMWQKVGKAMYAPPIRGQIRCSS
 NITGILLTRDGGNNNET---NRTETFRPGGGNMKDNWRSELYKYKVVKIEPLGIAPTAK
 RRVVQREKRAVGITIGAMFLGFLGAAGSTMGAASLTTLTVQARQLLP GIVQQNNLLKAIEA
 QQHLLRLTVWGIKQLQARVLAVERYLRDQQLLGIWGC SGKIICTTAVPWNASWSNKSLE
 IWDNMTWMEWEKEIDNYTNLIYNLLEESQNQQEKNELELELDKWANLWNWFDISNWLWY
 IKIFIMIIGGLVGLRIVFAVLSIVNRVRQGYSPFSFQTRLPA PRGPDRPEGIEEEGGERD
 RDRSVRLVDGFLALIWDRLSLCLFLYHRLRDL LLIAARIVELLG-----RGWEALK
 YWWNLLQYWSQELKNSAISLLNATAVAVAEAGTDRVIEALQRACRAILHIPRRIRQGLERL
 LL-

>ENV-B.syn3.3

MRVKGIRKNCQHLWRWGIML-----LGMLMICSAAEQWLWVTVYYGVPVWKEATTTLF
 CASDAKAYDKEVHNVWATHASVPTDPNPQEVVLENTENFNMWKNNMVDQMHEDIISLWD
 QSLKPCVKLTPLCVTLNCTD-LRNATNGNDTNTTSSS-REMMGGGEMKNCSFNVTTSIRD
 KVQKEYALFYKLDVIPI-DSRNNNSNSTE--YNSYRLINCNTSVITQACPKVSFEPIPIH

26/109

Fig. 10 cont'd-1

YCTPAGFAILKCKDKKFNGTGPCTKVSTVQCTHGIRPVVSTQLLNGLSLAE EEEV IIRSEN
 FTNNAKTIIVQLKEAVEINCTRPSNNTRKSIPIGPGRFYTTGDIIGDIRKAHCNISRAN
 WNNTLRQIVEKLGEQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYC NSTQLFNSTWNF
 --NGTWNKN---FNNTWNNTTEGTNDTITLPCRIKQIINRWQEVGKAMYAPPISGQIRCSS
 NITGLILTRDGGNNGNET--NGTEIFRPGGGMNRDNWRSELYRYKVVKIEPLGVAPTKAK
 RRVVQREKRAVG-IGALFLGFLGAAGSTMGAASITLTVQARQLLSGIVQQQNLLRAIEA
 QQHLLQLTVWGIKQLQARILAVE RYLKDQQLGIWGC SGKLICTTAVPWNTSWSNRS LNE
 IWNNMTWMEWEREIDNYTSLIYTLIEESQNQQEKNEQE LLDKWASLWNWFSITNWLWY
 IRIFIMIVGGLIGLRIVFAVLSVVRVRQGYSPLSFQTHLPAQRGPDRPEGTEEEGGERD
 RDRSGRLVDGFLAIIWVDLRSCLFSYHRLRDL LLIVTRIVELLG-----RRGWEVLK
 YWWNLLQYWIQELKNSAVSLFNAIAIAVAEGTDRIIEVVQRAYRAILHIPTRIRQGLERA
 LL-

>ENV-B.syn4.1

MRVTGIRKQNYQHLWRWGTM LLWRWGTM LLGILMICS AAGKLWVTVYYGVPVWKDATTTLF
 CASDAKAYDTEVHNWATHASVPTDPNPQEVVLENTEDFNMWKNNMVDQM HEDIISLWD
 QSLKPCVELTPLCVTLNCTDDVRNVT-NNATNTNSSW-GEPMEKGEIKFCSFNITTSIRN
 KVQKQYALFYKLDVVPI-DNDSNNTN-----YRLISCNTSVITQACPKVTFEPIPIH
 YCAPAGFAILKCNKTFNGTGPCTKVSTVQCTHGIRPVVSTHLLNGLSLAE EEEV IIRSEN
 FTDNTKTIIVQLKEAVEINCTRPNNNTRKGIHIGPGRFYTTGEIIGDIRQAHCNLSRAK
 WNDTLKQIVIKLREQFG-NKTIIFNQSSGGDPEIVMHTFNCGGEFFYCNTTQLFNSTW--
 -----QN---ETSGSINITDIGENITLPCRIKQIVNMWQKVGKAMYAPPIKGOISCSS
 NITGLLLTRDGGNNGNET--NGTEIFRPGGGMKDNWRSELYRYKVVKIEPLGVAPTRAK
 RRVVQREKRAVT-LGAMFLGFLGAAGSTMGAASMTLTVQARLLLSGIVQQQSNLLRAIEA
 QQHLLRLTVWGIKQLQARILAVE RYLQDQQLGIWGC SGKLICTTAVPWNASWSNKSQDE
 IWNNMTWMQWEKEIDNYTGLIYTLLEESQIQQEKNEQE LLELDKWASLWNWFDITNWLWY
 IKIFIMIVGGLIGLRIVFTVLSIVNRVRKGYSPLSFQTHLPA PRGPDRPGGIEEEEGGEQD
 RDRSGPLVNGFLALI WVDLRSCLFSYHRLRDL LLIVARIVELLG-----RRGWEVLK
 YWWNLLQYWSQELKNSAISLLNATAVAVAEGTDRVIEALQRACRAILHIPRRIRQGLERL
 LL-

>ENV-B.syn4.2

MRVKGIRKNCQHLWRWGILL-----LGMLMICSAAEQ LWVTVYYGVPVWRDANATLF
 CASDAKAYDTEAHNVWATHACVPTDPNPQEVELKNVTENFNMWKNNMVEQM QEDIISLWD
 QSLKPCVKLTPICVTLNCTDYVKNIT-NNATSTNSSW-GKPM EKGEIKNCSFNVTTSIRD
 KVQKEYALFYRLDVVPI-DNDSNNNDSTNTNYTNYRLISCNTSTITQACPKVSFQPIPIH
 YCAPAGFALLKCNDKKFNGTGPCKNVSTVQCTHGIPVSTQLLNGLSLAE EEEI VIRSEN
 FTNNAKTIIVQLNESVINCTRPNNNTRKSIHIGPGRFYATGDIIGDIRKAHCNISRAN
 WNNTLRQIVEKLGEQFG-NKTIVFNQSSGGDVEIVMHSFNCGGEFFYC NSTQLFNSTWNF
 --NGTWNKN---FNNTWNNTTEGTNDTITLPCRIKQIINMWQGVGKAMYAPPISGQIRCSS
 NITGLILTRDGGNN-NET--NRTETFRPGGDMRDNRSELYKYKVVKIEPLGIAPT KAR
 RRVVQREKRAVG TIGAMFLGFLGTAGSTMGAASLT LTVQARLLLSGIVQQQNLLKAIEA
 QQHLLQLTVWGIKQLQARLLAVE RYLGDQQLGLWGC SGKLICTTTVPWNASWSNKS LDK
 IWDNMTWMEWEREIDNYTGLIYNLLEKSQNQQEKNELE LLELDKWANLWNWFDITKWLWY
 IKIFIMIIGGLIGLRIVFAVLSVVRVRQGYSPLSLQTRLPTQRGPDRPEGIAEEGGERD
 RDRSGPLVDGFLAIIWVDLRSCLFSYHHLRDL LLIVTRIVELLG-----RRGWEALK
 YWWNLLLYWSQELKNSAVNLLNTTAIAVAEGTDRIIEVLQRIYRAFLHIPRRIRQGFERA
 LL-

>ENV-B.syn4.3

MRVKEIRKQNYQHLWKWGTM L-----LGMLMICS AAGNLWVTVYYGVPVWKEANTTLF
 CASDAKAYETE VHNWATHACVPIDPNPQEVVLGNVTENFNMGKNNMVEQM HEDIISLWD
 ESLKPCVKLTPLCVTLNCTDELKNATFRSNTTTNSSW--EKMEKGEIKNCSFNITTNMRD
 KMQKEYALFYKLDVIPI-DSRNNSNNSTE--YNSYRLINCNTSVITQACPKISFEPIPIH
 YCTPAGFAILKCKDKKFNGKGPCTNVSTVQCTHGIRPVVSTQLLNGLSLAEKEVVIRSDN
 FTNNAKTIIVQLNVSVEINCTRPNNNTRRSIPIGPGRV FYTTEDIIGDIRQAHCNISRAQ
 WNNTLKHIVEKLKGQFGNNKTIIVFNHSSGGDPEIVMHSFNCRGEFFYCKSTKLFNSTWTR

27/109

Fig. 10 cont'd-2

N-NGTWTRN---DTERSNSTE---EHITLPCRIKQIINMWQEVGKAMYAPPIRGQIRCSS
 NITGILLTRDGGNDT-----SGTEIFRPGGDMKDNWRSELYKYKVVRIEPLGVAPTEAK
 RRVVQREKRAVG-IGAVFLGFLGAAGSTMGA AVTLTVQARLLLSGIVQQQNLLRAIEA
 QQRLQLTVWGIKQLQARVLAVERYLRDQQLLGIWGCSGKI ICTTAVPWNTSWSNRSLSNE
 IWNNMTWMEWEKEIDNYTNLIYNLLEESQNQQEKNEQELLALDKWANLWNWFDISNWLWY
 IKIFIIIVGGLVGLRIVFAVLSIVNRVRQGYSPLSFQTRLPA PRGPDRPEGIEEEGGERD
 RDRSVRLVDGFLALIWDLLRSLCLFSYHRLRDLLLI----VELLG-----RRGWEILK
 YWWNLLQYWQELKNSAVSLLNATAIAVAEGTDRVIEVVQRAYRAILHIPTRIRQGLERA
 LL-

>ENV-B.syn4.4

MRVKETRKNYQHLWRWGIML-----LGMLMICSATEKLWVTVYYGVPVWKEATTTLF
 CASDAKAYDKEVHNVWATHACVPTDPSPQEVVLENTENFNMWKNDMVEQMHEDIINLWD
 QSLKPCVRLTPLCVTLNCTN-VNVTNLKNETNTKSSSGGEKMEEGEMKNCSFNITTSIRD
 KVQKQYALFYKLDVPI-DNDNTS-----YRLISCNTSVIKQACPKVSFEPIPIH
 FCAPAGFAILKCNDKKFNGTGPCTNVSTVQCTHGIRPVISTQLLNGLSLAE EEVVIRSEN
 FTDNAKTIIVQLNETVEINCTRPSNNTRKSIPIGPGRFYTTGDIIGDIRQAYCNISRAK
 WNNTLKQIVTKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYC NSTKLFNSTWTW
 N-NSTW--N---NTKRSNDTE---EIITLPCRIKQIINRWQEVGKAMYAPPIEGQIRCLS
 NITGLLLTRDGGTNNT----NTNETFRPGGGMRDNWRSELYKYKV VQIEPLGVAPTAK
 RRVVQREKRAVG-IGALFLGFLGAAGSTMGAASITLTVQARQLLSGIVQQQNLLRAIEA
 QQHMLQLTVWGIKQLRARVLAVERYLRDQQLLGIWGCSGRLICTTNVPWNTSWSNKSLSNE
 IWDNMTWMQWEREIDNYTSLIYTLIEESQNQQEKNEQDLLALDKWASLWNWFSITNWLWY
 IRIFIMIVGGLVGLRIVFTVISIVTRVRQGYSPLSFQTRLPTPRGPDRPEGTEEEGGERD
 RDRSGRLVDGFLALFWDDLRLSLCLFLYHRLRDLLLIAARIVEL LG-----RRGWELLK
 YWWNLLQYWIQELKNSAVSLFNAIAIAVAEGTDWVIEISQRAFR AVLHIPVRIRQGLERA
 LQ-

>ENV-B.syn6.1

MRVTGIRKNYQHLWRWGTMLLWRWGTMLLGILMICSAAAGKLWVTVYYGVPVWKDATTTLF
 CASDAKAYDTEAHNVWATHACVPIDPNPQEVVLENTENFNAWKNNMVEQMHEDMISLWD
 QSLQPCVRLTPLCVTLNCTDDVRN-----ATSTNSSW-GKPM EKGEIKNCSFNITTSIRD
 KVQKQYALFYKLDVPI-DNDSNNTN-----YRLISCNTSIIITQACPKITFEPIPIH
 YCTPAGFALLKCNDKKFNGTGPCTKVSTVQCTHGIRPVVSTHLLNGLSLAE EEVVIRSEN
 FTNNAKTIMVQLNVSV EINCTRPSNNTRKSIHIGPGRFYTTGDIIGDIRKAHCNISRA N
 WNNTLRQIVEKLGEQFGNNKTIVFNHSSGGDLEIVTHSFICGGEFFYC NSTKLFNSTWTW
 N-NSTW--N---NTKRSNDTE---EIITLPCRIKQIINMWQEVGKAMYAPPIRGKIRCSS
 NITGLLLTRDGGTNNT----NTNETFRPGGDMRDNWRNELYKYKVVRIEPLGIAPTEAK
 RRVVQREKRAVG-IGAMFLGFLGTAGSTMGAASVALTVQARQLLP GIVQQQNLLRAIDA
 QQHLLQLTVWGIKQLQARILAVERYLRDQQLLGFWGCSGKLICTTNVPWNTSWSNKSYSQ
 IWENMTWMEWEREINNYTGLIYNLLEKSQNQQEKNEQELLELDKWASLSWFDISNWLWY
 IKIFIIIVGGLVGLRIVFAVLSIINRVVRQGYSPLSFQTHLPAPRGPD RPEGIAEEGGERD
 RDRSGRLVNGFLALIWVDLRLSLCLFSYHHLRDLLLI----VEL LG-----RRGWEVLK
 YWWNLLLYWSQELKNSAISLLNATAVAVAEGTDRVIEALQRA CRAILHIPRRIRQGLERL
 LL-

>ENV-B.syn6.2

MRVKETRKNYQHLWKWGTM-----LGILMICSATENLWVTVYYGVPVWKEATTTLF
 CASDAKAYDKEVHNVWATHACVPTDPNPQEVELKNVTENFNMWKNNMVEQMQEDIISLWD
 QSLKPCVRLTPLCVTLNCTD-LRNATNGNDTNTTSSS-REMMGGGEMKNCSFNITTNIRD
 KVQKEYALFYKLDIVPI-DNDNTN-----YRLISCNTSVVTQACPKVSFEPIPIH
 FCAPAGFAILKCNDKKFNGKGPCTNVSTVQCTHGIRPVISTQLLNGLSLAE EEVVIRSEN

28/109

Fig. 10 cont'd-3

FTNNVKTIIIVQLNETVEINCTRPNNNTRRSIPIGPGRVFYTTEDIIGDIRQAHCNLSRTQ
 WNNTLKQIVTKLREQFG-NKTIIIFNQSSGGDPEIVMHTFNCGGEFFYCNTTKLFNSTW--
 --NDTTINR----TEGSNNTR----NITLPCRIKQIINLWQEVGKAMYAPPIQGGQISCSS
 NITGLLLTRDGGNN-NET--NRTETFRPGGGMNRDNWRSELYKYKVVKIEPLGVAPTAK
 RRVVQREKRAVG-IGALFLGFLGAAGSTMGAASVTLTVQARQLLSGIVQQRNLLRAIEA
 QQRMQLTLTVWGIKQLRARVLAVERYLKDQQLMGIWGCSGKLICTTTVPWNASWSNKSLE
 IWDNMTWMEWEREIDNYTSLIYTLIEESQNQQEKNELELLELDKASLWNWFSITNWLWY
 IRLFIMIVGGLVGLRIVFTVISIVTRVRQGYSPLSFQTRLPTPRGPDRPGGIEEGGEQD
 RDRSIRLVDGFLALIWDRLSLCLFSYHRLRDLWI----VELLG-----RRGWEALK
 YLWNLLQYWSQELKKSASVSLFNATAIAVAEGTDWVIEVIQRAFRAFIHIPTRVRQGLERA
 LQ-

>ENV-B.syn6.3

MRVKGIRKNCQHLWRWGILL-----LGMLMICSATEKLWVTVYYGVPVWKETTTTLF
 CASDAKAYVAEKHNWATHACVPTDPNPREVVMGNVTEEFNIWNNSMVEQMHEDIISLWE
 QSLKPCVKLTPLCVSLKCTDL-----KNDTNTNSSSGRMIMEKGEIKNCSFNITTGIRG
 KVQ-EYSLFYKLDVVQM-DEDNTS-----YRLINCNTSVITQACPKVSFQPIPIH
 YCAPAGFAILKCKDKKFNGTGSCKNVSTVQCTHGIRPVISTQLLNGSLAEGEVVIRSEN
 FTDNAKTIIIVQLKDPVKINCTRPNNNTRKSIPIGPGRAFYATGDIIGDIRQAHCNISTTK
 WNKTLGQVVKKLREQFK-NKTIVFKQSSGGDPEVVMHSFNCGGEFFYCNTSOLFNSTW--
 -----N---STSLFNSTN---GTITLQCRIKQIINRWQEVGKAMYAPPIEQIRCLS
 NITGLLLVRDGGINVTNN--TGTEVFRPGGDMRDNRSELYKYKVIEPLGVAPTRAK
 RRVVQREKRAVG-LGAMFLGFLGAAGSTMGAASITLTVQARQLLSGIVQQRNLLRAIEA
 QQHMLQLTVWGIKQLQARVLAVERYLQDQQLGIWGCSGKLICTTTVPWNWSNKSLE
 IWDNMTWMEWEKEIDNYTGLIYTLLEESQNQQEKNEHELLELDKASLWNWFNITNWLWY
 IKIFIMIIGGLIGLRIVFAVLSIVNRVRQGYSPISFQTRLPAAPRGPDPRPDGIEEGGDRD
 RDRSGRLVDGFLTLIWDRLSLCLFSYRRLRDLLLIAARIVELG-----HRGWEALK
 YWWNLLQYWIQELKNSAVNLLNTTAIAVAEGTDRVIEVVQRAYRAILNIPTRIRQGFERA
 LL-

>ENV-B.syn6.4

MRVKEIRKNCQRLWRWGTM-----LGMLMICSAAEQLWVTVYYGVPVWRDANATLF
 CASDAKAYDTEVHNWATHASVPTDPNPQEVVLGNVTENFNMWKNDMVEQMHEDVISLWD
 QSLKPCVKLTPICVTLNCTDYVKNT-NNATSTNSSW-GEPMEKGEIKNCSFNITTSMD
 KVQKTYALFYKLDVPI-DNDSNNNDSTNTNYTNRLISCNTSVIKQACPKVSFDPPIPIH
 YCTPAGFAILKCRDKKFNGTGPKNVSTVQCTHGIRPVVPTQLLNGSLAEIEIVIRSEN
 FSDNAKTIIIVHLNESVEINCTRLNNNTRKSIHMGPGRAFYATGEIIGDIRQAHCNISRAK
 WNNTLKQIAIKLREQFGNKTIIIFKQSSGGDPEIVTHSFNCGGEFFYCNTSOLFNSTWNF
 --NGTWNKN---FNNTWNTEGTNDTITLPCRIKQIINMWQKVGKAMYAPPIISGQIRCTS
 NITGLLLTRDGGN---DT--SGTEIFRPGGGMKDNWRSELYKYKVQIEPLGVAPTEAK
 RRVVQREKRAVG-IGAVFLGFLGAAGSTMGAASVTLTVQARLLLSGIVQQNNLLKAIEA
 QQHLLRLTLVWGIKQLQARLLAVERYLGDQQLGLWGCSGKLICTTAVPWNWSNRSLE
 IWNMTWMEWEKEIDNYTNLIYNLLEESQNQQEKNEKELLELDKWANLWNWFDISNWLWY
 IRIFIMIVGGLIGLRIVFIVLSVNVNRVRQGYSPSLQTRLPTQRPDRPEGTEEEGGERD
 RDTSGRLVDGFLAIWVDRLSLFLFSYHRLRDLILLIVTRIVELG-----RRGWEILK
 YWWNLLQYWGQELKNSAVSLNATAITVAEGTDRVIEVLQRAGRAILHIPTRIRQGLERI
 LL-

>ENV-B.syn6.5

MRVKGIRRNQHLWRWGIML-----LGMLMICSATEQLWVTVYYGVPVWKEANTTLF
 CASDAKAYKTEAHNVWATHASVPTDPNPQEIIVLENTENFNMWKNNMAEQMHEDIINLWD
 QSLKPCVELTPLCVTLNCTDELKNATFRSNTTTNSSW--EKMEKGEIKNCSFNVTTSIRD
 KMQKEYALFYRLDVPI-DNDNTS-----YRLISCNTSVITQACPKISFEPIPIH
 YCVPAGFAILKCNKTFNGTGPCNTVSTVQCTHGIPVSTQLLNGSLAEEDVVIRSEN
 FTDNTKTIIIVQLKEAVEINCTRPNNNTRKGIHIGPGRAFYTTGEIIGNIRQAHCNLSRAK
 WNDTLKQIVIKLREQFG-NKTIVFNQSSGGDVEIVMHSFNCGGEFFYCNTTOLFNSTW--
 --NANDIRN---VTRGSNRTTGGNDTLILPCRIKQIVNMWQEVGKAMYAPPIKQIKCSS

29/109

Fig. 10 cont'd-4

NITGLLLTRDGGNNGNET--NGTEIFRPGGGDMRNNNRSELYKYKVVRIEPLGVAPTAKR
 RRVVQREKRAVT-LGAMFLGFLGAAGSTMGAASMTLTQARLLLSGIVQQQSNLLRAIEA
 QQRLQLQTVWGIKQLQARILAIERYLKDQQLLGIWGCSGKI ICTTAVPWNASWSNKSQDE
 IWNNMTWMQWEREIDNYTGLIYNLIEESQNQQEKNEQELLALDKWANLWNWFDITKWLWY
 IKIFIMIVGGLIGLRIVFTVLSIVNRVRKGYSPLSFQTRLPAQRGPDRPEGIEEGGERD
 RDRSGPLVDGFLAIFWVDLRSFLFSYHRLRDLLIVARIVELLG-----RRGWELLK
 YWWNLLQYWSQELKSSAVSLLNATAIAVAEGTDRIEVLQRAYRAILHIPVRIRQGLERA
 LL-

>ENV-B.syn6.6

MRVKGIRKNYQHLWRWGMML-----FGMLMICSAGNLWVTVYYGVPVWREATTTLF
 CASDAKAYETEVEHNVWATHACVPTDPSPEVLENTEDFNMWKNNMVDQMHEIDIISLWD
 ESLKPCVKLTPLCVTLNCTDDVRNVT--NNATNTNSSW-GEPMEKGEIKFCSFNITTSIRN
 KVQKQYALFYKLDVIPI-DSRNNNSNNSTE--YNSYRLINCNSSTITQACPKVTFEPIPIH
 YCAPAGFAILKCNKKFNGTGPCNNVSTVQCTHGIRPVVSTQLLNGSLAEKEVVIRSDN
 FTNNAKTIIVQLNESVVINCTRPNNNTRKRISMGPGRVYTTGEIIGDIRRAHCNISRAQ
 WNNTLKHIVEKLGKQFGNNKTI-FNHSSGGDPEIVMHSFNCRGEFFYCKSTKLFNSTWTR
 N-NGTWTRN---DTERSNSTE---EHITLPCRICKIINMWQGVGKAMYAPPIRGQIRCSS
 NITGLILTRDGGNNDT----RGTEIFRPGGGDMKDNWRSELYRYKVVKIEPLGIAPTAKK
 RRVVQREKRAVGTIGAMFLGFLGTAGSTMGAASLTQARLLLSGIVQQQNNLLRAIEA
 QQHLLQLQTVWGIKQLQAKVLAVERYLRDQQLLGIWGCSGRLICTTNVPWNASWSNKS LDK
 IWNNMTWMEWDREINNYTSLIYSLIEESQNQQEKNEQDLLALDKWASLWNWFDITNWLWY
 IKIFIMVVGGLVGLRIIFAVLSIVNKVRQGYSPSLQTHLPARRGPDRPEGIEEGGERD
 RDRSVRLVDGFLALFWDDLRLSLCLFLYHRLRDLLIVTRTVELLG-----RRGWEALK
 YCWNLLQYWSQELKNSAVSLFNAIAIAVAEGTDRIIEVVQRICRAIRHIPRRIRQGFERA
 LL-

>ENV-C.syn1.1

MRVRGIQRNWPQWWIWGILG-----FWMLMICNVVGNLWVTVYYGVPVWKEAKTTTLF
 CASDAKAYEKEVEHNVWATHACVPTDPNPQEIVLENTENFNMWKNDMVDQMHEIDIISLWD
 QSLKPCVKLTPLCVTLNCTDVKNATSNGTTTYNNSI-DS--MNGEIKNCSFNTTTEIRD
 KKQKVYALFYRLDIVPL-DNNSSE-----YRLINCNTSTITQACPKVSFDPIPIH
 YCAPAGYAILKCNKTFNGTGPCNNVSTVQCTHGKIPVVSTQLLNGSLAEKEEIIIRSEN
 LTNNAKTIIVHLNESVEIVCTRPNNNTRKSIRIGPGQTFYATGDIIGDIRQAHCNISEKQ
 WDQTLRYVSEKLKEHFP-NKTIKFAPSSGGDLEITTHSFNCRGEFFYCNTSKLFNSTY--
 --NSTQMHN---DTGS--NST-----ITLPCRICKIINMWQEVGRAMYAPPIAGNITCKS
 NITGLLLTRDGGTNN-----NNTETFRPGGGDMRDNRSELYKYKVVEIKPLGIAPTAKK
 RRVVEREKRAVG-IGAVFLGFLGAAGSTMGAASITLTQARQLLSGIVQQQSNLLRAIEA
 QQHMLQLQTVWGIKQLQTRVLAIERYLKDQQLLGIWGCSGKLICTTAVPWNSSWSNKSQTD
 IWDNMTWMQWDREISNYTDTIYRLLEDSONQQEKNEKD LLDLSDWKNLWNWFDITNWLWY
 IKIFIMIVGGLIGLRIFAVLSIVNRVRQGYSPSLFQTLTPNPRGPDRLGRIEEEGGEQD
 RDRSIRLVSGFLALAWDDLRLSLCLFSYHRLRDFILVTARAVELLGRSSLRGLQRGWALK
 YLGS LVQYWGLELKKSAISLLDTIAIAVAEGTDRIIELIQRICRAIRNIPRRIRQGFEEA
 LL-

>ENV-C.syn3.1

MRVMGIQRNCQQWWIWGSLG-----FWMLMIYNVMGNLWVTVYYGVPVWKEAKTTTLF
 CASDAKAYDTEVEHNVWATYACVPTDPNPQEMVLENTENFNMWKNNMVDQMHEIDIISLWD
 QSLKPCVKMTPLCVTLNCSNAKKD-----NTTI-DNE-MKGEIKNCSFNITTEL RD
 KKKQVYALFYKLDIVPL-NSNSSE-----YRLINCNTSAITQACPKVSFDPIPIH
 YCAPAGYAILKCNNETFNGTGPCNNVSTVQCTHGIRPVVSTQLLNGSLAEKEEIIIRSEN
 LTDNVKTIIVHLNESVEINCTRPNNNTRRSIRIGPGQAFYATGEIIGDIRQAYCNISGEK
 WNETLQRVGKKLKEHFP-NKTIKFAPSSGGDLEITTHSFNCRRFFYCNTSGLFNGTY--
 --NGNGTYN---GTGTDNST-----ITIPCRICKIINMWQEVGRAMYAPPIEGNITCKS
 NITGLLLVRDGGTENNTET--NNTETFRPGGGDMRDNRSELYRYRVVEIKPLGIAPTAKK
 RRVVERGKRAVG-IGAVFLGFLGVAGSTMGAASITLTQARQVLSGIVQQQSNLLRAIEA
 QQHLLQLQTVWGIKQLQTRVLAIERYLKDQQLLGIWGYSGKLICTTAVPWNSSWSNRSQED

30/109

Fig. 10 cont'd-5

IWNNMTWMQWDREINNYTNTIYRLLEDSQNQQEKNEQDLLALDSWKNLWNWFDITNWLWY
IRIFIMIVGGLIGLRIIFAVLSIVNRVRQGYSPSLQTLTPNPRELDRLGRIEEGGGEQD
RDRSIRLVSGFLALAWDDLRLSLCLFSYHRLRDFILIAARAAELLGRSSLKGLQRGWEILK
YLGSLIQYWGLELKKSAINLLDTIAIVVAEGTDRIIELIQRICRAICNIPRRIRQGFEEA
LQ-

>ENV-C.syn3.2

MRVRGILRNWQQWWIWGILG-----FWMLMICNVVGNLWVTVYYGVPVWREAKTTLF
CASDAKAYEREVHNWATHACVPTDPNPQELVLENTENFNMWKNDMVDQMHQDIISLWD
ESLKPCVKLTPLCVTLKCVNVTKNS--NATVNSNATV--NNNTMGEEIKNCSFNATTEIRD
KKQNVYALFYRLDIVPL--NENNDNSS-----YRLINCNTSTITQACPKVTFDPIPIH
YCTPAGYAILKCNKTFNGTGPCNVSTVQCTHGIKPVISTQLLNGSLAEEIIIRSEN
LTNNVKTIIIVHLNKSVEIVCTRPGNNTRKSVRIGPGQTFYATGDIIGDIRQAHCNISRTA
WNKTLQEVGKKLAHF--NKTIEFKPSSGGDLEVTTHSFNCRGEFFCYNTSKLFNSTYNS
TYNSTYNSN---STNSNSNST-----ITLQCRIKQIINMWQKVGRAIYAPPIAGNITCRS
NITGLLLTRDGGNNNDTGNNNDTEIFRPGGGDMKDNWRNELYKYKVVEVKPLGIAPTGA
RRVVEREKRAVG--LGAVFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQQNNLLRAIEA
QQHMQVTVWGVIKQLQARVLALERYLKDQQLLGLWGCSGKLICTTNVPWNSSWSNKS LTD
IWENMTWMQWDKEISNYTDTIYRLLEVSQNQQEKNEKDLLALDSWNNLWNWFSITKWLWY
IKIFIMIVGGLIGLRIIFGVLSIVKRVVRQGYSPSLFQTLTPNPRGPDRLGRIEEGGGEQD
KDRSIRLVNGFLALAWDDLRLNLCFLSYHQLRDFILIVARAVELLGHSSLRGLQRGWEALK
YLGSLVQYWGLELKRSAISLLDTTIAIVAEGTDRIIEVIQRICRAIRNIPTIRIRQGFEEA
LLQ

>ENV-C.syn3.3

MRVRGIQRNWPQWWIWGILG-----FWMIICRVVGNLWVTVYYGVPVWTEAKATLF
CASDAKAYEKEVHNWATHSCVPTDPNPQEIIVLGNVTENFNMWENDMVDQMHEDVISLWD
QSLKPCVKLTPLCVTLNCT-----NANVTVNATSDGS--IKEEIKNCSFNNTTTEIRD
KKQKVYALFYRPDIVPLSGSNSSE-----YILINCNTSTVTQACPKVSFEPIPIH
YCAPASYAILKCNKTFNGTGPCQNVSTVQCTHGIKPVVSTQLLNGSLAEGEIIIRSEN
LTNNAKTIIIVHLNESIEIVCTRPNNNTRKSIRIGPGQTFATGDIIGNIRQAHCNISEEK
WNKTLQEVSRKLEHFP--NKTIIIFNSSSGGDLEITTHSFNCGGEFFCYNTTKLFNDS---
-----ALSAFNKTS--NETITLPCRIKQIINMWQGVGRAMYAPPIAGNITCNS
SITGLLLTRDGGT-----NNTTEIFRPGGGNMKDNWRSELYKYKVVEIKPLGVAPTEAK
RRVVEREKRAVG--IGAVLLGFLGAAGSTMGAASITLTAQARQLLSGIVQQQSNLLKAIEA
QQHMLQLTVWGVIKQLQARVLAIERYLQDQQLLGIWGCSGKLICTTTVPWNSSWSNKSQTD
IWDNMTWMQWDREISNYTNTIYRLLEESQNQQEQNEKDLLALDKWQNLWSWFSITNWLWY
IKIFIIIVGGLIGLRIILGVLSIVRRVRQGYSPSLFQTLIPNPRGPDRLGGIEEGGEQD
RDRSVRLVSGFLSLAWDDLRLSLCLFCYHRLRDFILVTARAVELLGRSSLRGLQKGWEALK
YLGSLVQYWGLELKKSAISLLDTTIAIVAEGTDRIIEFIQRICRAIRNIPTIRIRQGLEAA
LQ-

>ENV-C.syn4.1

MRVRGILRNYQQWWIWGSLG-----FWMLMIYNVGGNLWVTVYYGVPVWKEAKTTLF
CASDAKAYDTEVHNWATHACVPTDPDPQEIIVLENTENFNMWENDMVDQMHEDIISLWD
ESLKPCVKLTPLCVTLKCTNVTST---GNTTRGNNTS--EN---REEMKNCSFNNTTTEIRD
KKQKVYALFYKPDVPL--KENSSE-----YILINCNTSTVTQACPKVSFDPIPIH
YCAPAGFAILKCNKTFNGTGPCNNVSTVQCTHGIKPVVSTQLLNGSLAEEIIIRSEN
LTDNAKTIIIVHLNESIEIVCTRPGNNTRKSIRIGPGQAFYATGDIIGDIRQAYCNISKAT
WNKTLQEVGKELAKHFP--NKTINFNSSSGGDLEITTHSFNCGGEFFCYNTTKLFNNSL--
-----LNNTADNST---STITLQCRIKQIINMWQGVQAMYAPPIAGNITCKS
NITGLLLLRDGGDTST---NGTEIFRPGGGNMKDNWRSELYKYKVVEVKPLGIAPTGA
RRVVEREKRAVG--IGAVLLGFLGAAGSTMGAASITLTAQARQVLSGTVQQQSNLLRAVEA
QQHMLQLTVWGVIKQLQTRVLAIERYLKDQQLLGIWGCSGKLICTTNVPWNSSWSNKSQEE
IWENMTWMQWDREISNYTGTIYRLLEESQNQQEKNEQDLLALDSWKNLWNWFDISNWLWY
IKIFIIIVGGLIGLRIIFGVLSIVKRVVRQGYSPSLFQTLIPNPRGPDRLERIEEGGEQD
RGRSIRLVSGFLAIAWDDLRLSLCLFSYHQLRDFILIAVARAVELLGHSSLRGLQRGWEALK
YLGSLVQYWGLELKRSAISLLDTTIAIVAEGTDRIIEFIQRICRAIRNIPTIRIRQGFEEA
LQ-

31/109

Fig. 10 cont'd-6

>ENV-C.syn4.2

MRVMGIQRNCQQWWIWGILG-----FWILMICNVMGNLWVTVYYGVPVWKEAKATLF
 CASDAKAYEKEVHNIWATHACVPTDPNPQELVLENTENFNMWDNDMDVQMHQDIISLWD
 QSLKPCVKLAPLCVTLNCTNATVTATRNQSDIMNTTS-ND----GEMKNCSFNVTTTELRD
 KKKKEYALFYRLDIVPL-NEGSGNANQNNSNYSYRLINCNSTSAITQACPKVTFDPIPIH
 YCTPAGYAILKCNDFNGTGPCNVSTVQCTHGIRPVVSTQLLNGSLAEGEIMIRSEN
 LTNNAKTIIVHLNKSVEIVCTRPNNNTRKSVRIGPGQTFYATNDIIGDIRQAHCNISEEK
 WNKTLLQQVGKKLAEHFP-NKTIEFKPSSGGDLEVTTHSFNCRGEFFYCNTSGLFNGTF--
 --DGT-----ESNSTSNAT-----ITIPCRIKQIINMWQKVGRAIYAPPIAGNITCRS
 NITGLLLVRDGGNDNKT---NDTETFRPGGGDMRDNRSELYKYKVVEVKPLGVAPTAK
 RRVVQREKRAVG-IGAVFLGFLGVAGSTMGAASMTLTVQARQVLSGIVQQQSNLLRAIEA
 QQHLLQLTVWGILKQLQARVLALERYLRDQQLLGMWGC SGKLICTTAVPWNSSWSNKSQED
 IWGNMTWMQWDKEISNYTNTIYRLLEDSQNQQERNEKDLLALDSWKNLWSWFDITNLWLY
 IKIFIMIIGGLIGLRIFAVLSIVNRVRQGYSPSLQTLTPNPRGPDRLGRIEEEGGEQD
 KDRSIRLVNGFLALAWDDLRLNCLFSYHRLRDFILIVARAVELLGRNSLRGLQRGWETLK
 YLGSLIQYWGLELKKS AISLLDTTAIAVAEGTDRIIELIQRICRAICNIPRRIRQGLEAA
 LQ-

>ENV-C.syn4.3

MRVRGIQRNWPQWWIWGILG-----FWMIIICRVVGNLWVTVYYGVPVWREAKTTLF
 CASNAKAYEKEVHNVWATHACVPTDPNPQEIIVLGNVTENFNMWKNDMDVQMHEDVISLWD
 QSLKPCVKMTPLCVTLNCTDVKNATSNGTTTYNNNSI-DS--MNGEIKNCSFNTTTELRD
 KKQKAYALFYRPDIVPLPGKDNKDNSEYEE--YILINCNSSTITQACPKVSFEPIPIH
 YCAPASYAILKCNNETFNGTGPCKNVSTVQCTHGKIPVISTQLLNGSLAEKEIIIRSEN
 LTNNVKTIIIVHLKESVEINCTRPNNNTRKSIRIGPGQTFYATGDIIGNIRQAHCNISREK
 WNTTLKRVKEKLKEHFP-NKTIKFAPSSGGDLEITHTTFNCRGEFFYCNTSKLFNSTYV--
 --NRTDMND---D--TGNNST-----ITLPCRIKQIINMWQEVGRAMYAPPIAGNITCNS
 SITGLLLTRDGGNNT-----ENTETFRPGGGNMKDNWRNELYKYKVVEIKPLGVAPTEAK
 RRVVEREKRAVG-LGAVFLGFLGAAGSTMGAASITLTVQARQLLSGIVQQQNNLLRAIEA
 QQHMLQLAVWGILKQLQARVLAIERYLQDQQLLGIWGC SGKLICTTSVPWNSSWSNRSQED
 IWNMTWMQWDREISNYTDTIYRLLEVSQNQQEQNEKDLLALDKWQNLWSWFSITNLWLY
 IRIFIMIVGGLIGLRIVFAVLSLVNRVRQGYSPSLFQTLTPSPRGPDRLGRIEEEGGEQD
 RDRSIRLVSGFLSLAWDDLRLSLCLFSYHRLRDFILIAARAAELLGRSSLRGLQRGWEILK
 YLGSLAQYWGLELKKSAINLLDTIAIAVAEGTDRIIEVIQRICRAIYNIPRRIRQGFEEAS
 LL-

>ENV-C.syn4.4

MRVRGIPRNWQQWWIWGILG-----FWMLMICNVVGNLWVTVYYGVPVWTEAKTTLF
 CASDAKAYEREVHNVWATYACVPTDPNPQEMVLENTENFNMWKNDMVEQMHEDIISLWD
 QGLKPCVKLTPLCVTLNCSNAKDD-----NTTI-DNE-MKGEIKNCSFNITTEL RD
 KKQQVYALFYKLDIVPL-NSNSSE-----YRLINCNTSTITQACPKVNFDPPIPIH
 YCAPAGYAILKCNKTFNGTGPCQNVSTVQCTHRIKPVVSTQLLINGS LAEGEIIIRSEN
 LTDNVKTIIIVHLNESVEIVCTRPNNNTRKSMRIGPGQTFYATGEIIGDIRQAHCNISKEK
 WNTTLQEVREKLREHFP-NKTIKFAPHSGGDPEITTHSFNCRGEFFYCNTSQLFNSTY--
 --NSTQMHN---DTGS--NST-----ITLPCKIKQIINMWQGVGRAMYAPPIEGNITCTS
 NITGLLLTRDGGT-----NNTIEFRPGGGDMRNNNRSELYKYKVVEIKPLGIAPTAK
 RRVVERGKRAVG-IGAVLFGFLGAAGSTMGAASIALTAQARQLLSGIVQQQSNLLKAIEA
 QQHMWQVTVWGILKQLQARVLAMERYLKDQQLLGLWGCSGKLICTTTVPWNSSWSNKSQTD
 IWDNMTWMQWDREINNYTNTIYKLEDSQNQQEKNEKDLLALDSWNNLWNWFSITKWLY
 IKIFIMIVGGLIGLRILGVLSIVRRVRQGYSPSLFQTLTPNPRELDRLGRIEEEGGEQD
 RDRSVRLVSGFLALAWDDLRLSLCLFCYHRLRDFILVTARAVELLGRSSLKGLQRGWEALK
 YLGNLVQYWGLELKKS AISLFDITAITVAEGTDRIIELVQRICRAIRNIPRRIRQGFEEA
 LL-

>ENV-C.syn6.1

MRVRGIQRNWPQWWIWGILG-----FWIIIMCRMVGMNMWVTVYYGVPVWREAKTTLF
 CASDAKGYEKEVHNAWATHACVPTGPNPQEMVLENTENFNMWKNNMVDQMHEDIINLWD
 QSLKPCVRLTPLCVTLKCVNVTKNS--NATVNSNATV--NNNTMGEEIKNCSFNATTEIRD
 KKQKAYALFYRPDIVPL-NENSSSENNSSE----YILINCNTSTITQACPKVSFDPIPIH
 YCAPASYAILKCNNETFNGTGPCQNVSTVQCTHGKIPVISTQLLNGSLAEEDIIIRSEN

32/109

Fig. 10 cont'd-7

LTNNAKTIIVHLNQSVEIVCTRPGNNTRKSMRIGPGQTFYATNDIIGNIRQAHCNISSEGK
 WNETLLRVKKKLEEHFP-NKTIKFEPSSGGDLEITTHTFNCRGEFFYCDTSTLFNHTY--
 --VSAYMNTDVSADRKNDTQ-SNSTITLPCRIRQIINMWQEVGRAIYAPPIAGNITCRS
 NITGLLLVRDGGNTT-----NSTETFRPEGGMKDNWRSELYKYKVVEIRPLGIAPTGA
 KRRVVEREKRAVG-IGAVFLGFLGVAGSTMGAASMTLTVQARQVLSGVVQQQSNLLQAI
 EAQQHLLQLTVWGIKQLQTRVLALERYLRDQQLLGIWGC SGKI ICTTAVPWNTSWSNKSQ
 EDIWNMTWMQWDREINNYTNTIYKLLSESONQQEKNEQDLLALDSWNSLWNWFSITKWLWY
 IRIFIIIVGSLIGLRIFGVLSIVKVRVQGYSPLLSQTLTPNPREDRLGRIEEGGGEQD
 RDRSVRLVNGFLALVWDDLRLSLCLFCYHRLRDFILVTARVVELLGRSSLRGLQKGWEAL
 KYLGLSVQYWGLELKKSAINLLDTIAIAVGEQTDRIIEVIQRICRAIYNIPRRIRQGF
 EASLL-

>ENV-C.syn6.2

MRVRGILRNYQQWWIWGSLG-----FWMLMIYNVGGNLWVTVYYGVPVWTDAKTTLF
 CASDAKAYDKEVHNVWATHACVPTDPNPQELVLENTENFNMWKNDMVNQMHEDIISLWD
 ESLKPCVKLTPLCVTLNCTNATVTATRNGSDIMNTS-ND----GEMKNCSFNITTEL
 RDKKRKEYALFYRLDIVPL-DENNSSEKSSSENSSEYYRLINCNTSAITQACPKVTFDPI
 PLHYCAPAGYAILKCKDKTFNGTGPCSNVSTVQCTHGIKPVVSTRLLNGLAEGEIIIRSEN
 LTNNVKTIIIVHLKEPVEINCTRPNNNTRESIRIGPGQTFYATGDIIGDIRQAHCNISREK
 WNKTQLQEVGKKLAEHFP-NKTIKFAPHSGGDLEITMHSFNCRGEFFYCNTSGLFN
 GTY---MPTYMPN---GTESNSNST-----ITIPCRIRQIINMWQEVGRAMYAPPIE
 GNITCNSNITGLLLVRDGGINKT-----NNTETFRPGGDMRNNWRSELYKYKVVEIK
 PLGVAPTEAKRRVVEREKRA-A-LGAMFLGFLGAAGSNMGAASITLTAQARQLLSGIV
 QQRSNLLRAIEAQQHLLQLTVWGVKQLQARVLAMERYLKDQQLLGLWGCSGKLICTTS
 VPWNSSWSNRSQEEIWNMTWMQWDREISNYTNTIYRLLEDSQONQQEKNEKDLLALDS
 WKNLWSWFDITNWLWYIKIFIMIIGGLIGLRIVFAVLSIVNRVRQGYSPLSFQTLTPS
 PRGPDRLGRIEEEGGEQDKDRSVRLVSGFLSLAWDDLRLSLCLFSYHRLRDLILIAA
 RAVELLGHSSLRGLQRGWEILKYLGSLAQYWGLELKRSAISLLDTIAITVAEGTDRIE
 IIEIQRICRAICNIPRRIRQGFETALL-

>ENV-C.syn6.3

MRVMGILRNCQQWWIWGVLG-----FWMLMICNVVGNLWVTVYYGVPVWKEAKTTLF
 CASNAKAYEREVHNIWATHACVPTDPNPQEMVLKNVTENFNMWKNDMVDQMHEDVISLWD
 QSLKPCVKLAPLCVTLNCTNVTVNDTLHQNT-----DMKNCSFNVTTEL
 RDKKQKVYALFYRLDVVPL-GDNNSS-----YRLINCNTSTIAQACPKVNFDP
 IPIHIYCTPAGYAILKCNKDTFNGTGPCKNVSTVQCTHEIKPVVSTQLLNGLAE
 EGIIRSENLTDNAKTIIVHLNESVEINCTRPGNNTRQSIRIGPGQAFYATGAIIGD
 IRQAHCNISKDEWEKTLKRVSEKLKEHFP-NKTIEFKPSSGGDLEVTTHSFNCRREFF
 YCNTSKLFNSTY---NSTQMHN---DTGS---NST-----ITLPCKIKQIINMWQGV
 GQAMYAPPIKGNITCKSNITGILLTRDGGNLT-----NGTETFRPGGDMKDNWRSE
 LYRYRVVEIKPLGIAPTGAARRVVQREKRAVG-IGALFLGFLGTAGSTMGAASLT
 LTVQARQLLSIVQQQSNLLRAIEAQQHMLQLTIWGIKQLQTRVLAVERYLKDQQLL
 GMWGCSGKLICTTAVPWNASWSNKSQEEIWGNMTWMQWDREISNYTDIIYRLLEES
 QONQQERNEKDLLALDSWNNLWNWFNITNWLWYIKIFIMIVGGVIGLRIFAVLSL
 VNRVRQGYSPLSFQTLTPNPREDRLGRIEEEGGEQGRDRSIRLVNGFLAIAWDDL
 RSLCLFSYRRLRDFILIAARAAELLGRSSLRGLQRGWETLKYLGSLIQYWGLELKK
 SAISLFDITIAIAVAEGTDRIEELIQRICRAIRNIPRRIRQGLEALLQ-

>ENV-C.syn6.4

MRVMGIQRNCQQWWIWGILG-----FWMLMIYNVVGNLWVTIYYGVPVWKEAKATLF
 CASDAKAYDTEVHNVWATHACVPTDPDPQEMVLGNVTENFNMWKNDMADQMHEDIISLWD
 QGLKPCVKLTPLCVTLHCTN-----TNITNENRTI-GDKLNE-EMKNCSFN
 TTTTEL RDKKQQVYALFYKPDVVPL-NGGEHNETGE-----YILINCNSSTITQACPKV
 SFEP
 IPIHIYCAPAGFAILKCNKDTFNGTGPCNNVSTVQCTHGIRPVVSTQLLNGLAE
 EEEIIIRSENLTDNVKTIIIVHLNKSVEIVCTRPGNNTRKSIRIGPGQTFATNDIIGD
 IRQAYCNISAEKWNKTLEERVEEKLKEHFP-NKTIKFNSSSGGDLEITTHSFNCRGEFF
 YCNTSNLFNGTY---HGTQSTN---ST---NST-----ITLQCRIKQIINMWQKV
 GRAMYAPPIAGNITCKSNITGLLLRLDGGTEN-----NDTETFRPGGGM
 RDNWRSELYKYKVVEVKPLGIAPTTAKRRVVERDKRAVG-IGAVLLGFLGAAGST
 MGAASMTLTVQARQLLSGIVQQQSNLLRAVEA

33/109

Fig. 10 cont'd-8

QQHMLQLTVWGIKQLQARVLALERYLKDQQLLGIWGCSGRLICTTAVPWNSSWSNKTQGE
IWENMTWMQWDKEINNYTNTIYRLLEESQTQQEQNEKDLLALDSWKNLWNWFDITKWLWY
IKIFIMVVGGLIGLRIIFAVLSIVNSVRQGYSPSLQTLTPNPRGPDRLERIEEEEGGEQD
RNRSIRLVNGFLALAWDDLRLSLCLFSYHHLRDFILVTARAVELLGRSSLKGLQRGWEALK
YLGNLVQYWGLELKKSAISLLDTTAAIAVAEGTDRIIELVQRICRAILNIPTIRIQGFEEA
LQ-

>ENV-C.syn6.5

MRVRGIPRNWPQWWTWGILG-----FWMIIICRVVGNLWVTVYYGVPVWTEAKTTLF
CASDAKAYEREVHNWATHSCVPTDPNPQEIVLGNVTENFNMWENDMVDQMHDIIISLWD
QSLKPCVKMTPLCVTLNCSNAKDD-----NTTI-DNE-MKGEIKNCSFNITTEIRD
KKQKVHALFYRLDIVPL-NEGSGNANQNNNSYDYRLINCNTSTVTQACPKVTFDPIPIH
YCAPARYAILKCNNTFNNGTGPCNNVSTVQCTHGIKPVVSTQLLSGSLAEIEIVIRSEN
LTNNAKIIIVHLNESVEIVCTRPNNNTRRSIRIGPGQTFYATGEIIGDIRQAHCNISAKQ
WNTTLERVKEKLREHFP-NKTIKFEPHSGGDPEITTHSFNCGGEFFYCNTSQLFNSTY--
--NSTYMSN--NTGENSNET-----ITLPCRIKQIINMWQQVGRAMYAPPIAGNITCNS
SITGLLLTRDGGNNNDTGNNNDTEIFRPGGGDMRDNRSELYKYKVVELKPLGIAPTEAK
RRVVKREKRAVG-IGAVFLGFLGAAGSTMGAASIALTAQARQVLSGIVQQQNNLLRAIEA
QQHVLQLTVWGIKQLQTRVLAIERYLKDQQLLSLWGCSGKLICTTTPVWNSSWSNKS LTD
IWDNMTWMQWDREISNYTGTIYRLLEDSQSQQEKNEKDLELDKWNLLWNWFDISNWLWY
IKIFIIIVGGLIGLRIIFAVLSIINRVRQGYSPLLFQTLTPNPRGLDRLGRIEEEGGEQD
KDRSIRLVNGFLALAWEDLRLSLCLFSYHQLRDFILIVARAVELLG-----RRGWEALK
YLGNLVLVWGLELKKSAVSLLDTTAAIAVAGGTDRIIEVVQRICRAIRNIPTIRIQGLEAA
LL-

>ENV-C.syn6.6

MRVRGILRNWQQWWIWGILG-----FWMVMICNVMGNLWVTVYYGVPVWQEAKTTLF
CASDAKAYEKEVHNWATHACVPTDPSPQEIVLENVTENFNMWKNDMVEQMHDIIISIWD
QSLKPCVTLTPLCVTLNCTDVKNATSNGTTTNNNSI-DS--MNGEIKNCSFNITTTEIRD
KKKQVYALFYKLDIVPL-NSNSSE-----YRLINCNTSAVTQACPKVSWDPIPIH
YCAPAGYAILKCNKTFNGTGPCTNVSTVQCTHRIKPVVTTQLLNGSLAEKEIIIRSEN
LTNNIKTIIVHLNESIEIVCTRPNNNTRKSVRIGPGQTFATGDIIGDIRKAHCNISEDK
WNETLQRVGKKLVEHFP-NKTIKFAPSSGGDLEVTTHSFNCKGEFFYCNTTKLFD-----
-----DSERINTTT--TTIILPCRIKQFINMWQGVGRAMYAPPIAGNITCTS
NITGLLLTRDGGT-----NTEIFRPGGGNMKDNWRNELYKYKVVEVKPLGVAPTKAK
RRVVEREKRAVG-LGAVFLGFLGAAGSTMGAASITLTVQARQLLFGIVQQQSNLLKAIEA
QQHMWQVTVWGIKQLQARVLAIERYLQDQQLLGIWGCSGKLICTTNVPWNSSWSNKSQTD
IWDNMTWMQWDKEISNYTDTIYRLLEVSQNQQEENEKDLLALDKWQNLWNWFSITNWLWY
IRIFIMIVGGLIGLRIILGVLSIVRRVRQGYSPSLFQTLIPNPRGPDRLGIEEEEGGEQD
RDRSIRLVSGFLALAWDDLRLNCLFSYHRLRDFILIVVRAVELLGRNSLRGLQRGWEALK
YLGSLGQYWGLEIKKSAISLLDTIAIVVAEGTDRIIEFIQRFCAIRNLPRRIRIQGFEEA
LL-

>ENV-M.syn1.1

MRVTGIRKNYQHLWRWGTMLLWRWGTMLLGMMLICSAGNLWVTVYYGVPVWKEATTTTLF
CASDAKAYDTEVHNWATHACVPTDPNPQEVVLENVTENFNMWKNNMVEQMHDIIISLWD
QSLKPCVKLTPLCVTLNCTDDVRNVT-NNATNTNSSW-GEPMEKGEIKNCSFNITTSIRD
KVQKEYALFYKLDVVPID-NDNSNNTN-----YRLISCNTSVITQACPKVSFEPIPIH
YCAPAGFAILKCNKKFNGTGPCTNVSTVQCTHGIKPVVSTQLLNGSLAEIEEVIRSEN
FTNNAKTIIVQLNESVEINCTRPNNNTRKSIRIGPGQTFYATGDIIGDIRQAHCNISRAQ
WNNTLKHIVEKLGKQFGNNKTIVFNHSSGGDPEIVMHSFNCGGEFFYCNTTKLFNSTWTR
N-NGTWTRN---DTERSNSTE---EHITLPCRIKQIINMWQEVGKAMYAPPIRGQIRCSS
NITGLLLTRDGGNDT-----SGTEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTKAK
RRVVQREKRAVG-IGAVFLGFLGAAGSTMGAASITLTVQARQLLSGIVQQQNNLLRAIEA
QQHLLQLTVWGIKQLQARVLAVERYLKDQQLLGIWGCSGKLICTTAVPWNSSWSNKS LNE
IWNMTWMEWEREIDNYTGLIYTLIEESQNQQEKNEQELLELDKWLWNWFDISNWLWY
IKIFIMIVGGLIGLRIVFAVLSIVNRVRQGYSPSLFQTRLPA PRGPDRPEGIEEEEGGERD
RDRSIRLVSGFLALAWDDLRLSLCLFSYHRLRDLILLIVTRIVELLG-----RRGWEALK

34/109

Fig. 10 cont'd-9

YWWNLLQYWSQELKNSAVSLLNATAIAVAEGTDRVIEALQRACRAILHIPRRIRQGLERA
LL-

>ENV-M.syn3.1

MRVRGIQRNWPQWWIWGILG-----FWMLMICNVVGNLWVTVYYGVPVWKEAKTTLF
CASDAKAYEKEVHNWVATYACVPTDPNPQEIHLNVTEEFNMWKNDMVDQMHEDIISLWD
QSLKPCVQLTPLCVTLNCTN-VNVTNLKNETNTKSSSGGEKMEEGEMKNCSFNITTSIRD
KVQKEYALFYKLDVVI-DNDSNNTN-----YRLISCNTSVITQACPKVTFEPIPIH
YCTPAGFAILKCKDKKFNGTGPKNVSTVQCTHGIKPVISTQLLNGSLAEEEEIIIRSEN
ITNNAKTIIVQLNESVEINCTRPGNNTRKSVRIGPGQTFYATGEIIGDIRQAHCNLSRAK
WNDTLKQIVIKLREQFG-NKTIVFNQSSGGDPEIVMHSFNCGGEFFYCNTTQLFNSTW--
-----N---STSLFNSTN---GTITLQCRIKQIINMWQEVGKAMYAPPIEGNITCKS
NITGLLLVRDGGT---EP--NDTETFRPGGGMKDNWRSELYKYKVVKIEPLGVAPTAK
RRVVQREKRAVGTIGAMFLGFLGAAGSTMGAASMTLTVQARLLLSGIVQQNNLLRAIEA
QQHLLQLTVWGIKQLQARILAVERYLKDQQLLGLWGCSGKLICTTAVPWNWSWSNKSQTD
IWDNMTWMEWEREIDNYTGLIYTLIEESQNQQEKNEQELLELDKASLWNWFDITKWLWY
IKIFIMIVGGLVGLRIVFAVLSIVNRVRKGYSPLSFQTLTPNPRGPDRLGRIEEEGEGEQD
RDRSIRLVSGFLALAWDDLRLSLCLFSYHQLRDFILIVARAVELLGRSSLRGLQRGWEALK
YLGSLVQYWGLELKKSAISLLDTIAIAVAEGTDRIIEVIQIRICRAIRNIPRRIRQGFERA
LL-

>ENV-M.syn3.2

MRVTGIRKNYQHLWRWGTMLLWRWGTMLLGILMICSAAAGKLWVTVYYGVPVWRDAETTLF
CASDAKAHETEVEHNIWATHACVPTDPNPQEVVLGNVTENFNMWKNNMVEQMHEDIISLWD
ESLKPCVKLTPICVTLNCTDDVRNVT-NNATNTNSSW-GEPMEKGEIKNCSFNMTTEL RD
KKQKVHALFYKLDIVPL-NSNSSE-----YRLINCNTSAITQACPKVSFEPIPIH
YCAPAGFAILKCNDDKKFNGTGPCNTVSTVQCTHGIKPVVSTQLLNGSLAEEEVVIRSEN
FTNNAKTIMVQLNVSVEINCTRPNNNTRKSIHIGPGRAFYTGTGDIIGDIRQAHCNISRAQ
WNNTLKHIVEKLGKQFGNNKTIVFNHSSGGDPEITTHSFNCGGEFFYCNSTKLFNSTWTR
N-NGTWTRN---DTERSSTE---EHITLPCRIKQIVNMWQVRVGQAMYAPPIRGQIRCSS
NITGLLLTRDGGNDT-----SGTEIFRPGGDMRNNWRNELYKYKVVRIEPLGVAPTAK
RRVVEREKRAVG-LGAVFLGFLGAAGSTMGAASLTTLTVQARQVLSGIVQQQSNLLKAIEA
QQHLLKLTVWGIKQLQARVLAVERYLRDQQLLGIWGCSGKLICTTTVPWNASWSNKSLE
IWNMTWMEWEKEIDNYTNLIYNLLEESQNQQEKNEQDLLALDKWANLWNWFDISNWLWY
IKIFIIIVGGLIGLRIVFAVLSIINVRVQGYSPSLQTLIPNPRGPDRPGGIEEEEGEQG
RDRSIRLVNGFLALAWDDLRLNCLFSYHRLRDLILLIVTRIVELG-----RRGWEALK
YWWNLLQYWSQELKNSAISLLNATAVAVAEGTDRVIEALQRACRAILHIPRRIRQGLERL
LL-

>ENV-M.syn3.3

MRVKETRKNYQHLWKWGTML-----LGMLMICSAAEQWLWVTVYYGVPVWKEATTTTLF
CASDAKAYDTEVHNWVATHACVPTDPSPQEVVLENVTENFNMWKNDMVEQMHTDIISLWD
QSLKPCVKLTPLCVTLNCTDYVKNT-NNATSTNSSW-GKPMKGEIKNCSFNITTSIRN
KVQKQYALFYKLDIVPI-DNDNTS-----YRLINCNTSTITQACPKVSFDPIPIH
YCAPAGYAILKCNKTFNGTGPCNNVSTVQCTHGIKPVVSTQLLNGSLAEGEIIIRSEN
LTNNAKTIIVHLNKSVEINCTRPSNNTRKSIIRIGPGQAFYATGDIIGDIRKAHCNISGTK
WNHTLEQVMEELKKHFP-NKTIKFNSSSGGDLEITTHSFNCRGEFFYCNTSGLFNSTW--
--NDTTINR----TEGSNNTR---NITLPCRIKQIINMWQGVGRAMYAPPIAGNITCKS
NITGILLTRDGGNNN-----STNETFRPGGDMRDNWRSELYKYKVVEIKPLGIAPTAK
RRVVEREKRAVG-IGAVFLGFLGTAGSTMGAASITLTVQARQLLSGIVQQQSNLLRAIEA
QQHMLQLTVWGIKQLQTRVLAIERYLKDQQLLGIWGCSGKLICTTNVPWNSSWSNKSQSE
IWDNMTWMQWDREISNYTDTIYRLLEDSQNQQEKNEKDLLALDSWKNLWNWFDITNWLWY
IRIFIMIVGGLIGLRIFAVLSIVNRVRQGYSPSLFQTRLPA PRGPDRPEGIEEEEGGERD
RDRSVRLVDGFLALIWDLRLSLCLFSYHRLRDFILIAARTVELLGHSSSLKGLRLGWEGLK
YLWNLLQYWIQELKNSAVSLLNATAIAVAEGTDRVIEVVQRAYRAILHIPTRIRQGLERA
LL-

35/109

Fig. 10 cont'd-10

>ENV-M.syn4.1

MRVKETRKNYQHLWKWGTML-----LGMLMICSAAEQLWVTVYYGVPVWKEATTTLF
 CASDAKAHETEVEHNIWATHACVPTDPNPQEVVLGNVTENFNMWKNDMVEQMHTDIISLWD
 QSLKPCVELTPLCVTLHCTDL-----GNATNT-----MEKGEIKNCSFNMTTEL RD
 KKQKVYALFYRLDIVPI-DNDNTS-----YRLINCNTSVIKQACPKVTFEPIPIH
 YCTPAGFAILKCNCKNFNGTGPCKNVSTVQCTHGIRPVVSTQLLNLNGSLAEKEIIIRSEN
 LTDNAKTIIVHLNKSVEINCTRPSNNTKRSVRIGPGQTFYATGDIIGDIRQAHCNISRAK
 WNNTLKQIVTKLREQFK-NKTIVFNQSSGGDLEITTHSFNCRGEFFYCNTTQLFNSTW--
 -----KN---DTEVSNNTK-GNDTITLPCRIKQIVNMWQEVGRAMYAPPIEGNITCNS
 NITGILLTRDGGNNGNET--NGTEIFRPGGGNMRDNWRNELYKYKVVEIKPLGVAPTEAK
 RRVVEREKRAVG-IGAVFLGFLGAAGSTMGAASITLTVQARQLLTGIVQQQSNLLRAIEA
 QQHMLQLTVWGKQLQTRVLAIERYLKDQQLLGLWGCSGKLICTTAVPWNSSWSNKTynd
 IWDNMTWMQWDREISNYTDTIYRLLEDSQNQQEKNEKDLLALDSWKNLWNWFDITNWLWY
 IKIFIMIVGGLIGLRIIFAVLSIVNRVRQGYSPLSFQTLTPNPRGPDRPGGIEEGGEGQG
 RDRSIRLVNGFLALAWDDLRLNCLFSYHQLRDFILIVARAVELLGRSSRLRLQRGWEALK
 YLGS LVQYWGLELKKSAISLLDTIAIAVAEGTD R VIEVVQRAYRAILHIPTRIRQGLERL
 LL-

>ENV-M.syn4.2

MRVRGIQRNWPQWWIWGILG-----FWMLMICNVVGNLWVTVYYGVPVWKEAKTTLF
 CASDAKAYEKEVHNWATHACVPTDPSPQEVVLENTENFDMWKNNMVEQMVEDVISLWD
 QSLKPCVKLAPLCVTLNCTDYVKNIT-NNATSTNSSW-GKPMKGEIKFCSFNITTSIRN
 KVQKQYALFYKLDVVQM-DEDNTS-----YRLISCNTSTITQACPKVTFDPIPIH
 YCAPAGFAILKCNKTFNGTGPCTNVSTVQCTHGKIPVVSTQLLNLNGSLAEKEIIIRSEN
 LTNNAKTIIVHLNESVEIVCTRPNNNTRKSIHIGPGRFYATGEIIGDIRQAHCNLSRAK
 WNDTLKQIVIKLREQFG-NKTIIFNQSSGGDPEITTHSFNCGGEFFYCNSTQLFNSTWNF
 --NGTWNKN---FNNTWNNTTEGTNDTITLPCIKQIINMWQVRGQAMYAPPISGQIRCSS
 NITGLILTRDGGN---DT--SGTEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTKAK
 RRVVQREKRAVG-IGALFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQQSNLLKAIEA
 QQHLLQLTVWGKQLQARILAVERYLKDQQLLGIWGCSGKLICTTTVPWNASWSNKSLE
 IWDNMTWMEWEREIDNYTGLIYNLIEESQTQQEKNEQELLELDKASLWNWFDITKWLWY
 IKIFIMIIGGLIGLRIVFTVLSIVNRVRKGYSPLSFQTLTHHQREPDRPERIEEGGEGQD
 RDRSGRLVDGFLAIIWVDLRLSLCLFSYHRLRDLILLIVTRIVELLG-----RRGWEVLK
 YWWNLLQYWSQELKNSAVSLLNATAIAVAEGTDRIEVIQRICRAIRNIPRRIRQGLERA
 LL-

>ENV-M.syn4.3

MRVKETQMNWPNLWKWGTLI-----LGLVIICSASDNLWVTVYYGVPVWKDAETTLF
 CASDAKAYDTEVHNWATYACVPTDPNPQEIHLNVTEEFNMWKNDMVDQMHEDIISLWD
 ESLKPCVKLTPLCVTLNCTDDVRNVT-NNATNTNSSW-GEPMEKGEIKNCSFNITTSIRD
 KVQKEYALFYKLDVPI-DNDSNNTN-----YRLISCNTSVITQACPKVSFEPIPIH
 YCAPAGYAILKCNCKKFNGTGPCNNVSTVQCTHGKIPVVTQLLNLNGSLAEKEIIIRSEN
 ITNNAKTIIVQLNESVVINCTRPNNNTRKSIIRIGPGQAFYATGDIIGNIRQAHCNISRAK
 WNNTLRQIVEKLGEQFGNNKTIVFNHSSGGDPEIVTHSFNCAGEFFYCNTTKLFNSTWTR
 N-NGTWTNRN---DTERSNTSE---EHITLPCRIKQIINMWQEVGKAMYAPPIRGQIRCSS
 NITGLLLTRDGGNNN-----STNETFRPGGGNMKDNWRSELYKYKVQIEPLGIAPTKAK
 RRVVEREKRAVG-LGAVFLGFLGTAGSTMGAASLTLTVQARQVLSGIVQQQRNLLRAIEA
 QQHLLKLTWVGKQLQARVLAIERYLQDQQLLGMWGCSGKLICTTNVPWNSSWSNKSQTD
 IWDNMTWLQWDKEISNYTSLIYTLIEESQNQQEKNEQDLLALDKWASLWSWFDISNWLWY
 IKIFIIIVGGLIGLRIVFAVLSIINRVQGYSPSLQTLIPNPRGPDRLGRIIEEGGEGQD
 RDRSIRLVSGFLALAWDDLRLSLCIFSYHRLRDFILIAARTVELLGHSSSLKGLRLGWEGLK
 YLGNLLLYWGQELKNSAINLLDTIAIAVAGWTD R VIEIGQ RAGRAILNIPRRIRQGFERA
 LL-

>ENV-M.syn4.4

MRVTGIRKNYQHLWRWGTMLLWRWGTMLLGILMICSAAAGKLWVTVYYGVPVWRDADTTLF
 CASDAKAYDTEAHNVWATHASVPTDPNPQEIIVLENTENFNMWKNNMVEQMHEDIISLWD
 QSLKPCVQLTPLCVTLNCTN-VNVTNLKNETNTKSSSGGEKMEEGEMKNCSFNITTEIRD
 KKQKVHALFYKLDIVPL-NSNSSE-----YRLINCNTSAITQACPKVSFDPIPIH

36/109

Fig. 10 cont'd-11

YCTPAGYAILKCNKKFNGTGPCKNVSSVQCTHGKIPVISTQLLNGSLAE E EVVIRSEN
 FTNNAKTIMVQLNVSVEINCTRPNNNTRRSIPIGPGRAFYTTGDIIGDIRKAHCNISRAQ
 WNNTLKHIVEKLKGQFGNNKTIIIFKPSSGGDPEIVMHSFNCGGEFFYCNTSGLFNSTW--
 -----N---STSLFNSTN---GTITLQCRIKQIINMWQGVGRAMYAPPIAGNITCKS
 NITGLLLVRDGGT---EP--NDTETFRPGGGDMKDNWRSELYKYKVVRIEPLGVAPTRAK
 RRVVEREKRAIG-LGAMFLGFLGAAGSTMGAASVTLTVQARLLLSGIVQQQNLLRAIEA
 QQHLLRLTVWGIKQLQARVLAVERYLRDQQLGIWGCSGKIICTTAVPWNTSWSNRSLSNE
 IWNNMTWMEWEKEIDNYTNLIYNLLEESQNQQEKNEQELLALDKWANLWNWFDISNWLWY
 IRI FIMIVGGLVGLRIVFAVLSIVKVRVQGYSPLSFQTRLPAPRGPDRPEGIEEEGGGERD
 RDRSVRLVDGFLALIWD DLRSCLFSYHHLRDL LLIVARIVELLG-----RRGWEALK
 YWWNLLQYWIQELKNSAISLLNATAVAVAEGTDRVIEALQRACRAILHIPRRIRQGFEEA
 LL-

>ENV-M.syn6.1

MRVMGIQRNCQQWWIWGILG-----FWMLMICNVMGNLWVTVYYGVPVWKEANTTLF
 CASDAKAYEREVHNWATHASVPTDPNPQEVVLENVTEDFNMWKNNMVEQM QEDVISLWD
 QSLQPCVKLTPLCVTLNCTD-LRNATNGNDTNTTSSS-REMMGGGEMKNCSFNITTEIRD
 KKQKVYALFYKLDVVPIDNDSNNTN-----YRLISCNTSAVTQACPKVTFDPIPIH
 YCTPAGFAILKCRDKKFNGTGPCNNVSTVQCTHGKIPVVTQLLNGSLAE E EVVIRSEN
 FTDNAKTIIIVQLKEAVEINCTRPNNNTRKGIHIGPGRAFYATGEIIGDIRQAHCVSRSE
 WNKTLLQVATQLRKHF--NKTIIIFNSSSGGDLEITTHSFNCRGEFFYCNTSGLFNSTW--
 --NDTTINR----TEGSNNTR---NITLPCRIKQFINMWQEVGRAMYAPPIAGNITCRS
 NITGLLLTRDGGNNGNET--NGTEIFRPGGGDMRNNWRSELYKYKVVEIKPLGIAPTAKAR
 RRVVQREKRAVG-IGAVFLGFLSAAGSTMGAASITLTVQARQLLTGIVQQQSNLLKAIEA
 QQHMLQLTVWGVKQLQARVLAVERYLRDQQLGIWGCSGRLICTTAVPWNTSWSNKSLSNE
 IWDNMTWMEWEREIDNYTSLIYTLIEESQNQQEKNEQELLELDKWANLWNWFSITNWLWY
 IRI FIMIVGGLIGLRIIFGVLSIVKVRVQGYSPLSFQTRLPAPRGPDRPEGIEEEGGGERD
 RDRSGRLVDGFLALIWD DLRSCLFSYHRLRDLILIAARIVELLGHSSSLKGLRLGW EALK
 YLWNLLLYWGQELKNSAISLLNTTAIVVAEGTDRVIEVLQRAGRAILNIPRRIRQGFEEA
 LL-

>ENV-M.syn6.2

MRVTGIRKNYQHLWRWGTM LLWRWGTM LLGILMICS AAGKLWVTVYYGVPVWREAKTTLF
 CASDAKAYEKEVHNWATYACVPTDPNPQEMVLENVTENFNMWKNNMVDQM HEDIISLWD
 ESLKPCVKLTPLCVTLHCTDL-----GNATNT-----MEKGEIKNCSFNITTEIRD
 KKQKVHALFYRLDVVPIDNDNTS-----YTLINCNTSVITQACPKVTFEPIPIH
 YCAPAGFAILKCNKKFNGTGPCNTVSTVQCTHGIRPVVSTQLLNGSLAE E EIIIRSEN
 LTDNAKTIIIVHLNESVEIVCTRPNNNTRKSVRIGPGQTFYATGAIIGDIRQAYCNISRAK
 WNNTLQIVTKLREQFGNNKTIIIFKPSSGGDLEITMHHFNCRGEFFYCNTTQLFNSTWNF
 --NGTWNKN---FNNTWNTEGTNDTITLPCIKQIINMWQGVGRAMYAPPISGQIRCSS
 NITGLLLTRDGGT-----NNTTEIFRPGGGNMRDNWRSELYKYKVVKIEPLGVAPTKAK
 RRVVQREKRAVG-IGALFLGFLGAAGSTMGAASMTLTVQARQVLSGIVQQQNRLLRAIEA
 QQHLLQLTVWGIKQLQARILAVERYLKDQKFLGLWGCSGKIICTTAVPWNASWSNKS LDD
 IWNNMTWMQWEREIDNYTG LIYSLIEESQTQQEKNEQELLQLDKWASLWNWFDITNWLWY
 IRLFIMIVGGLVGLRIVFTVLSIVNRVRKGYSPLSFQTLTHHQREPDRPERIEEGGGEQG
 RDRSVRLVSGFLALFWDDLRSCLFCYHRLRDFILIAARTVELLGHSSSLKGLRRGW EGLK
 YLWNLLQYWIQELKNSAISLLNATAVAVAEGTDRVIEALQRACRAILHIPRRIRQGLERL
 LL-

>ENV-M.syn6.3

MRVRGIQRNWPQWWIWGILG-----FWMIIICRVVGNLWVTVYYGVPVWKDAETTLF
 CASDAKS YETEAHNIWATHACVPTDPSPQEVVLGNVTENFNMWKNDMVEQM HEDIISLWD
 QSLKPCVELTPLCVTLNCTDDVRNVT-NNATNTNSSW-GEPMEKGEIKNCSFNITTSIRN
 KVQKQYALFYKLDVVQIIDNNSTNTS-----YRLINCNTSAITQACPKVSFDPIPIH
 YCAPAGYAILKCNKTFNGTGPCNVSTVQCTHGIRPVISTQLLNGSLAE E EVVIRSEN
 FTNNAKTIMVQLNVSVEINCTRPNNNTRKSIHIGPGRAFYTTGDIIGDIRQAHCNISRAQ
 WNNTLKHIVEKLKGQFGNNKTIIIFKPSSGGDPEIVMHSFNCRGEFFYCNTSKLFNSTWTR
 N-NGTWTRN---DTERS NSTE---EHITLPCRIKQIINMWQRVGQAMYAPPIAGNITCNS
 SITGLLLTRDGGN---DT--SGTEIFRPGGGNIKDNWRSELYKYKVVKIEPLGVAPTRAK

37/109

Fig. 10 cont'd-12

RRVVEREKRAVG-IGAMIFGFLGAAGSTMGAASMAITVQARQLLSGIVQQQSNLLMAIEA
 QQHLLKLTVWGIKQLRARVLAVERYLKDQQLGIWGCSGKHICTTNVPWNSSWSNKSLE
 IWNNMTWIEWEREINNYTGLIYNLLEKSQNNQEKNEQDLLALDKWASLWSWFDISNWLWY
 IKIFIIIVGGLIGLRIVFAVLSLVNRVRQGYSPSLQTLPTPRGPDRPEGTEEGGEQG
 RDRSIRLVSGFLALAWDDLRLSLCRFSYHRLRDFILIVARTVELLGRSSLKGLRLGWEGLK
 YLGNLLLYWGQELKISAIISLLDTTIAIAGWTDRIEIGQRLCRAIRNIPRRIRQGAERA
 LQ-

>ENV-M.syn6.4

MRVKETQMNWPNLWKWGTLI-----LGLVICSASDNLWVTVYYGVPVWRDADTTLF
 CASDAKAHETEVEHNVWATHACVPTDPNPQEIHLNVTEEFNMWKNMVEQMHTDIISLWD
 QSLKPCVRLTPLCVTLNCTDELKNATFRSNTTNSW--EKMERGEIKNCSFNITTSIRD
 KVQKEYALFYKLDIVPL-NSNSSE-----YRLINCNTSVIKQACPKISFDPIPIH
 YCAPAGFAILKCKDKKFNGTGPCQNVSTVQCTHRIKPVVSTQFLLNGSLAEEDIIIRSEN
 ITNNAKTIIIVQLNESVVINCTRPNNNTRRSIPIGPGRVFYTTEDIIGDIRQAHCNLSRAK
 WNDTLKQIVIKLREQFG-NKTIVFNQSSGGDLEIVMHSFNCGGEFFYCNSTQLFNSTWF-
 --NSTW-----STEGSNTE-GSDTITLPCRIKQIVNMWQGVGKAMYAPPIRGQIRCSS
 NITGILLTRDGGTNGT---NETETFRPGGGMKDNWRSELYRYKVVKIEPLGIAPTAK
 RRVVEREKRAIG-LGAMFLGFLGTAGSTMGAASLTTLTVQARQLMSGIVQQQNNLLRAIEA
 QQHMLKLTVWGIKQLQARVLALERYLKDQQLGLWGCSGKLICTTTVPWNSSWSNKSQTD
 IWDNMTWMQWDREISNYTNTIYRLLEDSQNNQEKNEKDLLALDSWKNLWNWFDITKWLWY
 IKIFIMIVGGLIGLKIVFAVLSIINVRVRQGYSPSLFQTLIPNPRGPDRPGGIEEGGEQD
 RDRSIRLVNGFLALIWDLRSLFLFSYHRLRDLILLIVTRIVELLG-----RRGWEALK
 YWWNLLQYWSQELKNSAINLLDTIAIAVAEGTDRIIEVIQRICRAIRNIPTRIRQGLERA
 LL-

>ENV-M.syn6.5

MRVKGIRKNYQHLWKWGTM-----LGMLMICSATEKLWVTVYYGVPVWKEATTTLF
 CASDAKAYDTEVEHNVWATYACVPTDPNPQELVLENTENFDMWKNMVEQMHEIDIINLWD
 QSLKPCVKLTPLCVTLNCTDYVKNIT-NNATSTNSW-GKPMKEGEIKNCSFNMTTELRLD
 KKQKVYSLFYKLDVVQM-DEDNTS-----YRLISCNTSVITQACPKISFEPIPIH
 YCTPAGYAILKCNCKNFNGTGPCKNVSSVQCTHGKIPVISTQLLLNGSLAEEDIIIRSEN
 LTNNVKTIIIVHLNKSVEINCTRPSNNTRTSIRIGPGQAFYATGDIIGDIRKAHCNISRAK
 WNNTLRQIVEKLGEQFGNNKTIVFNHSSGGDPEITTHSFNCGGEFFYCNTTKLFNSTWTW
 N-NSTW--N---NTRKSDTE---EITLPCRIKQIINMWQEVGKAMYAPPIQGVIRCES
 NITGLILTRDGGNNN-----STNETFRPGGDMRDNRSELYKYKVRIEPLGVAPTEAK
 RRVVQREKRAVGTIGAMFLGFLGAAGSTMGAASVTLTVQARLLLSGIVQQQNNLLKAIEA
 QQHLLRLTVWGIKQLQARVLAIERYLQDQQLGIWGCSGKLICTTNVPWNSTWSNRSLE
 IWNNMTWMEWEKEIDNYTNLIYNLLEESQIQQEKNEQELLALDKWANLWNWFDISNWLWY
 IRIFIIIVGGLVGLRIVFAVLSIVNKVRQGYSPSLFQTHLPAQRGPDRPEGIEEGGEQD
 RDRSVRLVDGFLAIIWVDLRSLCLFSYHHLRDLILLIVARIVELLG-----RRGWEVLK
 YWWNLLKYWSQELKNSAVSLLNATAIAVAEGTDRIIELIQRICRAICNIPRRIRQGFERA
 LL-

>ENV-M.syn6.6

MRVKETRKNYQHLWRWGIML-----LGMLMICSAAEQWLWVTVYYGVPVWKEAKTTLF
 CASNAKAYDTEAHNVWATHACIPTDPNPQEIIVLENTESFNMWKNDMVDQMHEDEVISLWD
 QSLKPCVQLTPLCVTLNCTN-VNVTNLKNETNTKSSSGGEKMEEGEMKNCSFNVTTELRLD
 KKKKEYALFYRLDIVPL-NEGNSNSSY-----YRLINCNTSTITQACPKVSFEPIPIH
 FCAPAGFAILKCNCKKFNGTGPCKNVSTVQCTHGKIPVSTQLLLNGSLAEKEIIIRSEN
 LTNNAKIIIVQLNESVEINCTRPGNNTRKSIRIGPGQTFYATGDIIGNIRQAHCNISRTQ
 WNNTLKQIAIKLREQFG-NKTIIIFNQSSGGDPEIVTHSFNCGGEFFYCKSTKLFNSTW--
 -----N---STSLFNSTN---GTITLQCRKQIINRWQEVGKAMYAPPIEGNITCKS
 NITGLLLVRDGGINVTNN--TGTEVFRPGGDMKDNWRNELYKYKVVEIKPLGVAPTRAR
 RRVVEREKRAVG-LGAVFLGFLGAAGSTMGAASVTLTVQARQLLFGIVQQQSNLLRAIEA
 QQRMLQLTVWGIKQLQTRVLAIERYLKDQQLGMWGCSGKLICTTAVPWNSSWSNKTIND
 IWDNMTWLQWDKEISNYTDTIYRLLEESQNNQEKNEKDLELDKWASLWNWFDITNWLWY
 IKIFIMIIGGLIGLRIFAVLSIVNRVRQGYSPSLFQTLTPNPRGPDRLGRIIEEGGEQD
 KDRSIRLVNGFSALIWDLRNLCLFSYHQLRDFILVTARAVELLGRSSLRGLRQGWALK
 YLGSVLQYWGLELKKSAISLLDTIAIAVANWTDRIEIVVQRAYRAILHIPTRIRQGFEEA
 LQ-

38/109

Fig. 10 cont'd-13

>POL-B.syn1.1

FFRENLAFFPQGKAREFSSEQTRANSPTRR-----ELQVWGRDNNNSLSEAGA
 DR----QGTVS-FSFPQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMNLPGRWKPK
 MIGGIGGFIKVRQYDQIPIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFPISPIE
 TVPVKLKPGMDGPKVKQWPLTEEEKIKALVEICTEMEKEGKISKIGPENPYNTPVFAIKKK
 DSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDKDFRK
 YTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQNPDIYIYQYMD
 LYVGSDELEIGQHRTKIEELRQHLLRWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIVL
 PEKDSWTVNDIQKLVGKLNWASQIYAGIKVKQLCKLLRGTKALTEVIPLTEEALELAEN
 REILKEPVHGVYDPSKDLIAEIQKQGQGWYQIYQEPFKNLKTGKYARMGAHTNDVK
 QLTEAVQKIATESIVIWGKTPKFKLPQKETWETWWTEYWQATWIPWEFVNTPLVLKLW
 YQLEKEPIVGAETFYVDGAANRETKLGKAGYVTDGRGRQKVSLTDTTNQKTELQAIHLAL
 QDSGLEVNIVTDSQYALGIIQAQPDKSESELVSQIIEQLIKKEKVYLAWVPAHKGIGGNE
 QVDKLVSAIRKVLFLDGIDKAQEEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKCQLK
 GEAMHGQVDCSPGIWQLDCTHLEGKIILVAVHVASGYIEAEVIPAETGQETAYFLLKLAG
 RWPVKTIHTDNGSNFTSTTVKAACWWAGIKQEFPIPNPQSQGVVESMNKELKKIIGQVR
 DQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERIVDIIATDIQTKELQKQITKIQNFRVY
 YRDSRDPLWKGPALLWKGEAVVIQDNSDIKVVPRRKAKIIRDYGKQMGDDCVASRQD
 ED-

>POL-B.syn3.1

FFRENLAFFPQGKAREFPSEQTRANSPTSR-----ELQVWGGDNNNSLSEAGD
 DR----QGTVS-FSFPQITLWQRPIVTIKIGGQKKEALLDTGADDTVLEEMNLPGRWKPK
 IIGGIGGFIKVKQYDQILIEICGHKAIGTVLVGPTPANIIGRNLLTQIGCTLNFPISPIE
 TVPVRLKPGMDGPKVKQWPLTEEEKIKALIEICTELENEGKISKIGPENPYNTPVFAIKKK
 DGTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDKEFRK
 YTAFTIPSINNETPGIRYQYNVLPQGWKGSPSIFQSSMTKILEPFRKQNPDIYIYQYMD
 LYVGSDELEIEQHRTKIEELRQHLLKWGFYTPDKKHQKEPPFLWMGYELHPDKWTVQPIVL
 PEKDSWTVNDIQKLVGKLNWASQIYAGIKVKQLCRLLRGTKALTEVVPLTEEALELAEN
 REILKEPVHGVYDPSKDLVAEIQKQGLGQWYQIYQEPYKNLKTGKYAKMRGAHTNDVK
 QLTEAVQKIATESIVIWGKTPKFRPLQKETWEAWWMEYWQATWIPWEFVNTPLVLKLW
 YQLEKEPIVGAETFYVDGAANRDTKLKAGYVTDGRGRQKVSLTDTTNQKTELQAIHLAL
 QDSGLEVNIVTDSQYAIGIIQAQPDSESELVSQIIEQLIKKEKVYLAWVPAHKGIGGNE
 QVDKLVSSGIRKVLFLDGIDKAQEDHEKYHSNWKAMASDFNLPPVVAKEIVACCDKCQLK
 GEAMHGQVDCSPGIWQLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFILKLAG
 RWPVTTIHTDNGSNFTSTAVKAACWWAGIKQEFPIPNPQSQGVVESINKELKKIIGQVR
 DQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERMIDIATDIQTTELQKQITKLQNFRVY
 FRDSRDPLWKGPALLWKGEAVVIQDNSEIKVVPRRKAKIIRDYGKQMGDDCVASRQD
 ED-

>POL-B.syn3.2

FFREDLAFLQGKAREFSSEQTRANSPTRG-----ELQVWGRDNNNSLSEAGA
 DR----QGTVS-LSFPQITLWQRPLVTVKIGGQLKEALLDTGADDTVLEEMSLPGRWKPK
 MIGGIGGFIKVRQYDQILVEICGHKAIGTVLIGPTPVNIIGRDLLTQIGCTLNFPISPID
 TVPVKLKPGMDGPKVKQWPLTEEEKIKALVEICTEMEKEGKISKIGPENPYNTPIFAIAKKK
 DSTKWRKLVDFRELNRRRTQDFWEVQLGIPHPAGLKQKKSXTVLVDVGDAYFSVPLDEDFRK
 YTAFTIPSTNNETPGIRYQYNVLPQGWKGSPAIFQSSMTRILEPFRKQNPDLVIYQYMD
 LYVGSDELEIGQHRTKIEELREHLLKWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIML
 PEKDSWTVNDIQKLVGKLNWASQIYPGKIKVKQLCKLLRGAKALTEVIPLTKEAELELAEN
 REILKEPVHGAYDPSKDLIAEIQKQGQGWYQIYQDPFKNLKTGKYARMGAHTNDVR
 QLTEAVQKITTESIVIWGKIPKFKLPQKETWETWWTEYWQATWIPWEFVNTPLVLKLW
 YQLEKEPIIGAETFYVDGAANRETKLGKAGYVTNKGGRQKVVSITDTTNQKTELQAILLAL
 QDSGLEVNIVTDSQYAIGIIQAQPDKSESELVSQIIEELIKKEKVYLTWVPAHKGIGGNE
 QIDKLVSAIRKVLFLDGIDQAQEEHEKYHSNWRAMASDFNIPPVVAKEIVASCDKCQLK
 GEAIHGQVDCSPGIWQLDCTHLEGKIILVAVHVASGYIEAEIIPETGQETAYFLLKLAG
 RWPVKTVHTDNGSNFTSTTVKAACWWAGVKQEFPIPNPQSQGVVESMNNELKKIIGQIR
 DQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERIVDIIATDIQTKELQKQITKIQNFRVY
 YRDNRDPLWKGPALLWKGEAVVIQENSDIKVVPRRKVKIIRDYGKQMGDDCVASGQD
 ED-

39/109

Fig. 10 cont'd-14

>POL-B.syn3.3

FFREDLAFPQGEAREFSSEQTRANSPTRR-----ELQVWGRDSNSLSEAGA
 DR----QGTVS-FNFPQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEDMNLP GKWKPK
 MIGGIGGFIKVRQYDQIPIEICGHKAVGTVLVGPTPVNIIGRNLLTQLGCTLNFPISP
 TVPVKLKPGMDGPKVKQWPLTEEEKIKALVEICTELEKEGKISKIGPENPYNTPVFAIKKK
 DSTKWRKVVDFRELNKKQTQDFWEVQLGIPHPSGLKKKKSVTVLDVGDAYFSVPLDKDFRK
 YTAFTIPSVNNETPGVRYQYNVLPQGWKGSPAIFQCSMTKILEPFRKQNPDIVIYQYMDD
 LYVGSDEIGQHRAKIEELRQHLLRWGFTTPDKKHQNEPPFLWMGYELHPDKWTVQPIEL
 PEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGTKALTEVIPLTEEALELAEN
 REILREPVGHVYDPTKDLIAEIQKQEQGWYQIYQEPFKNLKTGKYARTRGAHTNDVK
 QLTEAVQKVATESIVIWGKTPKFKLPIQKETWEAWWTEYWQATWIPEWEFVNTPLVLKLW
 YQLEKEPIEGAETFYVDGASNRETKLGKAGYVTNRGRQKVPLTDTTNQKTELQAIYLAL
 QDSGSEVNIVTDSQYALGIIQAQPDKSESELVNQIIEQLIKKEKIYLAWVPAHKGIGGNE
 QVDKLVSA GIRKILFLDGDIDKAQEEHEKYHNNWRAMASDFNLPPIVAKEIVASCDKQCLK
 GEAIHQVDCSPGIWQLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFILKLAG
 RWPVKTIHTDNGSNFTSATVKAACWWAGIKQEFGI PYNPQSQGVVESMNKELKKIIGQVR
 DQAEHLKTAVQMAVFIHNFKRKGGIGEYSAGERIIDIIATDIQTRELQKQITKIQNFRVY
 YRDSRDPLWKGPALLWKGE GAVVIQDNSDIKVPRRKAKIIRDY GKQ MAGDDCVAGRQD
 ED-

>POL-B.syn4.1

FFRENLAFPQGEAREFSSEQNRANSPTTR-----ELQVWGGDNNLSLSEAGA
 DR----QGTVS-LSFPQITLWQRPLVTIRIGGQLKEALLDTGADDTVLEEMSLPGRWKPK
 MIGGIGGFIKVRQYDQILIEICGHKAIGTVLIGPTPVNIIGRNLLTQLGCTLNFPISPID
 TVPVKLKPGMDGPKVKQWPLTEEEKIKALVEICTEMEKEGKISKIGPENPYNTPIFAIAIKKK
 DSTKWRKLVDFRELNRRTQDFWEVQLGIPHPSGLKKKKSVTVLDVGDAYFSVPLDKDFRK
 YTAFTIPSINNETPGVRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQNPDMVIYQYMDD
 LYVGSDEIGQHRTKIEELRQHLLRWGLTTPDKKHQKEPPFLWMGYELHPDKWTVQPIKL
 PEKDSWTVNDIQKLVGKLNWASQIYAGIKVRQLCKLLRGAKALTEVIPLTAEAELELAEN
 REILREPVGHVYDPSKDLIAEIQKQGGQWYQIYQDPFKNLKTGKYAKMRGAHTNDVK
 QLTEAVQKVATESIVIWGKTPKFKLPIQKETWEAWWTEYWQATWIPEWEFVNTPLVLKLW
 YQLEKEPIEGAETFYVDGAANRDTKLKAGYVTDGRGRQKVSLTDTTNQKTELQAIHLAL
 QDSGLEVNIVTDSQYALGIIQAQPDKSESELVSQIIEELIKKEKVYLAWVPAHKGIGGNE
 QIDKLVSA GIRRVLFLDGDIDQAQEEHEKYHSNWRAMASDFNLPPIVAKEIVASCDKQCLK
 GEAIHQVDCSPGIWQLDCTHLEGKIIILVAVHVASGYIEAEVIPAETGQETAYFLLKLAG
 RWPVKTIHTDNGSNFTSTTVKAACWWAGIKQEFGI PYNPQSQGVVESMNKELKKIIEQVR
 DQAEHLKTAVQMAVFIHNFKRKGGIGEYSAGERIIDIIATDIQTRELQKQITKIQNFRVY
 YRDNRDPLWKGPALLWKGE GAVVIQDNSDIKVPRRKAKIIRDY GKQ MAGDDCVAGRQD
 ED-

>POL-B.syn4.2

FFRENLAFPQKGAREFPSEQTRANSPTSR-----ELQVWGRDNNLSLSEAGD
 DR----QGTVS-FSFPQITLWQRPLVTVKIGGQLKEALLDTGADDTVLEEMNLPGRWKPK
 MIGGIGGFIKVRQYDQIPIEICGHKAVGTVLVGPTPVNIIGRDLLTQIGCTLNFPISP
 TVPVKLKPGMDGPKVKQWPLTEEEKIKALIEICTELENEGKISKIGPENPYNTPVFAIKKK
 DGTKWRKLVDFRELNKRQTQDFWEVQLGIPHAGLKQKKSVTVLVDVGDAYFSVPLDKEFRK
 YTAFTIPSTNNETPGIRYQYNVLPQGWKGSPAIFQCSMTKILEPFRKQNPDLVIYQYMDD
 LYVGSDEIEQHRTKIEELREHLLKWGFTTPDKKHQNEPPFLWMGYELHPDKWTVQPIML
 PEKDSWTVNDIQKLVGKLNWASQIYAGIKVKQLCKLLRGTKALTEVVPLTEEALELAEN
 REILKVPVGHVYDPSKDLVAEIQKQGLGQWYQIYQEPFKNLKTGKYARTRGAHTNDVR
 QLTEAVQKIATESIVIWGKTPKFKLPIQKETWEAWWMEYWQATWIPEWEFVNTPLVLKLW
 YQLEKEPIVGAETFYVDGAANRETKLGKAGYVTDKGRQKVPLTDTTNQKTELQAINLAL
 QDSGSEVNIVTDSQYALGIIQAQPDSESELVSQIIEQLINKEKVYLAWVPAHKGIGGNE
 QVDKLVSSGIRKVLFLDGDIDKAQEDHEKYHSNWRAMAGDFNLPPVVAKEIVACCDKQCLK
 GEAMHGQVDCSPGIWQLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFILKLAG
 RWPVKTVHTDNGSNFISTTVKAACWWAGVKQEFGI PYNPQSQGVVESMNNELKKIIGQVR
 DQAEHLKTAVQMAV FVHNFKRKGGIGGYTAGERIVDIIASDIQTRELQKQITKIQNFRVY

40/109

Fig. 10 cont'd-15

YRDSRDPIWKGPAKLLWKGEAVVIQDNSEIKVVP RRKAKI IRDYGKQMAGDDCVASRQN
ED-

>POL-B.syn4.3

FFREDLAF LQ GKAREFSSEQTRANSPTRR-----ELQVWGRDNNSPSEAGA
DR----QGTVS-FNFPQITLWQRPIVTIKIGGQLKEALLDTGADDTVLEDNLP GKWKPK
MIGGIGGFIKVRQYDQILVEICGHKAIGTVLVGPTPANI IGRNLLTQIGCTLNFPISPIE
TVPVKLKSGMDGPKVKQWPLTEEEKIKALIEICTEMEKEGKISKIGPENPYNTPVFAIKKK
DSTKWRKLVDFRELNKR TQDFWEVQLGIPHP SGLKKKKS VTLVDVGDAYFSVPLDEDFRK
YTAFTIPSVNNETPGIRYQYNVLPQGWKGSPAIFQSSMTRILEPFRKQNPDI VIYQYMD
LYVGS DLEIGQHRAKIEELRQHLLRWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIEL
PEKDSWTVNDIQKL VGKLNWASQIYPGIKVRQLCKLLRGTKALTEVIPLTEEA ELELAEN
REILKEPVHGVYYDPSKELIAEIQKQEQGWTYQIYQEPFKNLKTGKYARMRGTH TNDVK
QLTEAVQKITTESIVIWGRTPKFKLP IQKETWESWWTEYWQATWIPEWEFVNT PPLVKLW
YQLEREPIAGAE TFYVDGASNRET KL GKAGYVTNRGRQKVVS L P DTTN QKTELQAIY LAL
QDSGLEVNIVTDSQYAIGIIQAQPDKSESELVNQII EQLIKKEKIYLA WVP AHKGIGGNE
QVDKLVSN GIRKILFLD GIDKAQDEHEKYHSNWKAMASDFNLPPVVAKEIVACCDKCQLK
GEAMHGQVDCSPGIWQLDCTHLEGKIILVAVHVASGYIEAEVIPAETGQETAYFLLKLAG
RWPVKI IHTDNGSNFTSTAVKAACWWAGIKQEF GIPYNPQS QGVVESINKELKKI IKQVR
DQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERMIDIATDIQTTELQKQITKLQNF RVY
FRDSRDPLWKGPAKLLWKGEAVVIQDNNDIKVVP RRKVKI IRDYGKQMAGDDCVASGQD
ED-

>POL-B.syn4.4

FFREDLAF PQGKARELSSEQTRANSPTRG-----ELQVWGRDSNSLSEAGA
DR----PGTVS-FSFPQITLWQRPLVTIKIGGQKKEALLDTGADDTVLEEINLPGRWKPK
IIGGIGGFIKVKQYDQIPIEICGHKVIGTVLVGPTPANI IGRNLLTQLGCTLNFPISPIE
TVPVRLKPGMDGPKVKQWPLTEEEKIKALVEICTELEKEGKISKIGPENPYNTPVFAIKKK
DSTKWRKVVDRELNKK TQDFWEVQLGIPHPAGLKKKKS VTLVDVGDAYFSVPLDENFRK
YTAFTIP SINNETPGIRYQYNVLPQGWKGSPSIFQSSMTKILEPFRKQNP EIV IYQYMD
LYVGS DLELGQHRTKIEELRQHLLKWGFYTPDKKHQKEPPFLWMGYELHPDKWTVQPIVL
PEKDSWTVNDIQKL VGKLNWASQIYPGIKVKQLCRLLRGAKALTEVIPLTKEAELELAEN
REILKEPVHGAYYDPTKDLIAEIQKQEGQWTYQIYQEPYKNLKTGKYARMRG AHTNDVK
QLTETVQKITTESIVIWGKIPKFKLP IQKETWETWWTEYWQATWIPEWEFVNT PPLVKLW
YQLEKEPIIGAETFYVDGAASRET KL GKAGYVTNKG RQKVVS I TDTTN QKTELQAI L LAL
QDSGLEVNIVTDSQYAIGIIQAQPDKSESEI VSQII EQLIKKEKVYLTVWPAHKGIGGNE
QVDKLV SAGIRKVLFLD GIDKAQEEHEKYHNNWRAMASDFNI PPVVAKEIVASCDKCQLK
GEAMHGQVDCSPGIWQLDCTHLEGKVILVAVHVASGYIEAEI IPTETGQETAYFILKLAG
RWPVTTIHTDNGSNFTSATVKAACWWAGVKQEF GIPYNPQS QGVIESMNKELKKI IGQIR
DQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERIVDIIATDIQTTELQNQITKI QNF RVY
YRDSRDPLWKGPAKLLWKGEAVVIQENS DIKVVP RRKVKI IRDYGKQMAGDDCVASRQD
ED-

>POL-B.syn6.1

FFREDLAF PQGEAREFCSEQTRANSPATR-----ELQVWGRDNNTSLSEAGA
DR----PGTVS-FSFPQITLWQRPIVTVKIEGQLKEALLDTGADDTVLEEMNLP GKWKPK
MIGGIGGFIKVRQYDQV SIEICGHKAIGTVLVGPTPANI IGRNLLTQIGCTLNFPISPIE
TVPVKLKPGMDGPKVKQWPLTEEEKIKALVEICTELEKEGKISKIGPENPYNTPVFAIKKK
DSTKWRKVVDRELNKR TQDFWEVQLGIPHP SGLKKKKS VTLVDVGDAYFSVPLDENFRK
YTAFTIP SINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQNPDI I IYQYMD
LYVGS DLEIGQHRAKIEELRQHLLKWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIEL
PEKDSWTVNDIQKL VGKLNWASQIYAGIKVKELCKLLRGTKALTEVVPLTEEA ELELAEN
REILKEPVHGVYYDPSKDLIAELQKQGGQGWTYQIYQEPYKNLKTGKYARTRG AHTNDVR
QLTEAVQKIATEGIVIWGKTPKFKLP IQKETWEAWWTEYWQATWIPEWEFVNT PPLVKLW
YQLEKEPI LGAE TFYVDGASNRET KL GKAGYVTDRGRQKVVS L TDTTN QKTELQAIN LAL
QDSGLEVNIVTDSQYALGIIQAQPD RSESELVSQII EQLINKEKVYLAWVPAHKGIGGNE
QVDKLVSTGIRRVLFLD GIDKAQEEHEKYHSNWRAMASDFNLPPIVAKEIVASCDKCQLK
GEAIHGQVDCSPGIWQLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFILKLAG

41/109

Fig. 10 cont'd-16

RWPVKTIHTDNGSNFTSTTVKAACWWAGIKQEFGI PYNPQS QGVVES MNKELKKIIEQVR
DQAEHLKTAVQMAVFVHNFKRKGIGEYSAGERIVDIIATDIQTKELQKHITKIQNFRVY
YRDSRDPLWKGPAPKLLWKGE GAVVIQDNSEIKVVP RRKAKIIRDY GKQMAGDDCVASRQD
ED-

>POL-B.syn6.2

FFREDLAFPQ GKARELSSEQTRANSPTS PTRG-----ELQVWGRDSNSLSEAGA
DR----QGPVS-FSFPQITLWQRPIVTIKIGGQLKEALLDTGADDTVLED MNLPGRWKPK
MIGGIGGFIKVKQYDEILVEICGHKAIGTVLIGPTPVNIIGRNLLTQLGCTLNFPISPIE
TVPVKLKSGMDGPKVKQWPLTEEEKIKALVEICTELEKEGKISKIGPENPYNTPIFAIKKK
DSTKWRKLVDFRELNRKTQDFWEVQLGIPHPAGLKKKKS VTVLDVGDAYFSVPLDKDFRK
YTAFTIPSVNNETPGIRYQYNVLPQGWKGSPAI FQCSMTKILEPFRKQNPDMVIYQYMDD
LYVGS DLEIGQHRIKIEELREHLLKKGFTTPDKKHQNEPPFLWMGYELHPDKWTVQPIVL
PEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGTKALTEVIPLTEAELELAEN
REILREPVHGVYDPTKDLIAEIQKQGQGWTYQIYQEPFKNLKTGKYARMGAHTNDVK
QLTEAVQKITTESIVIWGKIPKFRLP IQKETWEAWWIEYWQATWIPEWEFVNT PPLVKLW
YQLEREPIAGAETFYVDGAANRETKLGKAGYVTNRGRQKVVSITDTTNQKTELQAILLAL
QDSGLEVNIVTDSQYALGIIQAQPKDSESELVNQIIEQLIKKEKIYLAWVPAHKGIGGNE
QIDKLVSAGIRKVLFLDGDIDKAQDEHEKYHSNWRAMAGDFNLPPVVAKEIVACCDKCQLK
GEAMHGQVDCSPGIWQLDCTHLEGKIIILVAVHVASGYIEAEVIPAETGQETAYFLKLAG
RWPVTTIHTDNGSNFTSATVKAACWWAGVKQEFGI PYNPQS QGVVIES MNKELKKIIGQVR
DQAEHLKTAVQMAVFIHNFKRKGIGEYSAGERIIDIIATDIQTKELQKQITKIQNFRVY
YRDSRDP IWKGPAPKLLWKGE GAVVIQDNSEIKVVP RRKAKIIRDY GKQMAGDDCVAGRQD
ED-

>POL-B.syn6.3

FFRENLAFPQGEAREFSSEQTRANSPTRG-----ELQVWGRDSNSLSEAGD
DR----QGTVS-FSFPQITLWQRPLVTIKIGGQKKEALLDTGADDTVLEEMNLPGRWKPK
IIGGIGGFIKVTQYDQIPIEICGHKAVGTVLVGPTPVNIIGRDLLTQIGCTLNFPISPIE
TVPVKLKSGMDGPKVKQWPLTEEEKIKALIEICTEMEKEGKISKIGPENPYNTPVFAIKKK
DGTKWRKLVDFRELNRKTQDFWEVQLGIPHPAGLKKKKS VTVLDVGDAYFSVPLDREFRK
YTAFTIPSLNNETPGIRYQYNVLPQGWKGSPSIFQSSMTKILEPFRKQNPDLVIYQYMDD
LYVGS DLELGQHRTKIEELRQHLLKKGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIKL
PEKDSWTVNDIQKLVGKLNWASQIYPGIKVKQLCRLLRGAKALTEVVPLTKEAELELAEN
REILKEPVHGAYYDPTKDLIAEVQKQELGQWTYQIYQEPFKNLKTGKYARMGAHTNDVK
QLTETVQKITTESIVIWGKTPKFRLP IQKETWESWWTEYQATWIPEWEFVNT PPLVKLW
YQLEKEPITGAETFYVDGAANRETKIGKAGYVTDKGRQKVVS L P D T T N Q K T E L Q A I H L A L
QDSGSEVNIVTDSQYALGIIQAQPD RSESEVNQIIEQLIKKEKVYLAWVPAHKGIGGNE
QVDKLVSNGIRKILFLDGDIDKAQEEH ERYHSNWKAMASDFNLPPVVAKEIVACCDKCQLK
GEAIHGQVDCSPGIWQLDCTHLEGKVIILVAVHVASGYIEAEVIPAETGQETAYFILKLAG
RWPVKI IHTDNGSNFTSTAVKAACWWAGIKQEFGI PYNPQS QGVVES INKELKKI IKQVR
DQAEHLKTAVQMAVFIHNFKRKGIGGYSAGERIIDIIASDIQTKELQNQITKIQNFRVY
YRDSRDPVWKGPAPKLLWKGE GAVVIQDNSDIKVVP RRKAKIIRDY GKQMAGDDCVASRQN
ED-

>POL-B.syn6.4

FFRENLAFPQRKAREFSSEQTRANSPTRR-----ELQVWGGDNNSLSEAGA
DR----QGTVS-LSFPQITLWQRPLVTIKVGGQLKEALLDTGADDTVLEEINLPGRWKPK
MIGGIGGFIKVRQYDQILVEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFPISPIE
TVPVKLKP GMDGPKVKQWPLTEEEKIKALVEICTEMEKEGKISKIGPENPYNTPIFAIKKK
DSTKWRKLVDFRELNRRTQDFWEVQLGIPHPAGLKKKKS VTVLDVGDAYFSVPLDEDFRK
YTAFTIP SINNETPGVRYQYNVLPQGWKGSPAI FQASMTKILEPFRKQNPDIVIYQYMDD
LYVGS DLEIGQHRIKIEELRQHLLRWGLTTPDKKHQKEPPFLWMGYELHPDKWTVQPIML
PEKDSWTVNDIQKLVGKLNWASQIYAGIKVRQLCKLLRGAKALTEVIPLTAEAELELAEN
REILKVPVHGVYDPSKELIAEIQKQEQGWTYQIYQDPFKNLKTGKYARMRGHTNDVR
QLTEAVQKITTESIVIWGKIPKFKLP IQKETWETWWTEYQATWIPEWEFVNT PPLVKLW
YQLEKEPIIGAETFYVDGAASRETKLGKAGYVTD RGRQKVISLTD T T N Q K T E L Q A I H L A L
QDSGVEVNIVTDSQYALGIIQAQPKDSESEIVSQIIEQLIKKEKVYLTWVPAHKGIGGNE

42/109

Fig. 10 cont'd-17

QVDKLVSTGIRKVLFLDGDIDQAQEEHEKYHSNWRTMASDFNLPPIVAKEIVASCDKCQLK
GEAMHGQVDCSPGIWQLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAG
RWPVKTIHTDNGPNFISTTVKAACWWAGIKQEFGI PYNPQS QGVVESMNNELKKIIGQVR
EQAEHLKTAVQMAVFIHNFKRKGIGGYSAGERMIDIATDQITTELQKQITKLQNFVRVY
FRDSRDPLWKGP AKLLWKGE GAVVIQENS DIKVVPRRKAKIIRDYGKQ MAGDDCVAGRQD
ED-

>POL-B.syn6.5

FFRENLAFFPQ GKAREFPSEQTRANSPTS R-----ELQVWGRDNNSLSEAGA
NR----QGTVS-FSFPQITLWQRPLVTVKIGGQLKEALLDTGADDTVLEDMDLPGRWKPK
MIGGIGGFIKVRQYDQIPIEICGHKVI GTVLVGPTPANI IGRNLLTQIGCTLNFPISPIE
TVPVRLKPGMDGPKVKQWPLTEEEKIKALVEICTELEKEGKISKIGPENPYNTPVFAIKKK
DSTKWRKLVD FRELNKKTDQDFWEVQLGIPHP SGLKKKKS VTLVDVGDAYFSVPLDKEFRK
YTAFTIPSTNNETPGIRYQYNVLPQGWKGS PAIFQASMTKILEPFRKQNP EIVIYQYMDD
LYVGS DLEIEQHRTKIEELRQHLLRWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIVL
PDKDSWTVNDIQKLVGKLNWASQIYPGIKIRQLCKLLRGAKALTEVIPLTKEAELELAEN
REILKEPVHGVYDPSKDLIAEIQKQEGQWTYQIYQEPFKNLKTGKYAKMRGAHTNDVK
QLTEAVQKVATESI VIWGKTPKFKLP IQKETWEAWWMEYWQATWIP EWEFVNT PPLVKLW
YQLEKEPIEGAETFYVDGAANRDTKL GKAGYVTNKG RQKVVTLTDTTNQKTELQAIHLAL
QDSGLEVNIVTDSQYALGIIQAQPDKSESEIVNQIIEQLIKKEKVYLAWVPAHKGIGGNE
QVDKLVSSGIRKVLFLDGDIDKAQEDHEKYHSNWRAMANDFNLPVVAKEIVACCDKCQLK
GEAMHGQVDCSPGIWQLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFILKLAG
RWPVKTVHTDNGSNFTSNTVKAACWWAGIKQEFGI PYNPQS QGVVESMNNKQLKQIIGQVR
DQAEHLKTAVQMAVFIHNFKRKGIGGYTAGERIVDIATDQITRELQKQITKI QNFVRVY
YRDSREPLWKGP AKLLWKGE GAVVIQD NSDIKVVPRRKAKIIRDYGKQ MAGDDCVASGQD
ED-

>POL-B.syn6.6

FFREDLAFLQ GKAREFSSEQTRAISPTRR-----ELQVWGRDNNSPSEAGA
DR----QGTVS-FNFPQITLWQRPLVTIRIGGQLKEALLDTGADDTVLEEMSLPGRWKPK
MIGGIGGFIKVRQYDQILIEICGHKAVGTVLIGPTPVNI IGRNLLTQIGCTLNFPISPID
TVPVKLKPGMDGPKVKQWPLTEEEKIKALIEICTELENEGKISKIGPENPYNTPIFAIAKKK
DSTKWRKLVD FRELNKR TDQDFWEVQLGIPHP SGLKKKKS VTLVDVGDAYFSIPLDEDFRK
YTAFTIPSINNETPGTRYQYNVLPQGWKGS PAIFQSSMTRILEPFRKQNP DIVIYQYVDD
LYVGS DLEIGQHRTKIEELREHLLRWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPITL
PEKDSWTVNDIQKLVGKLNWASQIYAGIKVKQLCKLLRGTKSLTEVVPLTAEAELELAEN
REILKEPVHGAYYDPSKDLVAEIQKQGLGQWTYQIYQEPFKNLKTGKYAKMRGHTNDVK
QLTEAVQKIATESI VIWGRTPKFKLP IQKETWDAAWTEYWQATWIP EWEFVNT PPLVKLW
YQLEKEPIVGAETFYVDGAANRETRLGKAGYVTDGRGRQKVVTPLDTTNQKTELQAIYLAL
QDSGLEVNIVTDSQYALGIIQAQPDKSESELVSQIIEELIKKEKVYLAWVPAHKGIGGNE
QVDKLV SAGIRRVLFLDGDIDKAQEEHEKYHNNWRAMASDFNIPVVAKEIVASCDKCQLK
GEA IHGQVDCSPGIWQLDCTHLEGKII LVAVHVASGYIEAEI IPTETGQETAYFLLKLAG
RWPVKTIHTDNGRNFTSNSVKAACWWAGIKQEFGI PYNPQS QGVVESMNNRELKKIIGQIR
DQAEHLKTAVQMAVFIHNFKRKGIGGYSAGERIVDIASDIQTKELQKQITKI QNFVRVY
YRDNRDPLWKGP AKLLWKGE GAVVIQD NNDIKVVPRRKVKIIRDYGKQ MAGDDCVASRQD
ED-

>POL-C.syn1.1

FFRENLAFFPQGEAREFPSEQTRANSPTS RANSPTS R-----ELQV--RGDNPRSEAGA
ER----QGT---LNFPQITLWQRPLVSIKVG GQIKEALLDTGADDTVLEEINLP GKWKPK
MIGGIGGFIKVRQYDQILIEICGKKAIGTVLVGPTPVNI IGRNMLTQLGCTLNFPISPIE
TVPVKLKPGMDGPKVKQWPLTEEEKIKALTAICEEMEKEGKITKIGPENPYNTPVFAIAKKK
DSTKWRKLVD FRELNKR TDQDFWEVQLGIPHPAGLKKKKS VTLVDVGDAYFSVPLDEGFRK
YTAFTIPSINNETPGIRYQYNVLPQGWKGS PAIFQSSMTKILEPFRAQNPEI VYQYMDD
LYVGS DLEIGQHRAKIEELREHLLK WGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIQL
PEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGAKALTDIVPLTEAELELAEN
REILKEPVHGVYDPSKDLIAEIQKQGHQWTYQIYQEPFKNLKTGKYAKMRTAHTNDVK

43/109

Fig. 10 cont'd-18

QLTEAVQKIAMESIVIWGKTPKFRLPIQKETWETWWTDYWQATWIPWEFVNTPLVCLW
 YQLEKEPIAGAETFYVDGAANRETKIGKAGYVTDGRQKIVSLTETTNOQKTELQAIQLAL
 QDSGSEVNIVTDSQYALGIIQAQPDKSESELVNQIIIEQLIKKERVYLSWVPAHKGIGGNE
 QVDKLVSSGIRKVLFLDGIDKAQEEHEKYHSNWRAMASEFNLPPIVAKEIVASCDKCQLK
 GEAIHGQVDCSPGIWQLDCTHLEGKIILVAVHVASGYIEAEVIPAETGQETAYYILKLAG
 RWPVKVIHTDNGSNFTSAAVKAACWWAGIQQEFGI PYNPQSQGVVESMNKELKKIIGQVR
 DQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERIIDIIATDIQTKELQKQIIKIQNFRVY
 YRDSRDPWKGPAKLLWKGEAVVIQDNSDIKVVPRRKVKI IKDYGKQMAGADCVAGRQD
 EDQ

>POL-C.syn3.1

FFRENLAFPQGEAREFPPEQTRANSPT-RANSPTS-----KLQV--RGDNPCSEAGA
 ER----QGT---FNFPQITLWQRPLVTIKVGGQVKEALLDTGADDTVLEEINLPGKWKPK
 MIGGIGGFIVKVRQYDQIVIEICGKKAIGTVLIGPTPVNIIGRNMLTQLGCTLNFPISPIE
 TVPVKLPKMGDGPVKIKQWPLTEEKIKALTAICDEMEKEGKIEKIGPENPYNTPIFAIKKK
 DSTKWRKLVDFKELNKRTQDFWEVQLGIPHPAGLKKRSVTVLDVGDAYFSVPLDESFRK
 YTAFTIP SINNETPGIRYQYNVLPQGWKGSIPAIFQSSMTRILEPFRAKNPEIYIYQYMD
 LYIGSDLEIGQHRAKVEELREHLLRWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIQL
 PEKDSWTVNDIQRLVGKLNWASQIYPGIKVRQLCKLLRGTKALTDIVPLTEEALELAEN
 REILKEPVHGVYDPSKDLIAEIQKQGDQWYQIYQESFKNLKTGKYAKMRSATNDVK
 QLTEAVQKIALESIVIWGKAPKFRPLPIQKETWEIWWTDYWQATWIPDWEFVNTPLVCLW
 YQLEKEPIAGAETFYVDGAANRETKIGKAGYVTDGRQKVVTLTETTNOQKTELQAIQLAL
 QDSGLEVNIVTDSQYALGIIQAQPDKSESELVNQIIIEELIKKERVYLSWVPAHKGIGENE
 QVDKLVSNIRKVLFLDGIDKAQEEHEKYHSNWRAMANEFNLPPVVAKEIVASCDKCQLK
 GEAIHGQVDCSPGMWQLDCTHLEGKIVLVAVHVASGYVEAEVIPAETGQETAYFILKLAG
 RWPVKIHTDNGSNFTSNAVKAACWWAGIQQEFGI PYNPQSQGVVESMNKELKKIIGQVR
 EQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERIIDIIASDIQTKELQKQITKIQNFRVY
 YRDSRDPVWKGPAKLLWKGEAVVIQDNGDIKVVPRRKAKI IKDYGKQMAGDDCVAGRQD
 EDQ

>POL-C.syn3.2

FFRENLAFPQGEAREFPSEQTRANSPTSRSPTSRTNSPTSRELQV--RGDNPRSEAGV
 ER----QGT---LNFPQITLWQRPLVSIKVGQIKEALLDTGADDTVLEDINLPGKWKPR
 MIGGIGGFIVKVRQYDQIPIEICGKKAIGTVLVGPTPVNIIRNMLTQLRCTLNFPISPIK
 TVPVKLPKMGDGPVKVQWPLTEEKIKALTEICEEMEKEGKITKIGPENPYNTPVFAIKKK
 DSTKWRKLVDFRELNRKRTQDFWEVQLGIPHPAGLKKKSVTVLDVEDAYFSVPLDEGFRK
 YTAFTIPSTNNETPGIRYQYNVLPQGWKGSIPAIFQCSMTKILEPFRTQNPDIVIYQYMD
 LYVGSDEIGQHRAKIEELRAHLLKWGLTTPDKKHQKEPPFLWMGYELHPDKWTVQPIKL
 PEKDSWTVNDIQKIVGKLNWASQIYPGIKVKQLCKLLRGAKALTDIIPLEEALELAEN
 REILKEPVHGAYDPSKDLVAEIQKQGHQWYQIYQEPYKNLKTGKYAKMRTAHTNDVR
 QLTEAVQKIAQESIVIWGKTPKFRLPIQKETWETWWTDYWQATWIPWEFINTPLVCLW
 YQLEKEPIVGAETFYVDGAANRETKMGKAGYVTDKGRQKIVSLTETTNOQKTELQAIQLAL
 QDSGPEVNIVTDSQYALGIIQAQPDSESELVNQIIIEQLINKERIYLSWVPAHKGIGGNE
 QVDKLVSSGIRKVLFLDGIDKAQEDHEKYHNNWRAMASDFNLPPIVAREIVASCDKCQLK
 GEAMHGQVDCSPGVWQLDCTHLEGKIILVAVHVASGYMEAEVIPAETGQETAYYILKLAG
 RWPVKVIHTDNGSNFTSTAVKAACWWAGIKQEFGI PYNPQSQGVVEAMNKLKKIIEQVR
 DQAEHLKTAVQMAVLIHNFKRKGGIGGYSAGERIVDIIATDIQTRELQKQIIKIQNFRVY
 YRDSRDPWKGPAKLLWKGEAVVIQDNSDIKVVPRRKVKI IKDYGKQMAGADCVASRQD
 ED-

>POL-C.syn3.3

FFRENLAFPQGKAREFPSEQARANSPTSRSPTS-----ELQV--RRDNPRSEAGA
 ER----QGT---LNCPQITLWQRPLVSIKIGGQTREALLDTGADDTVLEEISLPGKWKPK
 MIGGIGGFIVKVRQYDQILIEICGKKAIGSVLVGPTPVNIIGRNLLTQLGCTLNFPISPIE
 TIPVKLPKMGDGPVKVQWPLTEEKIKALTAICEEMEKEGKISKIGPENPYNTPIFAIKKK
 DSTKWRKLVDFRELNRKRTQDFWEIQLGIPHPAGLKKKSVTVLDVGDAYFSVPLDEDFRK
 YTAFTIP SINNATPGIRYQYNVLPQGWKGSPIFQSSMTKILEPFRAQNPEIYIYQYMD
 LYVGSDEIEQHRAKIEELREHLLKWGFTTPDKKHQKEPPFLWMGCELHPDKWTVQPIQL

44/109

Fig. 10 cont'd-19

PEKESWTVNDIQKLVGKLNWASQIYAGIKVKQLCRLLRGAKALTDIVPLTAEAELELAEN
 REILREPVHGVYDPSKELIAEIQKQGQDQWTYQIYQEPFKNLKTGKYAKRRTAHTNDVK
 QLAEAVQKIAMESIVIWGKIPKFRPLPIQKETWEAWWTDYWQATWIPWEFVNTPPVLVCLW
 YQLEKDPIAGVETFYVDGAANRETKLGKAGYVTDGRQKIVSLSETTNQKTELHAIQLAL
 QDSGSEVNIVTDSQYALRIIQAQPKSESEIVNQIIEQLIKKERVYLAWVPAHKGIGENE
 QVDKLVSKGIRKVLFLDGLIEKAQEEHRYHSNWRAMASEFNLPPIVAKEIVASCDKCQLK
 GEATHGQVDCSPGIWQLDCTHLEGKVIIVAVHVASGYIEAEVIPAETGQETAYFLLKLAG
 RWPVKTIHTDNGSNFTSAAVKAACWWAGIHQEFGI PYNPQSQGVVESMNKELKKIIGQVR
 DQAEHLKTAVLMAVFIHNFKRKGGIGDYSAGERIIDIIATDIQTRELQKQIIKIQNFRVY
 YRDNRDPIWKGPAPKLLWKGEAVVLQDNSDIKVI PRKAKIIRDYGKQ MAGADCVAGRQD
 ENQ

>POL-C.syn4.1

FFRENLAFFPQKAREFPSEQARANSPTS RANSPTS R-----ELQV--RRDNPRSEAGA
 ER----QGT---LNL PQITLWQRPLVSIKVGGOIKEALLDTGADDTVLEDINLPGKWKPK
 MIGGIGGFIKVRQYDQIPIEICGKKAIGTVLVGPTPVNIIGRNMLTQLGCTLNFPISPIE
 TIPVKLKPGMDGPKVKQWPLTEEEKIKALTEICEEMEKEGKIEKIGPENPYNTPVFAIKKK
 DSTKWRKLVDFRELNRKTQDFWEVQLGIPHPISGLKKKSVTVLDDVEDAYFSVPLDENFRK
 YTAFTIPSVNNETPGIRYQYNVLPQGWKGS PAIFQCSMTKILEPFRTQNP EIVIYQYMD
 LYVGS DLEIGQHRAKIEKLREHLLKWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIKL
 PEKDSWTVNDIQRLVGKLNWASQIYAGIKVRQLCKLLRGAKALTDIVPLTKEAELELAEN
 REILKEPVHGVYDPSKELIAEIQKQGQDQWTYQIYQEPFKNLKTGKYAKRRTAHTNDVK
 QLAEAVQKITMESIVIWGRTPKFRPLPIQKETWEAWWTDYWQATWIPWEFVNTPPVLVCLW
 YQLEKEPIAEAEETFYVDGAANRETKMGKAGYVTDKGRQKIVSLTETTNQKTELHAIQLAL
 QDSGPEVNIVTDSQYALGIIQAQPKSESELVSQIIEQLINKERIYLSWVPAHKGIGGNE
 QVDKLVSKGIRKVLFLDGLDKAQEEHRYHSNWRAMANEFNLPPIVAREIVASCDKCQLK
 GEATHGQVDCSPGVWQLDCTHLEGKIIIVAVHVASGYVEAEVIPAETGQETAYFLLKLAG
 RWPVKTIHTDNGSNFTSAAVKAACWWAGVQQEFGIPYNPQSQGVVESMNKELKKIIGQIR
 DQAEHLKTAVQMAVFIHNFKRKGGIGDYSAGERIIDIIATDIQTRELQKQIIKIQNFRVY
 YRDSRPVWKGPAPKLLWKGEAVVLQDNSDIKVVPRRKVKI IKDYGKQ MAGADCVAGRQD
 ENQ

>POL-C.syn4.2

FFRENLAFFPEGEAREFPSEQTRANSPT-RANSPTS R-----KLQV--RGDNPRSEAGV
 ER----QGT---LNF PQITLWQRPLVSIKIGGQTREALLDTGADDTVLEEIKLPGNWKPK
 MIGGIGGFIKVRQYDQILIEICGKKAIGTVLIGPTPVNIIGRNLLTQLGCTLNFPISPIK
 TVPVKLKPGMDGPKVKQWPLSEEKIKALTEICEEMEKEGKISKIGPENPYNTPVFAIKKK
 DSTKWRKLVDFRELNRKTQDFWEIQLGIPHPAGLKKKSVTVLDVGDAYFSVPLDEDFRK
 YTAFTIP SINNATPGIRYQYNVLPQGWKGS PSIFQSSMTKILEPFRAKNPEIVIYQYMD
 LYVGS DLEIGQHRAKVEELREHLLKWGLTTPDKKHQKEPPFLWMGYELHPDKWTVQPIQL
 PDKDSWTVNDIQKLVGKLNWASQIYPGIKVKQLCKLLRGTKALTDIVPLTAEAELELAEN
 REILREPVHGVYDPSKDLVAEIQKQGNDQWTYQIYQEPYKNLKTGKYAKMRS AHTNDVK
 QLTEAVQKIALESIVIWGKAPKFRPLPIQKETWEIWWTDYWQATWIPDWEFVNTPPVLVCLW
 YQLEKDPIAGVETFYVDGAANRETKLGKAGYVTDGRQKIVSLSETTNQKTELQAIQLAL
 QDSGLEVNIVTDSQYALGIIQAQPKSESELVNQIIEELIKKEKVYLSWVPAHKGIGENE
 QVDKLVSSGIRKVLFLDGLIEKAQEEHEKYHNNWRAMASEFNLPPIVAKEIVASCDKCQLK
 GEATHGQVDCSPGMWQLDCTHLEGKIIIVAVHVASGYLEAEVIPAETGQDTAYYILKLAG
 RWPVKVIHTDNGTNFTSAAVKAACWWAGIQQEFGIPYNPQSQGVVESMNKELKKIIGQVR
 EQAEHLKTAVLMAVFIHNFKRKGGIGEYSAGERIIDMIATDIQTRELQNKQITKIQNFRVY
 YRDSRPDIWKGPAPKLLWKGEAVVIQDNGDIKVVPRRKVKI IRDYGKQ MAGDDCVAGRQD
 EDQ

>POL-C.syn4.3

FFRENLAFFPQGEAREFPPEQTRANSPTSRTNSPTS R-----ELQV--RGDNPHSEAGA
 ERQGT LQGT---LNC PQITLWQRPLVSI RVGGQIKEALLDTGADDTVLEEISLPGKWKPK
 MIGGIGGFIKVRQYDQIVIEICGKKAIGSVLVGPTPVNIIRRNMLTQLRCTLNFPISIE
 TVPVKLKPGMDGPRVKQWPLTEEEKIKALTAICEEMEKEGKITKIGPDNPNYNTPVFAIKKK

45/109

Fig. 10 cont'd-20

DSTKWRKLVDFKELNKRTQDFWEVQLGIPHPAGLKRKKS SVTVLDVGDAYFSVPLDEGFRK
 YTAFTIPSTNNETPGIRYQYNVLPQGWKGSPAIFQSSMIKILEPFRAQNPDIVIIYQYMDD
 LYIGSDLEIGQHRAKIEELRAHLLKWGFTTPDKKHQKEPPFLWMGCELHPDKWTVQPIQL
 PEKESWTVNDIQRLVGKLNWASQIYSGIKVRQLCKLLRGVKALTDIVPLTEEALELAEN
 REILKETVHGAYYDPSKDLIAEIQKQGHQDQWYQIYQEPFKNLKTGKYAKMRTAHTNDIK
 QLTEAVQKIAMESIVIWGKTPKFRLPIQKETWETWWTDYWQATWIPWEFINTPPLVKLW
 YQLEKEPIVGAETFYVDGAANRDTKLKAGYITDRGRQKVITLTETTNQKAEQAIQLAL
 QDSGSKVNIVTDSQYALGIIQAQPDSESELVNQIIIEQLIKKERVYLSWVPAHKGIGGNE
 QIDKLVSSGIRRVFLFDGIDKAQEDHEKYHSNWRAMASDFNLPPIVAKEIIASCDKCQLK
 GEAMHGQVDCSPGIWQLDCTHLEGKVIIVAVHVASGYIEAEVIPAETGQETAYYILKLAG
 RWPVKIIHTDNGSNFTSNAVKAACWWAGIHQEFPIPNPQSQGVVEAMNKLKIIIGQVR
 DQAEHLKTAVLMAVFIHNFKRGGIGGYSAGERIIDIIASDIQTKELQKQITKIQNFRVY
 YRDSRDPIWKGPALLWKGEAVVIQDNSEIKVIPRRKAKIIKDYGKQMGADCVAGGQD
 EN-

>POL-C.syn4.4

FFRENLAFFQGEAREFPSEQTRAISPTSR-----ELQV--RGDNPCSEAGA
 ER----QGT---FNFPQITLWQRPLVTIKVGGQVKEALLDTGADDTVLEEINLPGKWKPR
 MIGGIGGFIIKVRQYEQILIEICGKRAIGTVLVGPTPINIIGRNMLTQIGCTLNFPISPIE
 TVPVKLKPGMDGPKIKQWPLTEEEKIKALTAICDEMEKEGKITKIGPENPYNTPIFAIKKK
 DSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKRSVTVLDVGDAYFSVPLDESFRK
 YTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTRILEPFRAQNPEIVIIYQYMDD
 LYVGSDEIEQHRAKIEELREHLLRWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIQL
 PEKDSWTVNDIQKIVGKLNWASQIYPGIKVRQLCRLLRGAKALTDIIPLTEEALELAEN
 REILKEPVHGAYYDPSKDLIAEIQKQGDQWYQIYQESFKNLKTGKYAKMRTAHTNDVR
 QLTEAVQKIAQESIVIWGKIPKFRLPIQKDTWETWWTDYWQATWIPDWEFVNTPPLVKLW
 YQLEKEPIAGAETFYVDGAANRETKIGKAGYVTDGRGRQKVITLTETTNQKTELQAIHLAL
 QDSGSEVNIVTDSQYALRIIQAQPDKSESEIVNQIIIEQLINKERVYLSWVPAHKGIGGNE
 QVDKLVSNIGIRKVLFLDGIDKAQEEHEKYHSNWRAMASEFNLPPVVAKEIVASCDKCQKQK
 GEAIHGQVDCSPRIWQLDCTHLEGKVIIVAVHVASGYMEAEVIPAETGQETAYFILKLAG
 RWPVKVIHTDNGSNFTSTAVKAACWWAGIKQEFPIPNPQSQGVVESMNKLKIIIEQVR
 DQAEHLKTAVQMAVLIHNFKRKGIGGYSAGERIVDIIATDIQTKELQKQIIKIQNFRVY
 YRDNRPDIWKGPALLWKGEAVVIQDNSEIKVVP RRKAKIIRDYGKQMGADCVASRQD
 ED-

>POL-C.syn6.1

FFRENLAFFQGEAREFPSEQTRAISPTSR-----ELQV--RGNNPRSEAGA
 ER----QGT---LNLFPQITLWQRPLVSIKIGGQTREALLDTGADDTVLEEIKLPGNWKPK
 MIGGIGGFIIKVRQYDQILIEICGKRAIGTVLVGPTPVNIIGRNMLTQIGCTLNFPISPIE
 TVPVKLKPGMDGPKVKQWPLTEEEKIKALTEICEEMEKEGKITKIGPDNPYNTPVFAIKKK
 DSTKWRKLVDFKELNKRTQDFWEVQLGIPHPAGLKKKRSVTVLDVGDAYFSVPLDESFRK
 YTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTRILEPFRAQNPDIVIIYQYMDD
 LYVGSDEIEQHRAKIEELREHLLRWGFTTPDKKHQKEPPFLWMGCELHPDKWTVQPIQL
 PEKDSWTVNDIQKLVGKLNWASQIYSGIKVRQLCKLLRGVKALTDIVPLTKEAELELAEN
 REILREPVGHVYYDPAKDLIAEIQKQGDQWYQIYQEPFKNLKTGKYAKRRTAHTNDVK
 QLTEAVQKIATESIVIWGKIPKFRLPIQKDTWETWWTDYWQATWIPDWEFVNTPPLVKLW
 YQLEKEPIAEAEETFYVDGAASRETKMGKAGYVTDGRGRQKVITLTETTNQKTELQAIKLAL
 QDSGSEVNIVTDSQYALGIIQAQPDKSESEIVNQIIIEQLINKERVYLSWVPAHKGIGGNE
 QVDKLVSRGIRKVLFLDGIDKAQDEHEKYHSNWRAMASEFNLPPIVAREIVASCDKCQLK
 GEATHGQVDCSPGIWQLDCTHLEGKIIIVAVHVASGYIEAEVIPAETGQETAYYILKLAG
 RWPVKIIHTDNGSNFTSSAVKAACWWAGIQEFPIPNPQSQGVVESMNKLKIIIGQVG
 DQAEHLKTAVQMAVFIHNFKRKGIGGYSAGERIIDIIATDIQTKELQKQIIKIQNFRVY
 YRDSRDPIWKGPALLWKGEAVVIQDNSEIKVVP RRKVKIIRDYGKQMGADCMASRQD
 ED-

>POL-C.syn6.2

FFRENLAFFQGEARELPSEQTRANGPTSR-----ELQV--RGDNPCSEAGA
 ER----QGT---FNFPQITLWQRPLVTIKVGGQVKEALLDTGADDTVLEEINLPGKWKPK

46/109

Fig. 10 cont'd-21

MIGGIGGFIKVRQYDQIPIEICGKRAIGTVLVGPTPINIIGRNMLTQLGCTLNFPISPIE
 TVPVQLKPGMDGPRVKQWPLTEEEKIKALTEICKEMEKEGKISKIGPENPYNTPIFAIKKK
 DSTKWRKLVDRELNKRTQDFWEVQLGIPHPAGLKKKRSVTVLDVGDAYFSVPLDEDFRK
 YTAFTIP SINNETPGVRYQYNVLPQGWKGSPSIFQSSMTKILEPFRTQNPEIIVYQYMD
 LYIGSDLEIGQHREKIEELREHLLKWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIKL
 PEKDSWTVNDIQRLVGKLNWASQIYAGIKVRQLCKLLKGAKALTDIVTLTEEALELAEN
 REILKEPVYGVYDPSKDLVAEIQKQGNQWQTYQIYQESFKNLKTGKYAKMRTAHTNDIK
 QLTEAVQKIAQESIVIWGKTPKFRLPIQKETWEAWWTDYWQATWIPDWEFVNTPPVLKWL
 YQLEKEPMAGVETFYVDGAANRETKIGKAGYVTDGRGRQKVVTITETTNQKTELQAIYLAL
 QDSGSKVNIVTDSQYALGIIQAQPKSESELVSQIIEQLINKEKIYLSWVPAHKGIGGNE
 QVDKLVSSGIRKVLFLDGDIDKAQEDHERYHSNWRAMASDFNLPPIVAKEIVASCDQCQLK
 GEAMHGQVDCSPGMWQLDCTHLEGKIIIVAVHVASGYIEAEVISAETGQETAYYILKLAG
 RWPVKVHTDNGSNFTSAVKAACWWAGVQQEFGIPYNPQSQGVVESMNKELKRIIGQVR
 DQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERIIDMIATDIQTKELQKQIIQIQNFRVY
 YRDSRDPIWKGPALLWKGEAVVIQDKGDIKVVPRRKAKIIRDYGKQ MAGADCMAGRQD
 EDQ

>POL-C.syn6.3

FFREDLAFPPQGEARKFPPEQTRANSPTSRSR-----KLQV--RGDNPRSEAGV
 ER----QGT---LNFPQITLWQRPLVSIKVGQIREALLDTGADDTVLEEMSLPGKWKPK
 MIGGIGGFIKVKQYEQILIEICGKKAIGSVLVGPTPVNIIGRNLLTQLGCTLNFPISPIE
 TVPVKLKPGMDGPKVKQWPLTEEEKIKALTAICDEMEKEGKITKIGPENPYNTPVFAIKKK
 DSTKWRKLVDRELNKRTQDFWEVQLGIPHPAGLKKKSVTVLDVGDAYFSVPLDENFRK
 YTAFTIPSRNNETPGIRYQYNVLPQGWKGSPAI FQASMTKILEPFRAKNPEIIVYQYMD
 LYVGSDLEIGQHRAKIEELRDHLLKWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIEL
 PEKDSWTVNDIQKLVGKLNWASQIYPGIQVKQLCKLLRGAKALTDVVPPLTEEALELAEN
 REILKEPVHGAYYDPSKDLIAEIQKQGHQWQTYQIYQEPYKNLKTGKYAKRRAHTNDVK
 QLTEAVQKIAMESIVIWGKTPKFRLPIQKETWETWWTEYWQATWIPWEFVNTPPVLKWL
 YQLEKEPIAGAETFYVDGAANRETKMGKAGYITDRGRQKIIISLTETTNQKTELHAIQLAL
 QDSGSEANIVTDSQYALGIIQAQPDSESELVNQIIEQLIKKERVYLAWVPAHKGIGENE
 QVDKLVSSGIRKILFLDGDIDKAQEEHEKYHSNWKAMASEFNLPPVAREIVASCDKCQLK
 GEAMHGQVDCSPRIWQLDCTHLERKVLVAVHVASGYMEAEVIPAETGQETAYFILKLAG
 RWPVKVIHTDNGSNFTSNAVKAACWWAGIHQEFGIPYNPQSQGVVESMNKELKKIIEQVR
 DQAEHLKTAVQMAVLIHNFKRKGGIGGYTAGERIIDIATDIQTKELQNQITKIQNFRVY
 YRDNRPDIWKGPALLWKGEAVVLQDNSDIKVVPRRKVKIIRDYGKQ MAGADCVAGRQD
 ED-

>POL-C.syn6.4

FFRKNLAFPPQGEAREFPPEQTRANSPTSRSR-----ELQV--RGDNPLSEAGA
 ERQGTQGT---LNCQITLWQRPLVSIKVGQIKEALLDTGADDTVLEEISLPGKWKPK
 MIGGIGGFIKVRQYDQIVIEICGKKAIGAVLVGPTPVNIIRNMLTQLRCTLNFPISPIK
 TVPVKLKPGMDGPKVKQWPLSEEKIKALTAICDEMEKEGKITKIGPENPYNTPIFAIKKK
 DSTKWRKLVDRELNKRTQDFWEVQLGIPHPAGLKKKSVTVLDVEDAYFSVPLDEGFRK
 YTAFTIP SINNATPGIRYQYNVLPQGWKGSPAI FQSSMTKILEPFRTKNPDIVYQYMD
 LYVGSDLEIGQHRAKIEKLREHLLRWGLTTPDKKHQKEPPFLWMGYELHPDKWTVQPIQL
 PDKDSWTVNDIQKLVGKLNWASQIYAGIKVKQLCRLLRGAKALTDIIPLTEEALELAEN
 REILKEPVHGAYYDPSKDLIAEIQKQGGQWQTYQIYQEPYKNLKTGKYAKMRTAHTNDVK
 QLAEAVQKITMESIVIWGRTPKFRLPIQKETWETWWTDYWQATWIPWEFINTPPVLKWL
 YQLEKEPIVGAETFYVDGAANRETKLGKAGYVTDKGRQKIVSLTETTNQKTELQAIHLAL
 QDSGSEVNIVTDSQYALRIIQAQPKSESELVNQIIEQLINKERIYLSWVPAHKGIGGNE
 QVDKLVNSGIRKVLFLDGDIEKAQEEHEKYHSNWRAMANEFNLPPIVAREIVASCDKCQIK
 GEAMHGQVDCSPGVWQLDCTHLEGKVLVAVHVASGYVEAEVIPAETGQEAAYFILKLAG
 RWPVKTIHTDNGSNFTSTAVKAACWWAGIKQEFGIPYNPQSQGVVESMNKELKKIIGQVR
 DQAEHLKTAVQMAVLIHNFKRKGGIGDYAGERIIDIIATDMQTKELQKQIIKVQNFRVY
 YRDSRDPIWKGPALLWKGEAVVIQDNGDIKVVPRRKVKI IKDYGRQ MAGADCVASRQD
 ED-

47/109

Fig. 10 cont'd-22

>POL-C.syn6.5

FFRENLAFFPEGEAREFPSEQARANSPTSR-----ELQV--RRDNPRSEAGA
 EG----QGT---LNFPQITLWQRPLVSIRVGGQIKEALLDTGADDTVLEEINLPGRWKPK
 MIGGIGGFIKVRQYDQITIEICGKKAIGTVLVGPTPINIIGRNMLTQIGCTLNFPISPIE
 TVPVKCLKPGMDGPKIKQWPLTEEEKIKALKAICEEMEKEGKIEKIGPENPYNTPVFAIKKK
 DSTKWRKLVDRELNKRTQDFWEIQLGIPHPAGLKRKKS SVTVLDVGDAYFSVPLYEDFRK
 YTAFTIPSVNNETPGIRYQYNVLPQGWKGS PAIFQCSMTKILEPFRAQNPEIIVYQYMDD
 LYVGS DLEIGQHRAKVEELREHLLKWGLTTPDKKHQKEPPFLWMGYELHPDKWTVQPIQL
 PEKESWTVNDIQKIVGKLNWASQIYPGIKVRQLCRLLRGAKALTDIVPLTAEAELELAEN
 REILKEPVHGVYDPSKELIAEIQKQGDDQWTYQIYQEPFKNLKTGKYAKMRSATNDVK
 QLTEAVQKIALESIVIWGKAPKFRPLPIQKETWEIWWTDYWQATWIPEWEFINTPPLVKLW
 YQLEKEPIAGVETFYVDGAANRDTKIGKAGYVTRDRGRQKIVSLSETTNQKTELQAIQLAL
 QDSGLEVNIVTDSQYALGIIQAQPDNSESELVNQIIIEELIKKERVYLSWVPAHKGIGGNE
 QVDKLVSKGIRKVLFLDGDIDKAQEEHEKYHNNWRAMASDFNLPPVVAKEIVACCDKCQLK
 GEAIHGQVDCSPGIWQLDCTHLEGKVIIVAVHVASGYLEAEVIPAETGQETAYFLLKLAG
 RWPVKVIHTDNGPNFTSAAVKAACWWAGINQEFGI PYNPQSQGVVESMNKELKKIIGQVR
 EQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERIVDIIATDIQTKELQNQIIKIQNFRVY
 YRDSRDPWKGPAKLLWKGEAVVIQENSDIKVVP RRKAKIIKDYGKQ MAGDDCVAGRQD
 EDQ

>POL-C.syn6.6

FFRENLAFFQGEAREFPSEQTRANSPT-RANSPTSRTNSPTSRELQV--RGDNPHSEAGA
 ER----QGS---LNFPQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEDINLP GKWKPK
 MIGGIGGFIKVRQYEQIPIEICGKKAIGTVLIGPTPVNIIGRNLLTQLGCTLNFPISPIE
 TIPVKCLKPGMDGPKVKQWPLTEEEKIKALTAICEEMEKEGKITKIGPDNPYNTPVFAIKKK
 DSTKWRKLVDRELNKRTQDFWEVQLGIPHPAGLKKKKS SVTVLDVGDAYFSVPLDKDFRK
 YTAFTIPSTNNETPGIRYQYNVLPQGWKGS PAIFQSSMIKILEPFRAKNPELVIYQYMDD
 LYVGS DLEIMQHRAKIEELRAHLLKWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIVL
 PEKDSWTVNDIQKLVGKLNWASQIYPGIKVKQLCKLLRGTKALTDIVPLTEEALELAEN
 REILKETVHGAYDPSKDLIAEIQKQGYDQWTYQIYQEPFKNLKTGKYAKKRTAHTNDVR
 QLTEAVQKIAIESIVIWGKTPKFRPLPIQKETWETWWADYWQATWIPEWEFVNTPPLVKLW
 YQLEKDPIAGAETFYVDGAANRET KKGKAGYVTDKGRQKVVTLTETTQKAELOAIQLAL
 QDSGPEVNIVTDSQYALRIIQAQPDKSESGLVNQIIIEQLIKKEKVYLSWVPAHKGIGGNE
 QIDKLVSSGIRRVFLDGDIDKAQEDHEKYHSNWRAMAGEFNLPPVVAKEIVASCDKCQQK
 GEAIHGQVDCSPGIWQLDCTHLEGKIIILVAVHVASGYIEAEVIPAETGQDTAYYILKLAG
 RWPVKVIHTDNGTNFTSAAVKAACWWASIQQEFGIPYNPQSQGVVEAMNKLKKIIGQIR
 DQAEHLKTAVLMAVFIHNFKRKGGIGEYSAGERIIDIIASDIQTKELQKQITKIQNFRVY
 YRDSRDPVWKGPAKLLWKGEAVVIQDNSDIKVI PRKAKIIRDYGKQ MAGADCVAGGQD
 ED-

>POL-M.syn1.1

FFRENLAFFQGEAREFPSEQTRANSPTS RANSPTS R-----ELQV--RGDNPRSEAGA
 ER----QGT---LNFPQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEDINLP GKWKPK
 MIGGIGGFIKVRQYDQILIEICGKKAIGTVLVGPTPVNIIGRNMLTQIGCTLNFPISPIE
 TVPVKCLKPGMDGPKVKQWPLTEEEKIKALTEICTEMEKEGKISKIGPENPYNTPVFAIKKK
 DSTKWRKLVDRELNKRTQDFWEVQLGIPHPAGLKKKKS SVTVLDVGDAYFSVPLDESFRK
 YTAFTIP SINNETPGIRYQYNVLPQGWKGS PAIFQSSMTKILEPFRKNPDIVYQYMDD
 LYVGS DLEIGQHRAKIEELREHLLKWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIQL
 PEKDSWTVNDIQKLVGKLNWASQIYPGIKVKQLCKLLRGAKALTDIVPLTEEALELAEN
 REILKEPVHGVYDPSKDLIAEIQKQGQDQWTYQIYQEPFKNLKTGKYAKMRTAHTNDVK
 QLTEAVQKIATESIVIWGKTPKFRPLPIQKETWETWWTDYWQATWIPEWEFVNTPPLVKLW
 YQLEKEPIVGAETFYVDGAANRET KLGKAGYVTRDRGRQKVVSLETETTQKTELQAIQLAL
 QDSGSEVNIVTDSQYALGIIQAQPDKSESELVNQIIIEQLIKKEKVYLSWVPAHKGIGGNE
 QVDKLVSSGIRKVLFLDGDIDKAQEEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKCQLK
 GEAMHGQVDCSPGIWQLDCTHLEGKVIILVAVHVASGYIEAEVIPAETGQETAYFILKLAG
 RWPVKVIHTDNGSNFTSAAVKAACWWAGIQQEFGIPYNPQSQGVVESMNKELKKIIGQVR
 DQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERIIDIIATDIQTKELQKQITKIQNFRVY

48/109

Fig. 10 cont'd-23

YRDSRDPIWKGPALLWKGEAVVIQDNSDIKVVPRRKAKIIRDYKQMGAGDDCVAGRQD
EDQ

>POL-M.syn3.1

FFRENLAFFPQGEAREFPSEQTRANSPTSRSNSPTS-----ELQV--RGDNPRSEAGA
ER----QGT---LNFQITLWQRPLVSIKVGGOIKEALLDTGADDTVLEEINLPGKWKPK
MIGGIGGFIKVRQYDQIPIEICGKRAIGTVLVGPTPINIIGRNMLTQLGCTLNFPISPIK
TVPVKLKPMDGPKVKQWPLTEEEKIKALTAICEEMEKEGKITKIGPENPYNTPVFAIKKK
DGTKWRKLVDFRELNRKTQDFWEVQLGIPHPGSLKKKKSVSVDVGDAYFSVPLDEFRK
YTAFTIPSINNETPGIRYCYNVLPQGWKGSPAIFQCSMTKILEPFRAQNPEIVYQYMDD
LYIGSDLEIGQHRAKIEELREHLLRWGFTTPDKKHQKEPPFLWMGCELHPDKWTVQPIQL
PEKDSWTVNDIQKIVGKLNWASQIYPGIKVRQLCKLLRGAKALTDIVPLTEEALELAEN
REILREPVGHVYDPSKDLVAEIQKQGQDQWYQIYQEPFKNLKTGKYAKMRTAHTNDVK
QLTEAVQKIALESIVIWGKIPKFRLPQKETWEAWWMEYWQATWIPWEFINTPPLVKLW
YQLEKEPIAGAETFYVDGAANRETKIGKAGYVTDGRGRQKIVSLTETTQKAELOAIQLAL
QDSGPEVNIVTDSQYALGIIQAHPDKSESELVNQIIIEQLIKKERVYLSWVPAHKGIGENE
QVDKLVSNIGIRKILFLDGIDKAQEEHEKYHSNWRAMASEFNLPPIVAKEIVASCNKCQLK
GEALHGQVDCSPGMWQLDCTHLEGKVLVAVHVASGYMAEVI PAETGQETAYYILKLAG
RWPVKVVHTDNGSNFTSTAVKAACWWAGIQQEFGI PYNPQSQGVIESMNKELKKIIGQIR
DQAEHLKTAVQMAVFIHNFKRKGIGEYSAGERIIDIIATDIQTRELQKQIIKIQNFRVY
YRDSRDPIWKGPALLWKGEAVVLQDNSDIKVVPRRKVKI IKDYKQMGAGADCVAGRQD
ENQ

>POL-M.syn3.2

FFRENLAFFQGEARKFSSEQTGANSPTS-----ELRV--RRGDNPLSEAGA
ER----RGTVP SLSFPQITLWQRPLVTVKIGGQLIEALLDTGADDTVLEDINLPGKWKPR
MIGGIGGFIKVKQYDQILIEICGKKAIGTVLVGPTPVNIIGRNMLTQIGCTLNFPISPID
TVPVTLKPGMDGPKIKQWPLTEEEKIKALTEICTEMEKEGKISRIGPENPYNTPIFAIAKKK
DSTKWRKLVDFRELNRKTQDFWEVQLGIPHPAGLKKKRSVTVLDVGDAYFSVPLDKDFRK
YTAFTIPSTNNETPGIRYQYNVLPQGWKGSPSIFQSSMTRILEPFRAKNPEIVYQYMDD
LYVGSDEIEQHRTKIEELRQHLLRWGLTTPDKKHQKEPPFLWMGYELHPDRWTVQPIEL
PEKESWTVNDIQKLVGKLNWASQIYAGIKVKQLCKLLRGTKALTEVVPLTEEALELEEN
REILKDPVHGAYYDPSKDLIAEIQKQGHQWYQIYQEQYKNLKTGKYARKRSAHTNDVR
QLTEAVQKIATESIVIWGKTPKFRLPQRETWETWWTWQATWIPWEFVNTPLVKLW
YQLEKDPIAGVETFYVDGASNRETKKKGAGYVTDKGRQKVVSLETTNQKTELHAIHLAL
QDSGSEVNIVTDSQYALGIIQAQPDSESELVNQIIIEELIKKEKVYLSWVPAHKGIGGNE
QVDKLVSSGIRKVLFLDRIDKAQEEHRYHSNWRMTASDFNLPPIVAKEIVANCDKCQLK
GEAMHGQVDCSPGIWQIDCTHLEGKVIIVAVHVASGYIEAEVI PAETGQETAYFLLKLAG
RWPVKVIHTDNGSNFTSAAVKAACWWANVTQEFGI PYNPQSQGVVESMNKELKKIIGQVR
EQAEHLKTAVQMAVLIHNFKRKGIGGYSAGERIIDIIASDIQTKELQKQIIKVQNFRVY
YRDSRDPVWKGPALLWKGEAVVIQDNSDIKVVPRRKVKIIRDYKQMGAGDDCVAGRQD
EDQ

>POL-M.syn3.3

FFREDLAFFPQKAREFSSEQTRANSPTR-----ELQVWGRDNNSLSEAGA
DR----QGTVS-FSFPQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMNLPGRWKPK
MIGGIGGFIKVRQYDQILIEICGHKAIGTVLIGPTPVNIIGRNLLTQIGCTLNFPISPIE
TVPVKLKPMDGPRVKQWPLTEEEKIKALVEICTEMEKEGKISKIGPENPYNTPVFAIKKK
DSTRWRKLVDFRELNRKTQDFWEIQLGIPHPAGLKKKKSVTVDVGDAYFSVPLDEGFRK
YTAFTIPSVNNETPGVRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQNPDIVYQYMDD
LYVGSDEIEQHRTKIEELREHLLKWGFTTPDKKHQNEPPFLWMGYELHPDKWTVQPIVL
PEKDSWTVNDLQKLVGKLNWASQIYPGIKVKQLCRLLRGAKALTEVIPLTKEAELELAEN
REILKEPVHGHVYDPSKELIAEIQKQGQGWYQIYQEPYKNLKTGKYARMRGAHTNDVK
QLTEVVQKIAMESIVIWGKTPKFKLPQKETWETWWTWYQATWIPDWEFVNTPLVKLW
YQLEKEPIVGAETFYVDGAANRETKLGKAGYVTDGRGRQKVVSLETTNQKTELQAIHLAL
QDSGLEVNIVTDSQYALGIIQAQPDSESELVSQIIIEQLIKKEKVYLAWVPAHKGIGGNE
QVDKLVSAGIRKVLFLDGIDKAQEDHEKYHNNWRAMASDFNLPVVAKEIVASCDKCQLK

49/109

Fig. 10 cont'd-24

GEAIHGQVDCSPGIWQLDCTHLEGKIIILVAVHVASGYLEAEVIPAETGQETAYFILKLAG
RWPVKTIHTDNGSNFTSTTVKAACWWAGIKQEFPIPNPQSQGVVESINKELKKIIGQVR
DQAEHLKTAVLMAVFIHNFKRRGGIGGYSAGERIVDIIATDIQTKELQKQITKIQNFRVY
YRDSRDPLWKGPALLWKGEAVVIQDNSEIKVVP RRKAKIIRDY GKQMAGDDCVASRQD
ED-

>POL-M. syn4.1

FFRENLAFFQGEARKFSSEQTRANSPTRG-----ELQVWGRDNNPLSEAGA
ER----RGTVP SLSFPQITLWQRPLVTVKIGGQLIEALLDTGADDTVLEDINLPGKWKPK
MIGGIGGFIKVKQYDQILIEICGKKAIGTVLVGPTPVNIIGRNMLTQIGCTLNFPISPID
TVPVTLKPGMDGPRIKQWPLTEEEKIKALTEICKEMEEEGKISKIGPENPYNTPIFAIKKK
NSTRWRKLVDFRELNKKTQDFWEVQLGIPHPAGLKRKKS SVTVLDVEDAYFSVPLDESFRK
YTAFTIPSINNETPGVRYQYNVLPQGWKGS PAIFQCSMTKILEPFRIKNPEMVIYQYMD
LYVGS DLEIGQHRIKIEELRAHLLSWGFTTPDKKHQKDPFLWGMGYELHPDRWTVQPIEL
PEKDSWTVNDIQKLVEKLNWASQIYSGIKVRQLCRLLRGAKALTDIVPLTEEALELAEN
REILKEPVHGAYYDPSKDLVAEIQKQGQDQWTYQIYQEPFKNLKTGKYARKRSAHTNDVK
QLTEVVQKIATESIVIWGKTPKFRLP IQRETWETWWTEYWQATWIPWEFVNT PPLVKLW
YQLEKDPIAGVETFYVDGAASRETKLGKAGYVTDGRQKVVS LTETTNQKTELHAIHLAL
QDSGSEVNIVTDSQYVLGIIQAQPD RSESELVNQII EELIKKEKVYLSWVPAHKGIGGNE
QVDKLVS SGIKRVLFNLGIDKAQEEHERYHSNWRTMASDFNL PPIVAKEIVANCDKCQLK
GEAMHGQVDCSPGVWQLDCTHLEGKVIIVAVHVASGYIEAEVIPAETGQEAAYFILKLAG
RWPVKVHTDNGSNFTSAAVKAACWWANVRQEFPIPNPQSQGVVESMNNELKKIIGQIR
DQAEHLKTAVLMAVFIHNFKRRGGIGGEYSAGERIIDIIATDIQTRELQKQITKIQNFRVY
FRDSRDPIWKGPALLWKGEAVVIQDNSEIKVVP RRKVKIIRDY GKQMAGDDCVAGRQD
EN-

>POL-M. syn4.2

FFRENLAFFPQGEAREFPSEQARANSPTS RANSPTS R-----DLWDGGRDNL P-SEAGA
ER----QGT---LNFPQITLWQRPLVTVRIGGQLREALLDTGADDTVLEDIDLPGKWKPK
IIGGIGGFIKVRQYEQIPIEICGHKAIGTVLVGPTPINIIGRNMLTQLGCTLNFPISP I K
TVPVKLKPGMDGPRVKQWPLTEEEKIKALTAICEEMEKEGKITKIGPENPYNTPVFAIKKK
DSTRWRKLVDFRELNRR TQDFWEVQLGIPHPAGLKKKRSVTVLDVGDAYFSVPLDEGFRK
YTAFTIPSVNNETPGIRYQYNVLPQGWKGS PSI FQSSMTRILEPFRAKNPEI VIYQYIDD
LYVRS DLEIGQHRAKIEELREHLLRWGFTTPDKKHQKEPPFLWGMGCELHPDKWTVQPIQL
PEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGTKALTDIVTLTEEALELAEN
REILKDPVHG VYYDPSKELIAEIQKQGDDQWTYQIYQE QYKNLKTGKYAKRRTAHTNDVR
QLTEAVQKIALESIVIWGKIPKFRLP IQKETWEAWWMEYWQATWIPWEYVNT PPLVKLW
YQLEKEPIIGAETFYVDGAANRETKLGKAGYVTNRGRQKVVS LDTTNQKTELQAIQLAL
QDSGSEVN VVTD S QYALGIIQAHPDKSESELVNQII EQLIKKERVYLSWVPAHKGIGGNE
QVDKLVSAGIRKVLFLDGDIDKAQEDHERYHSNW RAMASDFNL P PVVAKEIVASCNKCQLK
GEAIHGQVDCSPGIWQLDCTHLEGKIIIVAVHVASGYMEAEVIPAETGQETAYFILKLAG
RWPVKIIHTDNGSNFTSATVKAACWWANVTQEFPIPNPQSQGVVESINKELKKIIGQVR
DQAEHLRTAVQMAVFIHNFKRRGGIGGYSAGERIVDIIATDIQTKELQKQITKI QKFRVY
YRDSRDPLWKGPALLWKGEAVVIQDN S DIKVI PRRKAKI IKDY GKQMAGADCVAGRQD
EDQ

>POL-M. syn4.3

FFRENLAFFPQ GKAREFPSEQTRANSPTS RANSPTS R-----ELQV--RGDNPRSEAGA
ER----QGT---FNFPQITLWQRPLVSIKVG GQIKEALLDTGADDTVLEEINLPGKWKPR
MIGGIGGFIKVRQYDQILIEICGKRAIGTVLVGPTPANIIGRNLLTQLGCTLNFPISP I E
TVPVKLKPGMDGPKIKQWPLTEEEKIKALTEICTEMEKEGKISKIGPENPYNTPIFAIKKK
DSTKWRKLVD FKELNKRTQDFWEVQLGIPHP SGLKKKKS VTVLDVGDAYFSVPLDEDFRK
YTAFTIPSTNNETPGIRYQYNVLPQGWKGS PAIFQASMTKILEPFRAQNPEI VIYQYMD
LYVGS DLEIEQHRAKVEELREHLLKWGFTTPDKKHQNEPPFLWGMGYELHPDKWTVQPIKL
PEKESWTVNDIQKLVGKLNWASQIYPGIKVKQLCRLLRGTKALTEVIPLTKEAELELAEN
REILREP VHG VYYDPTKDLIAEIQKQGH DQWTYQIYQEPHKNLKTGKYAKMRTAHTNDVK
QLAEAVQKIAMESIVIWGKIPKFKLP IQKETWETWWTDY WQATWIPDWEFVNT PPLVKLW

50/109

Fig. 10 cont'd-25

YQLEKEPIAGAETFYVDGAANRETKIGKAGYVTDGRGRQKIVSLTETTNQKAELQAIQLAL
 QDSGPEVNIVTDSQYALGIIQAQPKDSESEIVNQIEKLIKEDKVYLSWVPAHKGIGGNE
 QIDKLVSNGIRKVLFLDGLIEKAQEEHEKYHSNWRAMASEFNLPPIVAKEIVASCDKCQLK
 GEATHGQVDCSPGMWQLDCTHLEGKIIILVAVHVASGYIEAEVPTETGQETAYYILKLAG
 RWPVKVIHTDNGSNFTSTAVKAACWWAGIQQEFGI PYNPQGGVVE SMNKELKKIIGQVR
 EQAEHLKTAVQMAVLIHNFKRKGIGGYSAGERIIDIIASDIQTRELQKQIIKVQNFRVY
 YRDSRDPIWKGP AKLLWKGE GAVVLQDNSDIKVVPRRKVKI IKDYGKQMAGADCVASRQD
 EN-

>POL-M.syn4.4

FFREDLAFPPQ GKAREFSSEQTRANSPTRR-----ELQVWGRDNNLSLSEAGA
 DR----QGTVS-FSFPQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMNLPGRWKPK
 MIGGIGGFIKVRQYDQIPIEICGKKAIGTVLIGPTPVNIIGRNLLTQIGCTLNFPISPIE
 TIPVKLKPMDGPKVKQWPLTEEEKIKALVEICTEMEKEGKISRIGPENPYNTPVFAIKKK
 DGTKWRKLVDFRELNKRTQDFWEIQLGIPHPAGLKKKKS SVSLDVGDAYFSVPLDKDFRK
 YTAFTIP SINNETPGIRYCYNVLPQGWKGS PAIFQSSMTKILEPFRKQNPDIVIYQYMD
 LYIGSDLEIGQHRTKIEELRQHLLRWGLTTPDKKHQNEPPFLWMGYELHPDKWTVQPIVL
 PEKDSWTVNDLQKLVGKLNWASQIYAGIKVKQLCKLLRGAKALTEVVPLTEEALELEEN
 REILKEPVHGVYDPSKDLIAEIQKQGQGWYTYQIYQEPYKNLKTGKYARMGAHTNDVK
 QLTEAVQKIAQECIVIWGKTPKFKLPIQKETWETWMDYQATWIPWEFINTPPLVKLW
 YQLEKEPIVGAETFYVDGASNRETKKGKAGYVTDKGRQKVVTLTETTNQKTELQAIHLAL
 QDSGLEVNIVTDSQYALGIIQAQPKDSESELVSQIIEQLIKKEKVYLA WVP AHKGIGGNE
 QVDKLVSNGIRKILFLDGLDKAQEEHEKYHNNWRAMASDFNIPPVVAKEIVACCDKCQLK
 GEALHGQVDCSPRIWQLDCTHLEGKVIILVAVHVASGYLEAEVIPAETGQETAYFLLKLAG
 RWPVKTIHTDNGSNFTSTTVKAACWWAGIKQEFGI PYNPQSQGVIESMKNELKKIIEQVR
 DQAEHLKTAVQMAVFVHNFKRKGIGGDYSAGERIIDIIISTDIQTRELQKQIIKIQNFRVY
 YRDSRDPVWKGP AKLLWKGE GAVVIQDNNEIKVVPRRKAKI IRDYGKQMAGDDCVASRQD
 ED-

>POL-M.syn6.1

FFREDLAFPPQGEARKFPSEQTRANSPTRG-----ELQVWGRDNNLSLSEAGD
 DR----QGTVS-FNLPQITLWQRPLVTVRIGGQLIEALLDTGADDTVLEDNMLPGWKWKPK
 MIGGIGGFIKVRQYEQIPIEICGHKAIGTVLIGPTPVNIIGRNLLTQIGCTLNFPISPID
 TVPVTLKPGMDGPKIKQWPLTEEEKIKALTEICTEMEKEGKISRIGPENPYNTPVFAIKKK
 NSTRWRKLVDFRELNKRTQDFCEVQLGIPHPAGLKKKRSVTVLDVGDAYFSVPLDENFRK
 YTAFTIP SINNETPGVRYQYNVLPQGWKGS PAIFQASMTKILEPFRTKNP ELVIYQYMD
 LYVGS DLEIEQHRTKIEELRAHLLSWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIEL
 PEKDSWTVNDIQKIVGKLNWASQIYPGIKVRQLCRLLRGT KALTDIVPLTAEAELELAEN
 REILREP VHGVYDPSKELIAEIQKQGHQDQWYTYQIYQDPFKNLKTGKYARKRSAHTNDVR
 QLTEAVQKITTESIVIWGKTPKFKRLPIQRETWEAWWMEYWQATWIPWEFINTPPLVKLW
 YQLEKDPIVGAETFYVDGAASRETKLGKAGYVTNKG RQKVVS LNETTNQKTELHAIHLAL
 QDSGSEANIVTDSQYALGIIQAQPD RSESEVNVN QIIEELIKKEKVYLSWVPAHKGIGGNE
 QVDKLVS SGIRKVLFLDGLDKAQEDHERYHSNWR TMASDFNLPPIVAREIVASCDKCQKQK
 GEAMHGQVDCGPGIWQLDCTHLERKVIILVAVHVASGYIEAEVIPAETGQETAYFVLKLAG
 RWPVKVIHTDNGSNFTSA AVKAACWWANVTQEFGI PYNPQSQGVVE SMNNELKKIIGQVR
 EQAEHLKTAVLMAVFIHNFKRKRIGGYSAGERIVDIIASDIQTRELQNKITKIQNFRVY
 FRDSRDPIWKGP AKLLWKGE GAVVIQDNNDIKVVPRRKVKI IRDYGKQMAGDDCVAGRQD
 EN-

>POL-M.syn6.2

FFREDLAFQQGEARKFSSEQTRANSPTSR-----ELRVWG-GDNTLSETGA
 ER----QGT---LNFQITLWQRPLVTIKVGGQIKEALLDTGADDTVLEDINLP GKWKPR
 MIGGIGGFIKVRQYDQIPIEICGKKAIGSVLVGPTPVNIIGRNMLTQLGCTLNFPISPIK
 TVPVKLKPMDGPKVKQWPLSEEKIKALTAICDEMEKEGKITKIGPDNPNYTPVFAIKKK
 DGTKWRKLVD FKELNKRTQDFWEIQLGIPHPAGLKKKKS SVSLDVGDAYFSVPLDESFRK
 YTAFTIP SLNNETPGIRYCYNVLPQGWKGS PAIFQSSMTKILEPFRAQNPEI VIYQYIDD
 LYVRSDLEIGQHRAKIEELREHLLKWLTT PDKKHQKEPPFLWMGYELHPDRWTVQPIQL

51/109

Fig. 10 cont'd-26

PDKDSWTVNDLQKLVGKLNWASQIYPGIRVKQLCKLLKGAKALTDIVTLTEEALELAEN
 REILKNPVHGVYDPAKDLIAEIQKQGNDQWTYQIYQEPHKNLKTGKYAKMRTAHTNDVK
 QLTEVVQKIAMESIVIWGKVPKFRPLPIQKETWETWWTWTDYWQATWIPDWEFVNTPPLVKLW
 YQLEKEPIAGAETFFYVDGAANRETKMGKAGYVTDGRQKVVSLTETTNQKTELQAIQLAL
 QDSGPEVNIVTDSQYAIGIIQAQPKSESEIVNQIIIEQLIKKERVYLSWVPAHKGIGENE
 QVDKLVSTGIRRVFLDGDIDKAQEEHERYHSNWRAMASEFNLPPIVAKEIVASCDQCQLK
 GEAMHGQVDCSPGVWQLDCTHLEGKIIILVAVHVASGYMEAIEVIPAETGQETAYFILKLAA
 RWPVKVIHTDNGPNFTSATVKAACWWANITQEFGIPYNPQGQGVVESMNKELKKI IKQVR
 DQAEHLKTAVQMAVLIHNFKRKGGIGGYSAGERIIDIIASDIQTKELQKQIIKIQNFRVY
 YRDSRPDIWKGPAKLLWKGEAVVLQDNSDIKVVPRRKVKI IKDYGKQMAGADCVAGGQD
 ED-

>POL-M.syn6.3

FFRENLAFFPQKGAREFPSEQTRAISPTSR-----ELQVWGGDNNLSLSEAGA
 ER----QGTVS-FSFPQITLWQRPIVTIKIGGQLREALLDTGADDTVLEEMNLPGRWKPK
 MIGGIGGFIKVKQYDNILIEICGHKAVGTVLVGPTPANIIGRNLLTQLGCTLNFPISPIE
 TVPVKLKPGIDGPKVKQWPLTEEEKIKALVEICTEMEKEGKISKIGPENPYNTPIFAIKKK
 DSTRWRKLVDFRELNRRRTQDFWEVQLGIPHPAGLKRKKSVTVLVDVGDAYFSVPLDKEFRK
 YTAFTIPSINNETPGIRYQYNVLPQGWKGSPIFQSSMTKILEPFRIKNPEMVIYQYMDD
 LYIGSDLEIGQHRIKIEELREHLLKWGFTTPDKKHQKEPPFLWMGCELHPDKWTVQPIML
 PEKDSWTVNDIQKLVGKLNWASQIYAGIKVRQLCKLLRGTKALTEVVPLTEEALELAEN
 REILKEPVHGAYYDPSKDLIAEVQKQGQDQWTYQIYQEPFKNLKTGKYAKKRS AHTNDVK
 QLTEAVQKIALESIVIWGKAPKFRPLPIQKETWEAWWTEYWQATWVPEWEFVNTPPLVKLW
 YQLETEPIAGAETYYVDGAANRETKLGKAGYVTNRGRQKVVSLTDTTNQKTELQAIHLAL
 QDSGLEVNIVTDSQYALGIIHAQPKSESELVNQIIIEQLINKERIYLSWVPAHKGIGENE
 QVDKLVSKGIRKVLFLDGIIEKAQEEHEKYHSNWKAMASEFNLPPVVAKEIVACCDKCQLK
 GEALHGQVDCSPGMWQLDCTHLEGKIIIVAVHVASGYIEAEVIPTETGQETAYFLLKLAG
 RWPVKTIHTDNGSNFTSTTVKAACWWAGIKQEFGIPYNPQSQGVVESINKELKKIIGQIR
 DQAEHLKTAVLMAVFIHNFKRKGGIGGYTAGERIVDIIATDIIQTKELQKQITKVQNFRVY
 YRDSREPLWKGPAPKLLWKGEAVVIQDNEIKVVPRRKAKILRDYGKQMAGADCVASRQD
 EN-

>POL-M.syn6.4

FFRENLAFFQQGEAREFSSEQTRTNSPTSR-----ELWDGGRDNLPS-SEAGA
 ER----RGTVPSSLSPQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEINLPWKWKPK
 LIGGIGGFIKVRQYDQILIEICGKKAIGTVLVGPTPINIIGRNMLTQIGCTLNFPISPIE
 TIPVKLKPMDGPRVKQWPLTEEEKIKALIEICTEMEKEGKISRVPENPYNTPIFAIKKK
 NSNRWRKLVDFRELNKRTQDFWEVQLGIPHPGGLKKKKSVTILVDVGDAYFSVPLDEDFRK
 YTAFTIPSINNATPGIRYQYNVLPQGWKGSPIFQCSMTKILEPFRKQNP EIIYQYMDD
 LYVRS DLEIGQHRTKIEELRQHLLKWGFYTPDKKHQKEPPFLWMGYELHPDKWTVQPIKL
 PEKESWTVNDIQKLVEKLNWASQIYPGIKVKQLCRLLRGAKALTEVIPLTEEALELEEN
 REILKDPVHGVYDPTKDLIAEIQKQGDDQWTYQIYQEPYKNLKTGKYAKRRTAHTNDVR
 QLTEVVQKVATESIVIWGKIPKFKLPIQKETWEIWWTDYWQATWIPWEFVNTPHLVKLW
 YQLEKEPIIGAETFFYVDGASNRETKKGKAGYVTDGRQKIVSLTETTNQKTELQAIQLAL
 QDSGSEVNIVTDSQYALGIIQAHPDKSESELVSQIIIEQLIKKEKVYLAWVPAHKGIGGNE
 QIDKLVSNIRKILFLDGDIDKAQEEHEKYHNNWRAMASDFNLPPVVAKEIVASCNKCQLK
 GEAIHGQVDCSPRIWQLDCTHLEGKVMVAVHVASGYVEAEVIPTETGQDTAYFILKLAG
 RWPVKVVHTDNGSNFTSAAFKAACWWANVQQEFGIPYNPQSQGVVEAMNKLKKIIEQVR
 DQAEHLKTAVQMAVFVHNFKRKGGIGDY SAGERIIDIIATDIIQTRELOKQIIKIQNFRVY
 YRDNRDPDIWKGPAKLLWKGEAVVIQDNSDIKVI PRKAKIIRDYGKQMAGDDCMAGRQD
 EDQ

>POL-M.syn6.5

FFREDLAFLQGKAREFSSEQTRANSPTRR-----ELQVWGRDSNSLSEAGA
 DR----QGTVS-FNFPQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEDIDLPGWKWKPK
 IIGGIGGFIKVKQYDQILIEICGKRAIGTVLVGPTPVNIIGRNILTQIGCTLNFPISPID
 TVPVKLKPMDGPRIKQWPLTEEEKIKALTEICKEMEEEGKISKIGPENPYNTPVFAIKKK

52/109

Fig. 10 cont'd-27

DSTKWRKVVDFRELNKGTQDFWEVQLGIPHPAGLKQKKSVTVLDVEDAYFSVPLDKDFRK
 YTAFTIPSVNNETPGIRYQYNVLPQGWKGSPAIFQSSMTRILEPFRKQNPDIVIYQYMDD
 LYVGSdleIGQHRTKVEELRQHLLRWGFTTPDKKHQKDPFLWWMGYELHPDKWTVQPIVL
 PEKDSWTINDIQKLVGKLNWASQIYSGIKVRQLCKCLRGTKALTEVIPLTKEAELELAEN
 KEILKEPVHGVYDPSKDLVAEIQKQGQGWYQIYQEQYKNLKTGKYARMGAHTNDVK
 QLAEAVQKIATESIVIWGKIPKFRLP IQRETWETWWTEYWQATWIPWEYVNT PPLVKLW
 YQLEKEPIVGAETFYVDGAANRDTKLKGAGYVTDGRQKVPLTDTTNQKTELQAINLAL
 QDSGSKVNIVTDSQYVLGIIQAQPDRSESEIVNQIIEKLIKDKVYLSWVPAHKGIGGNE
 QVDKLVSAGIRKVLFLDGDIDKAQDEHEKYHSNWRAMASDFNLPPVIAKEIVASCDKCQLK
 GEATHGQVDCSPGIWQLDCTHLEGKVIIVAVHVASGYIEAEVISAETGQETAYIILKLAG
 RWPVKI IHTDNGSNFTSTAVKAACWWAGIQQEFGI PYS PQSQGVVESMNKQLKQIIGQVR
 DQAEQLKTAVQMAVFIHNFKRKGGIGEYSAGERIIDIIISTDIQTRELQKQITKIQNFRVY
 YRDSRDPVWKGPAKLLWKGEAVVIQDNSEIKVVP RRKAKIIRHYGKQ MAGDDCVASRQD
 EDQ

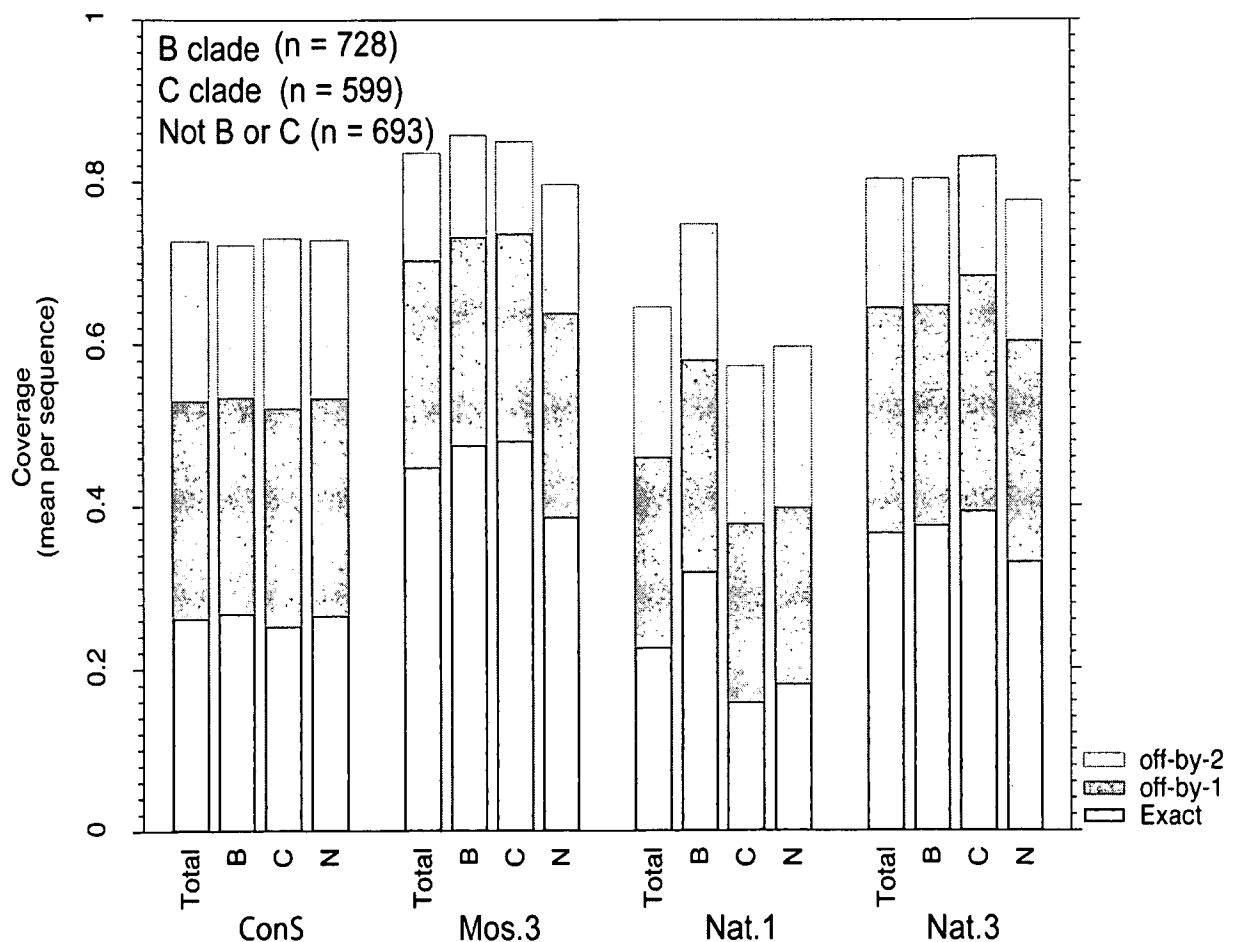
>POL-M.syn6.6

FFRENLAFFPQGEAREFPSEQARANSPTS RANSPTS R-----ELQV--RGDNPRSEAGA
 ERQGT LQGT---LNC PQITLWQRPLVSIKVGQVKEALLDTGADDTVLEEMSLPGKWKPK
 MVGIGGFIKVRQYDQILVEICGHKAIGTVLVGPTPVNI IRRNMLTQLRCTLNFPISPIE
 TVPVT LKPGMDGPKVRQWPLTEEEKIKALTAICEEMEKEGKITKIGPENPYNTPVFAIRKK
 DSTKWRKLVDFRELNKKTQDFWEVQLGIPHPSGLKKKKSVTVLDVGDAYFSVPLDEGFRK
 YTAFTIPSTNNETPGIRYQYNVLPQGWKGSPAIFQSSMIKILEPFRKNPEIYIYQYMDD
 LYVGSdleIGQHRAKVEELREHLLRWGFTTPDKKHQNEPPFLWWMGYELHPDKWTVQPIQL
 PEKDSWTVNDIQRLVGKLNWASQIYAGIKVKQLCKLLRGAKALTDIVPLTEEALELAEN
 REILKTPVHGVYDPSKDLIAEIQKQGQDQWSYQIYQEPFKNLKTGKYARTRGHTNDVR
 QLTEAVQKIAQECIVIWGKTPKFKLP IQKDTWETWWMDYWQATWIPKWEFVNT PPLVKLW
 YQLEKDPIAGVETFYVDGAANRETKIGKAGYVTDKGRQKVVTLTETTNQKTELHAIYLAL
 QDSGSEVNVTDSQYALGIIQAQPDRSESELVNQIIEKLIGKDKVYLSWVPAHKGIGENE
 QVDKLVSNIGIRKVLFLDGDIDKAQEDHEKYHSNWRAMANEFNLPPIVAKEIVANCDKCQLK
 GEAMHGQVDCSPGIWQIDCTHLEGKVIILVAVHVASGYLEAEVIPAETGQEAAYFILKLAG
 RWPVKTVHTDNGSNFTSNAVKAACWWANVRQEFGI PYN PQSQGVIESMNKELKKIIGQVR
 DQAEHLRTAVQMAVFIHNFKRKGGIGGYSAGERMIDIATDIQTTELQKQITKIQKFRVY
 YRDSRDP LWKGP AKLLWKGEAVVIQENS DIKVVPRRKAKI IKDYGKQVAGADCVAGRQD
 EDQ

53/109

Fig. 11

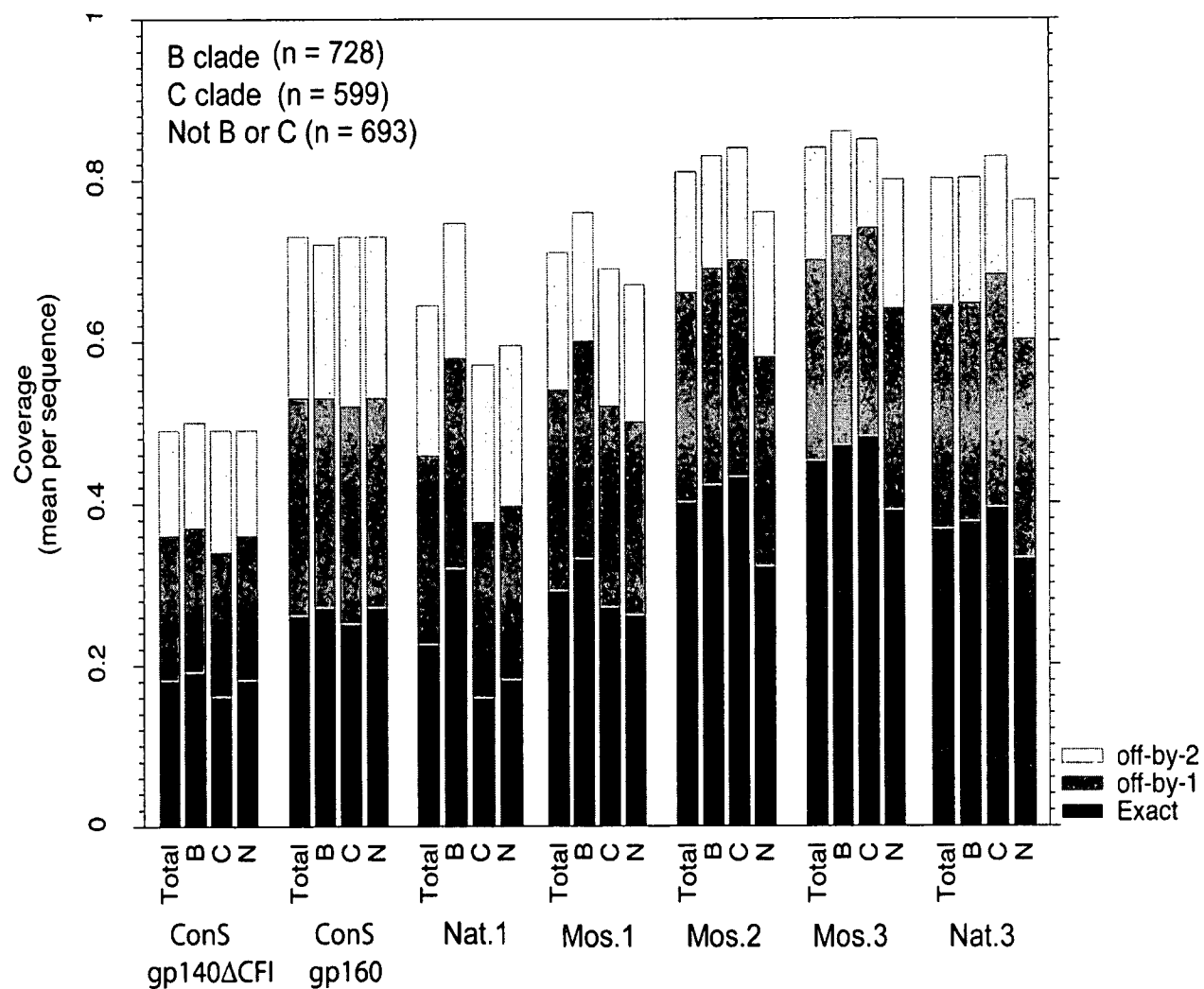
This plot is alignment independent, based on splintering all M group proteins, (database and CHAVI, one sequence per person) into all possible 9-mers, attending to their frequencies, and then looking for matches and near matches in each vaccine antigen or cocktail with the database



Vaccine	subset	Off-by0	Off-by1	Off-by2	(<3,>1)	unique	absent	rare < 3
ConSgp160	Total	0.2628	0.5301	0.7267	9	12	45	66
ConSgp160	B	0.2682	0.5344	0.7223	2	8	45	
ConSgp160	C	0.2526	0.5214	0.7302	1	0	45	
ConSgp160	N	0.2662	0.5332	0.7283	7	4	45	
Mos.3	Total	0.4485	0.7032	0.8358	15	164	8	179
Mos.3	B	0.4749	0.7319	0.8576	3	40	8	
Mos.3	C	0.4809	0.7363	0.8498	8	65	8	
Mos.3	N	0.3868	0.6383	0.7970	11	59	8	
Nat.1.acute	Total	0.2258	0.4598	0.6458	125	0	0	125
Nat.1.acute	B	0.3190	0.5803	0.7482	125	0	0	
Nat.1.acute	C	0.1589	0.3781	0.5726	0	0	0	
Nat.1.acute	N	0.1815	0.3979	0.5968	0	0	0	
Nat.3.acute	Total	0.3673	0.6449	0.8036	164	252	0	416
Nat.3.acute	B	0.3765	0.6483	0.8045	130	0	0	
Nat.3.acute	C	0.3940	0.6840	0.8307	19	102	0	
Nat.3.acute	N	0.3311	0.6036	0.7766	21	150	0	

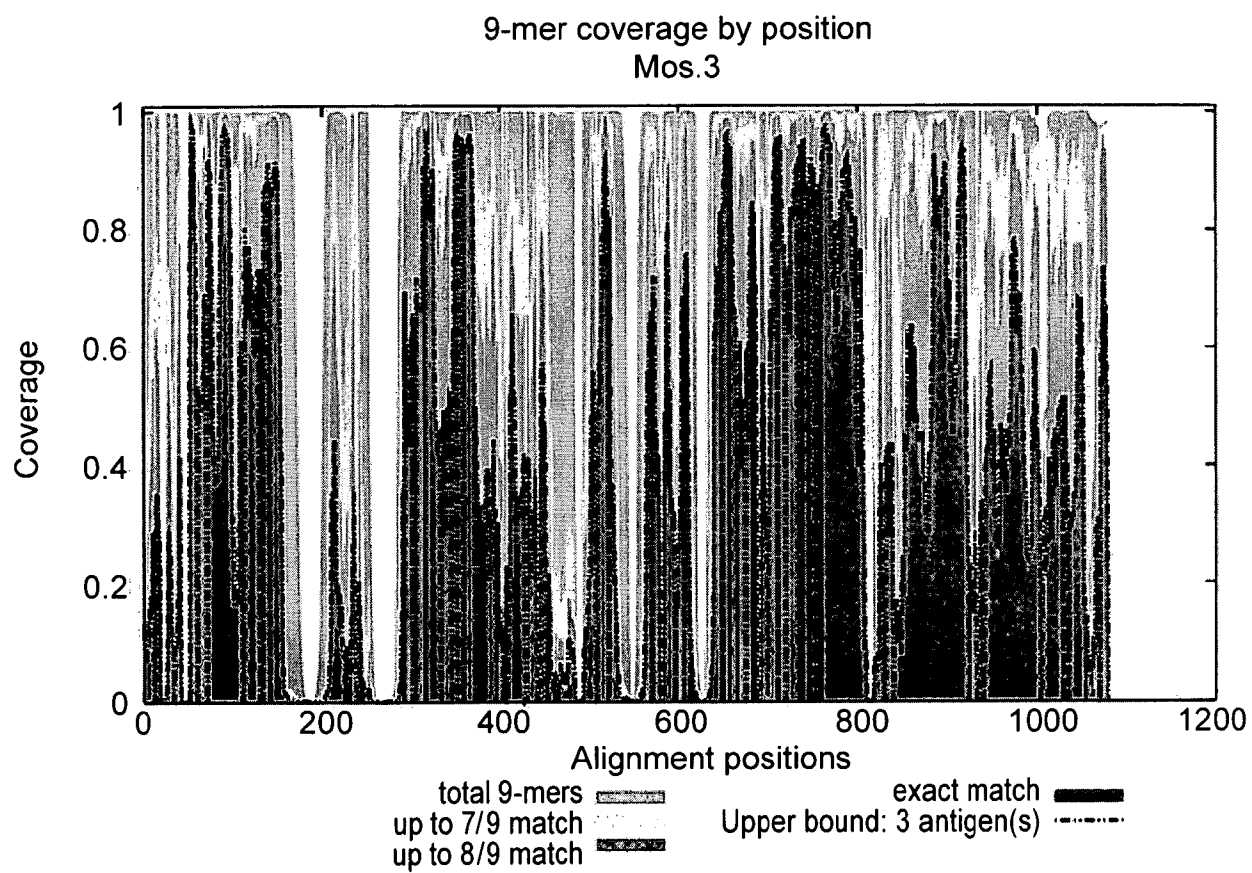
54/109

Fig. 12



55/109

Fig. 13



56/109

Fig. 14A

ConSgp160

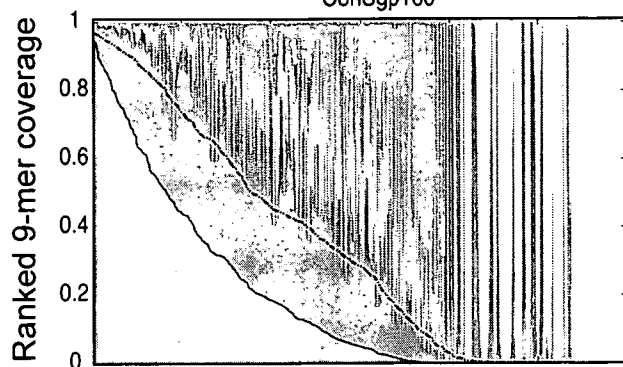


Fig. 14C

Mos.3

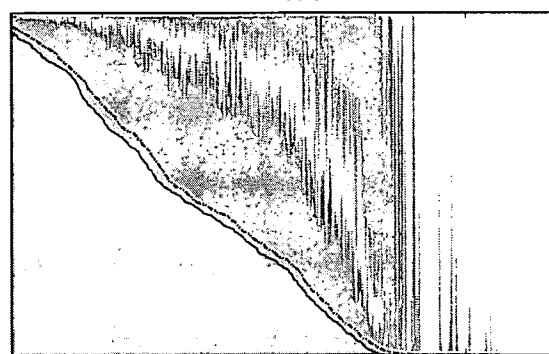


Fig. 14B

Nat.1.acute

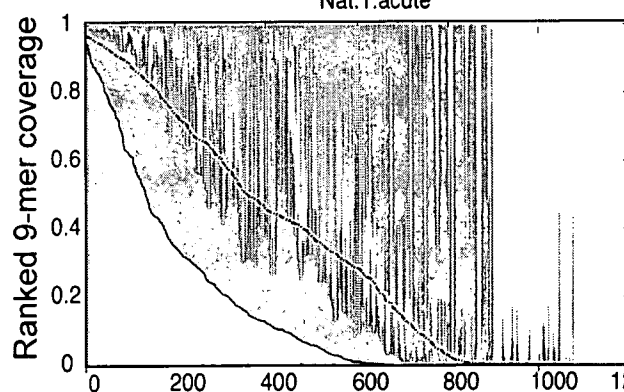
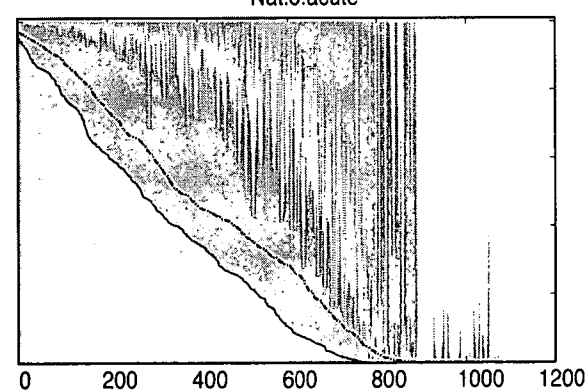
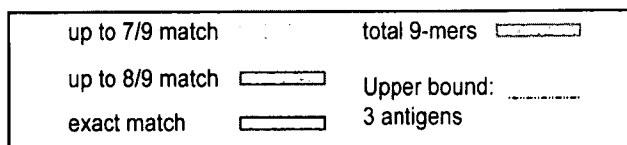


Fig. 14D

Nat.3.acute



Alignment positions (reordered by coverage ranking)



57/109

Fig. 15A

Con S

Sequences

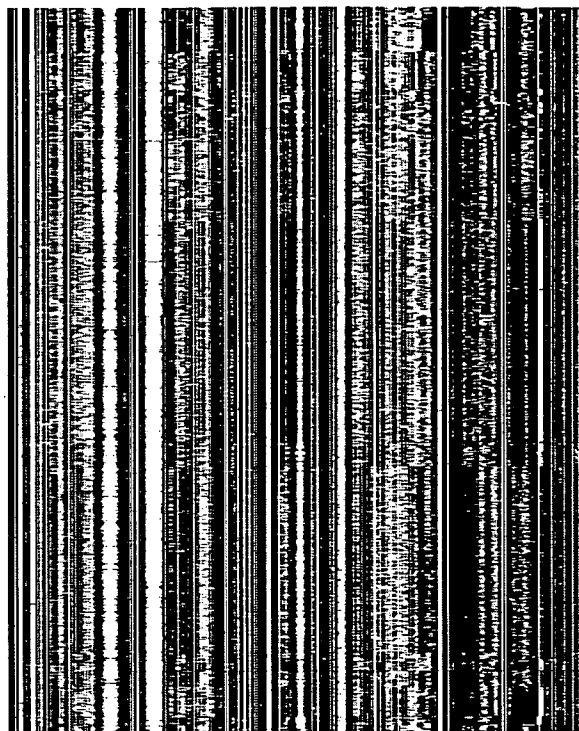


Fig. 15C

Mos.3

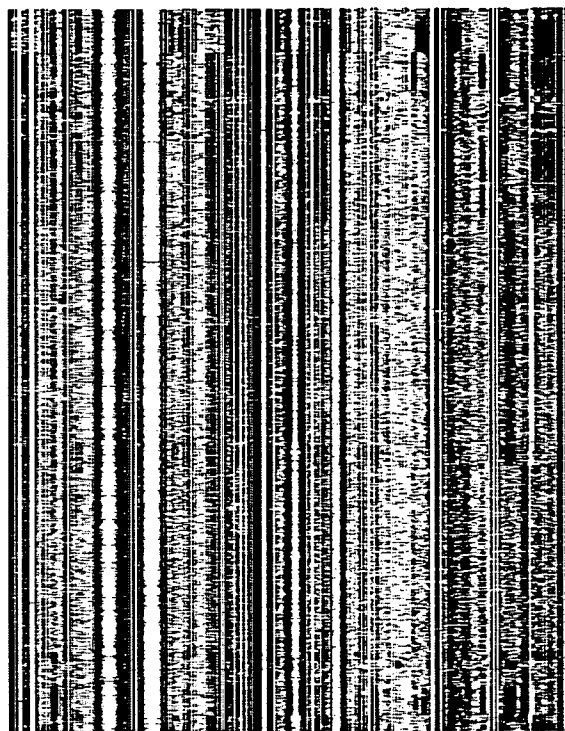


Fig. 15B

Nat.1

Sequences

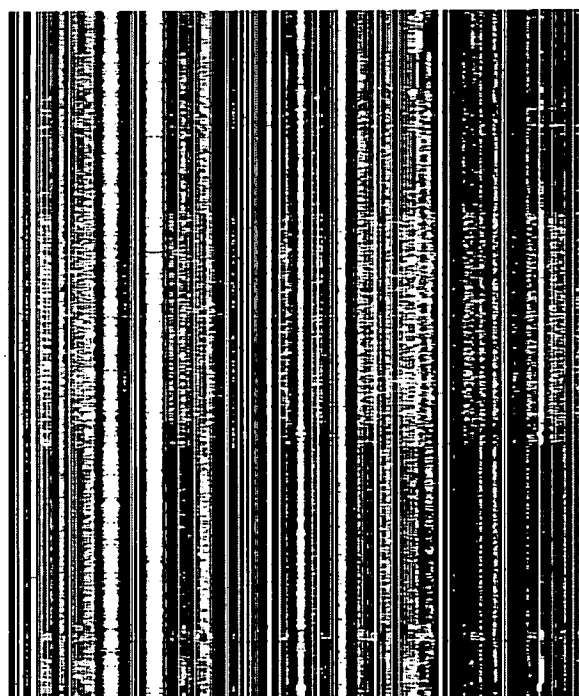
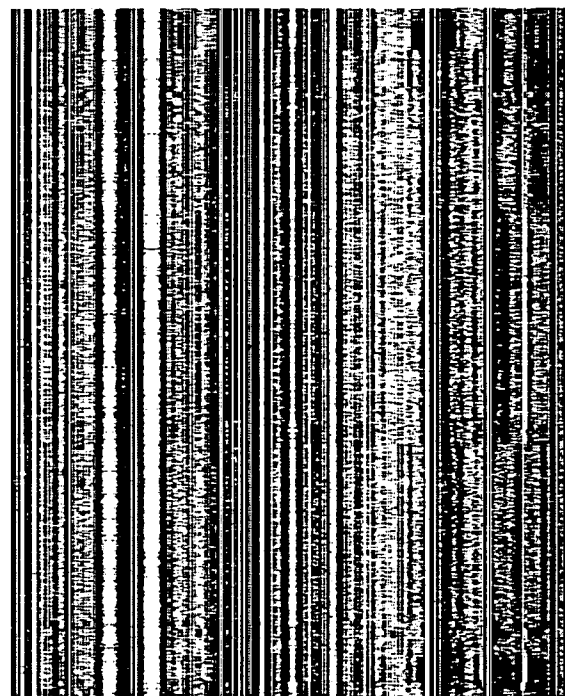


Fig. 15D

Nat.3



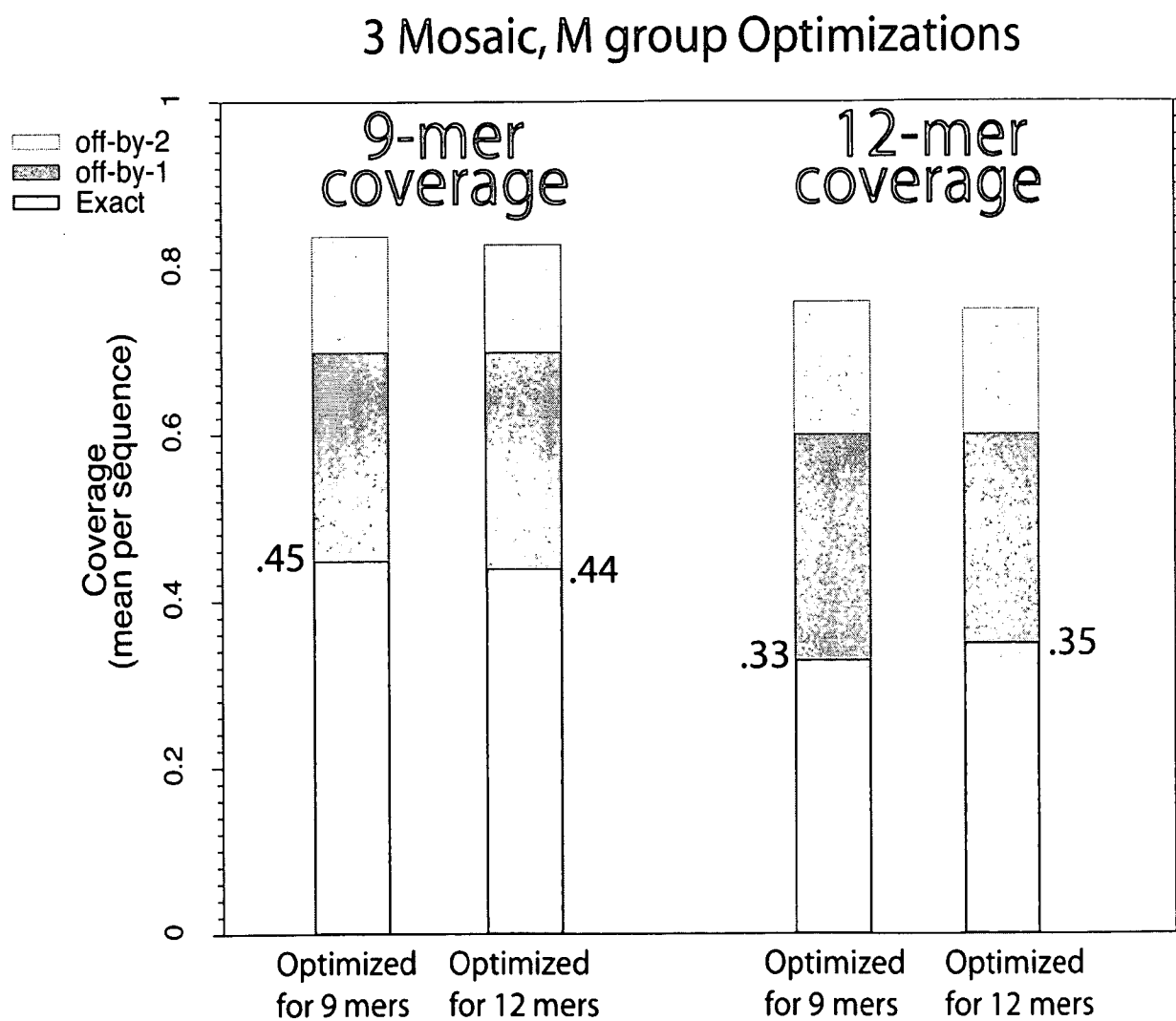
Env alignment position

■ B clade

0 1 2 3 4 5 6 7 8 9

58/109

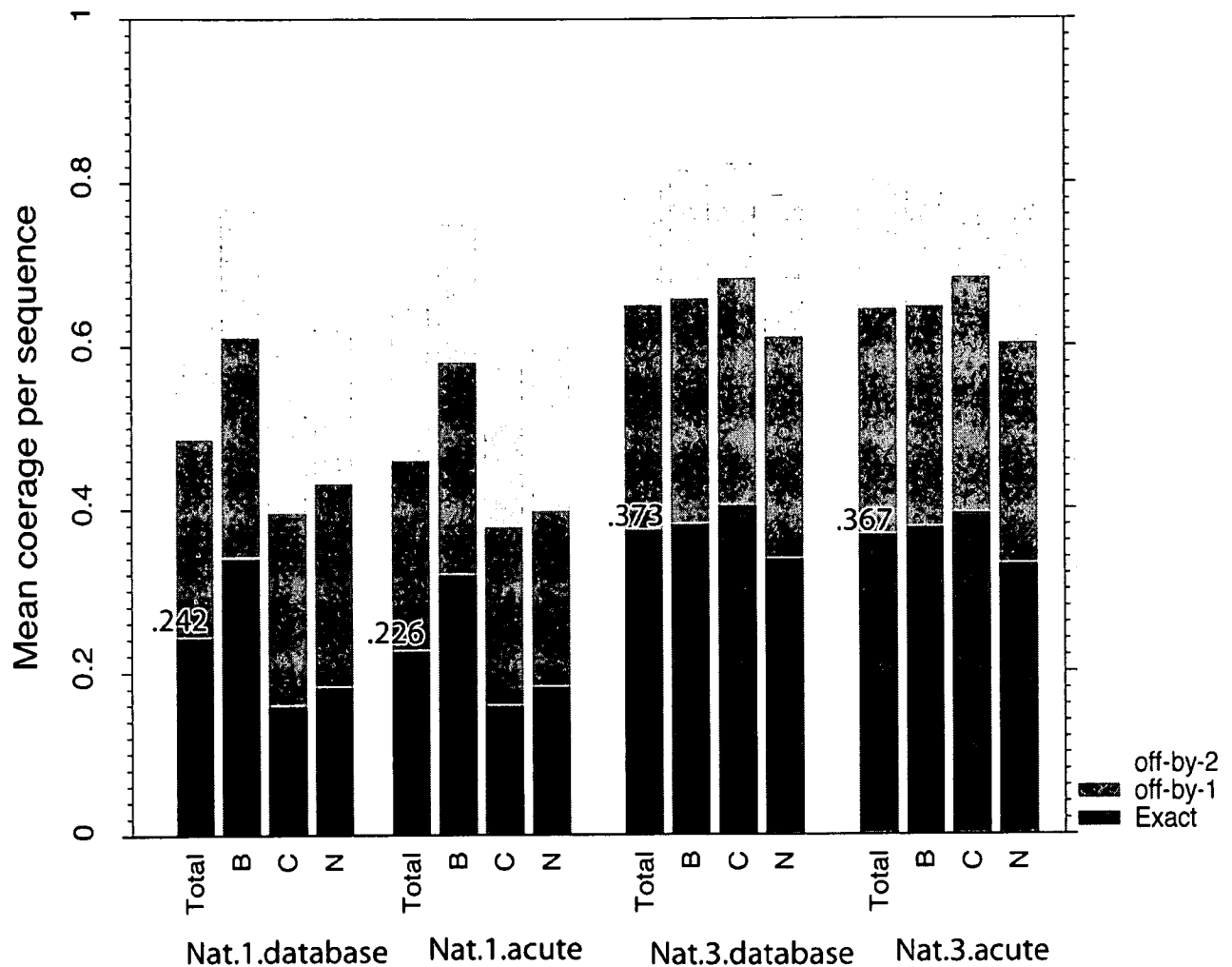
Fig. 16



59/109

Fig. 17

Coverage of the HIV database plus CHAVI sequences (N = 2020)



Nat.3.database

Option 1:

B YU2* -- 1986, USA

C DU467, South Africa,

A ML170 -- 1986 Kenya

Optimal for the set
after requiring inclusion
of one each of subtype
A, B and C.

Nat.3.acute

Option 2:

B 1059*

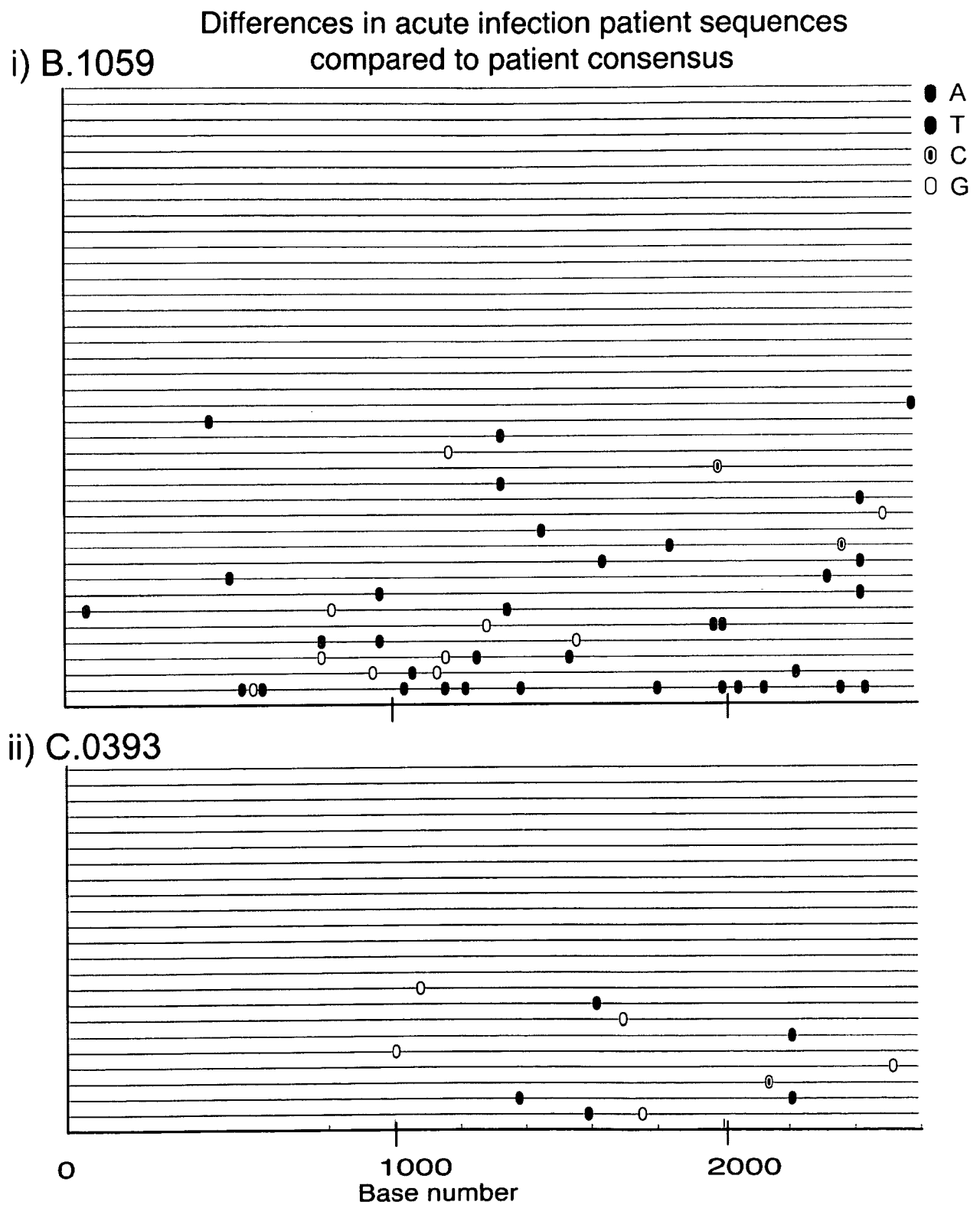
C 0393

A R66201FPB

Optimal for the set
after requiring inclusion
of one each of subtype
A, B and C as well as
restricting antigen selection
to SGA sequences sampled
during acute infection.

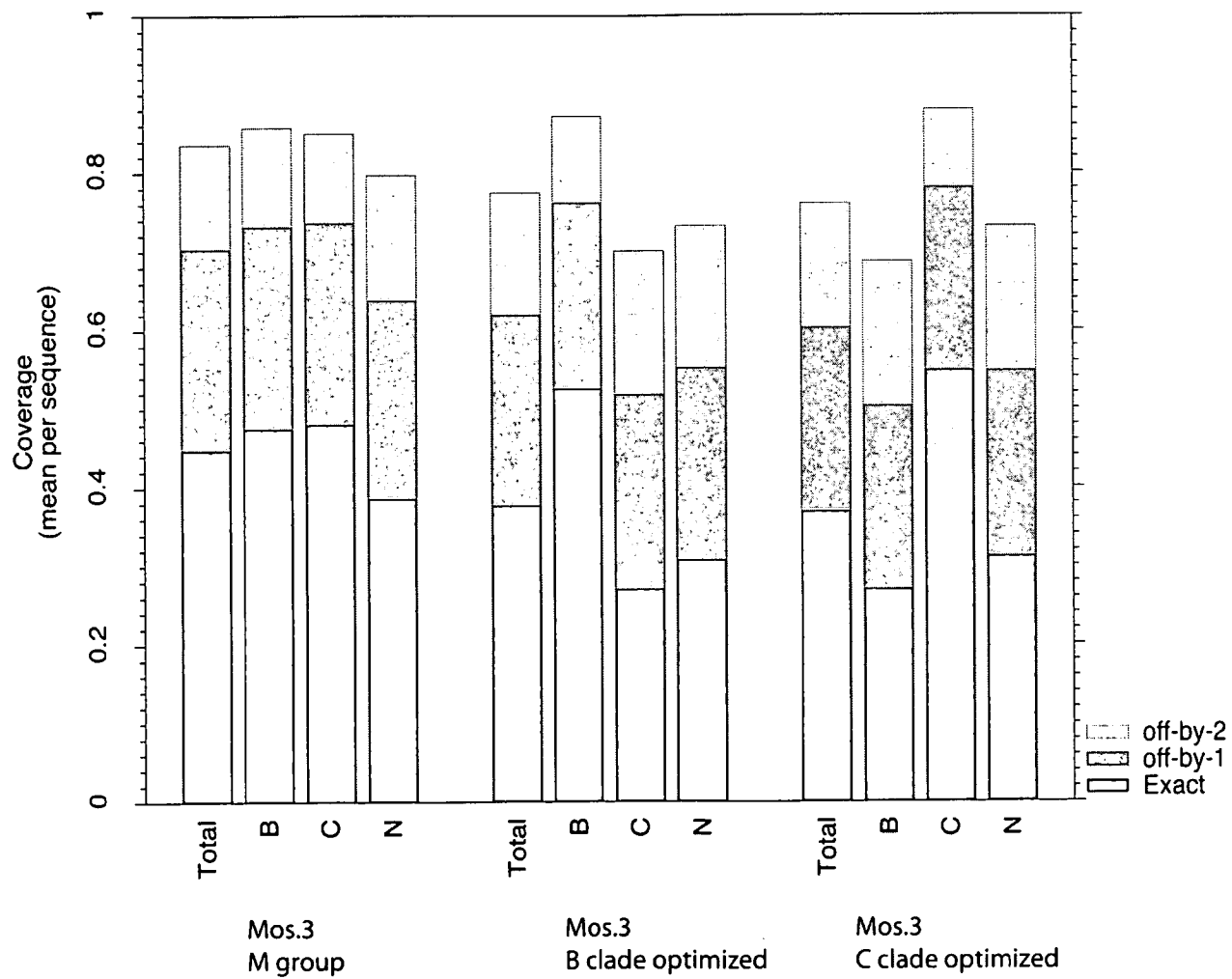
60/109

Fig. 18



61/109

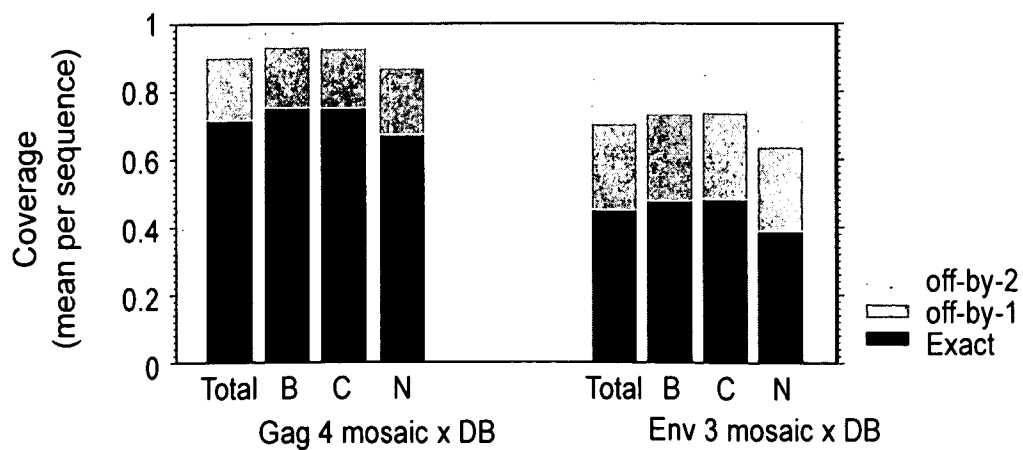
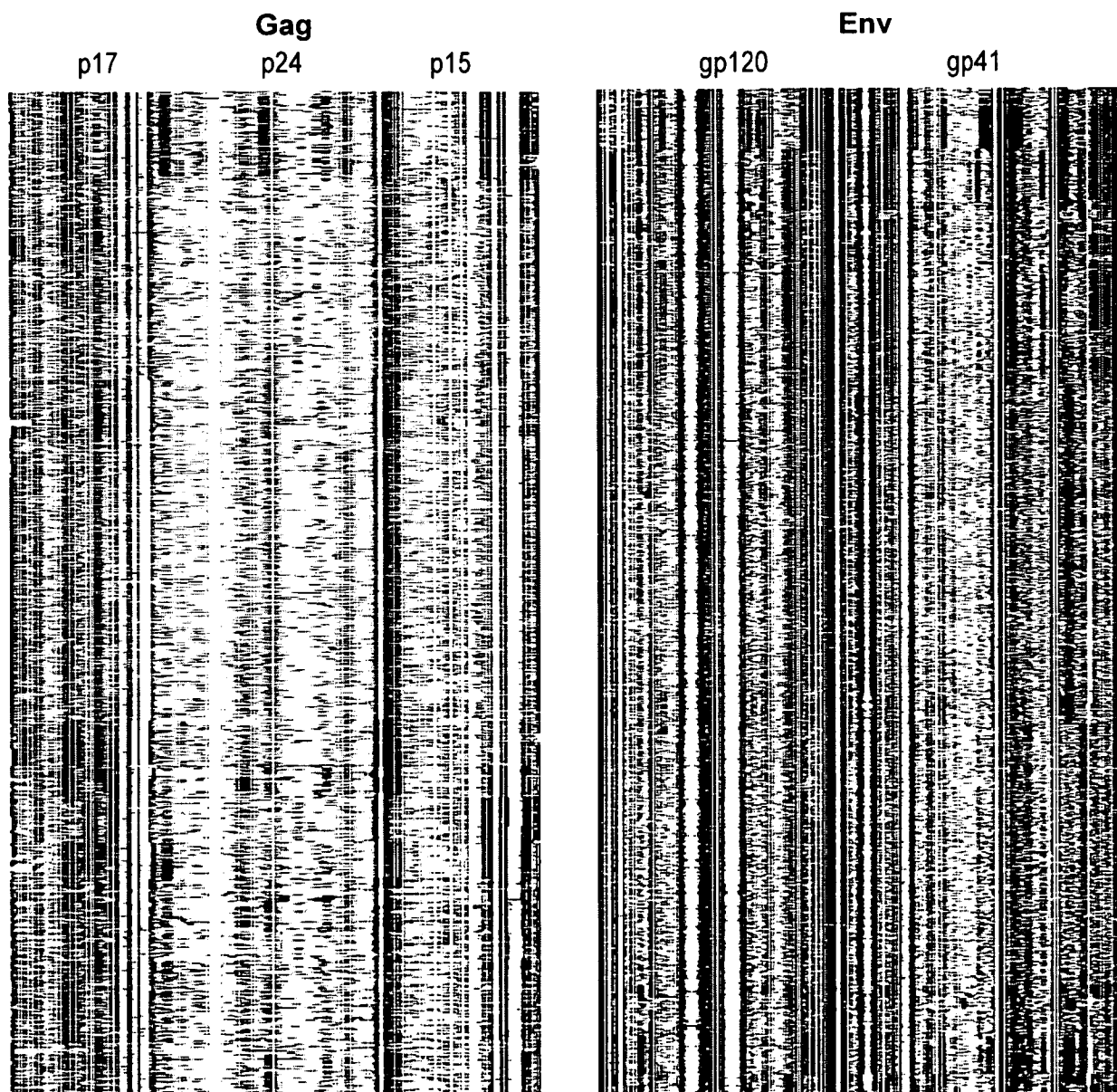
Fig. 19



62/109

Fig. 20

Proposed vaccine mosaic coverage of Gag and Env



63/109

Fig. 21

>nefM_4.1Dmyr

MAAKWSKSSIVGWPAVRERMRAEPAADGVGA VSRDLEKHGAITSSNTAATNADCAWLEAQEE
DSEVGFPVRPHVPLRPMTYKAAVDLSHFLKEQGGLEGLIYSQKRQDILDLWVYHTQGYFPDWQN
YTPGPGIRYPLTFGWCFKLVPVDPREVVEANKGENNCLLHPMSQHGMDDPKEVLMWKFDSRLA
FHHMAREKHPEFYKDC

>nefM_4.2Dmyr

MAAKWSKRSIVGWPAIRERMRRTEPA AEGVGAASQDLDKYGALTSSNTAQTNPDCAWLEAQEE
EEVGFPVRPQVPVRPMTYKGALDLSHFLKEKGGLDGLIYSRKRQEILDLWVYNTQGYFPDWQNY
TPGPGVRYPLTFGWICYKLVPVDPEEVEKANEGENNSLLHPMSLHGMEDPEREVLKWKFDSRLAL
KHRARELHPEFYKDC

>nefM_4.3Dmyr

MAANWSKSSIVGWPEIRERJRRTPAAEGVGAASRDLERHGAITSSNTATTNAACAWLEAQEED
VGFPVRPQVPLRPMTFKGAFDLGFFLKEKGGLEGLIYSKKRQDILDLWVYNTQGFFPDWQNYTPG
PGTRFPLTFGWCFELVPVDPKEVEEANEKENNCLLHPICQHGMDDEEREVLWKFDSRLARRHLA
REKHPEYYKDC

>nefM_4.4Dmyr

MASKWSKSSIVGWQPVRERJRQTPAAEGVGAVSQDLDKHGAVTSSNTAANNADCAWLQAQEE
EEEVGFPVKPQVPLRPMTYKGAFDLSFFLKEKGGLDGLIYSKKRQEILDLWVYHTQGFFPDWHNY
TPGPGTRYPLCFGWCFKLVPMDPAEVEEATEGENNSLLHPICQHGMEDEDREVLVWRFDSRLARR
HIARELHPEYYKDC

>Gag-M4.1Dmyr

MAARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQILGQLQ
PSLQTGSEELRSLYNTVAVLYCVHQRIDVKDTKEALEKIEEEQNKSQQKTQQAADGKVSQNY
IVQNAQGQMVHQAI SPRTLNAWVKVIEEKGSPEVIPMFSALAEGATPQDLNMLNTIGGHQAAM
QMLKDTINDEAAEWDRLHPVQAGPVAPGQIREPRGSDIAGTTSNLQEIQGWMTSNPPIPVGDYIKR
WIIMGLNKIVRMYSPTSILDIRQGPKEFRDYVDRFFRTLRAEQASQEVKNWMTETLLVQNSNPDC
KTILKALGPAATLEEMMTACQGVGGPGHKARVLAEMSQVQQPNIMMQRGNFKGQKRIKCFNC
GREGHLARNCRAPRKKG CWKCGREGHQMKDCTESKANFLGKIWP SNKGRPGNFLQSRPEPSAPP
AESFGFGEEITPSQKQEQDKELYPLASLSLFGNDPLSQ

>Gag_M4.2 Dmyr

MAARASVLSGGKLD AWEKIRLRPGGKKKYRLKHLVWASRELD RFALNPSLLETAEGCKQIMKQL
QPALKTGTEELKSLYNTVATLYCVHEKIDVRDTKEALDKIEEEQNKIQQKTQQAKEADGKVSQNY
PIVQNIQGQMVHQPI SPRTLNAWVKVVEEKAFSPEVIPMFTALSDGATPQDLNSMLNAVGGHQA
MQILKDTINEEAADWDRLHPVHAGPVAPGQMREPRGSDIAGTTSTLQEIQGWMTNPPIPVGEIYK
RWIILGLNKIVRMYS PVSILDIKQGPKEFRDYVDRFFKVLRAEQATQDVKNWMTDTLLIQNANPD
CKSILRALGPGATLEEMMTACQGVGGPSHKARILAEAMSQVTNSATIMMQRGNFRNQRKTVKCF
NCGKEGHLARNCKAPRKRG CWKCGKEGHQMKECTERQANFLGKIWPSSKGRPGNFPQSRPEPTA
PPEESFRFGEETTPSQKQEPIDKELYPLASLSLFGNDPSSQAS

64/109

Fig. 21 cont'd-1

>Gag_M4.3 Dmyr

MAARASILRGKGLDKWEKIRLRPGGKKRYMLKHLI WASRELERFALNPGLLETAEGCQQIIEQLQS
 TLKTGSEELKSLFNTVATLYCVHQRIEVKDTKEALDKVEEEQNKSKKKAQQAAADTGNSSQVSQ
 NYPIVQNLQGQMVHQALSPRTLNAWVKVIEEKAFSPEIIPMFTALSEGATPSDLNMTMLNTVGGHQ
 AAMQMLKDTINEEAAEWDRVHPVHAGPIPPGQMREPRGSDIAGTTSSLQEQA WMTSNPPVPVGE
 IYKRWIVLGLNKJVRMYSPTSILDIRQGPKPEFRDYVDRFFKTLRAEQASQDVKNWMTETLLVQN
 ANPDCKTILRALGPGASLEEMMTACQGVGGPSHKARVLAEAMSQTNSAILMQRSNFKGSKRIVKC
 FNCGKEGHIARNCRAPRKRGWCWCGQEGHQMKDCNERQANFLGKJWPSHKGRPGNFLQNRPEPT
 APPEPTAPPAESFRFEETTPAPKQELKDREPLTSLKSLFGSDPLSQAS

>Gag_M4.4 Dmyr

MAARASVLRGEKLDKWERJRLRPGGKKHYMLKHLV WASRELEKFALNPGLLETSEGCKQIIKQL
 QPALQTGTEELRSLFNTVATLYCVHAGIEVRDTKEALDKIEEIQNKSKQKTQQAAAGTGSSSKVSQ
 NYPIVQNLQGQMVHQPLSPRTLNAWVKVVEEKGFNPEVPMFSALSEGATPQDLNMMLNTVGGH
 QAAMQMLKETINEEAAEWDRLHPVHAGPIAPGQMREPRGSDIAGSTSTLQEQA WMTGNPPVPV
 GDIYKRWIILGLNKJVKMYSPTSILDIKQGPKEFRDYVDRFYKTLRAEQATQEVKNWMTDITLLVQ
 NANPDCKSILKALGTGATLEEMMSACQGVGGPAHKARVLAEAMSQANNTNIMMQRSNFKGPKRI
 IKCFNCGKEGHIAKNCRAPRKKGWCWCGKEGHQMKDCTERQANFLGRIWPSSKGRPGNFLQSRP
 EPTAPPAEPTAPPAESFKFEETTPAPKQEPKDREPLTSLRSLFGSDPLLQAS

>M_mos_3_1 (M_mos_Env_3_1)

MRVKGIRKNYQHLWRWGTMLLGLMLICSAAEQLWVTVYYGVPVWRDAETTLFCASDAKAYER
 EVHNVWATHACVPTDPNPQEI VLENVTEEFNMWKNNMVDQM HEDIISLWDESLKPCVKLTPLCV
 TLNCTDVNVTKTNSTSWGMMEKGEIKNCSFNMTTEL RDKKQKVYALFYKLDIVPLEENDTISNST
 YRLINCNTSAITQACPKVTFEPIPIHYCTPAGFAILKCN DKKFNGTGPCKNVSTVQCTHGIRPVVTTQ
 LLLNGSLAEEEIIRSENLTNNAKTIIVQLNESVVINCTRPNNNTRKSIRIGPGQTFYATGDIIGNIRQA
 HCNISREKWINTRDVRKKLQEHFNKTIIFNSSSGGDLEITHSFNCRGEFFYCNTSKLFNSVWGNS
 SNVTKVNGTKVKETITLPCJKQIINMWQEVGRAMYAPPIAGNITCKSNITGLLLVRDGGNVTNNT
 EIFRPGGGNMKDNWRSELYKYKVVEIKPLGIAPTKAKRRVVEREKRAVGLGAVFLGFLGAAGST
 MGAASMTLTVQARQLLSGIVQQQSNLLRAIEAQQHMLQLTVWG IQLQARILAVERYLRDQQLL
 GIWGC SGKLICTTNVPWNSSWSNKS LDEIWNMTWMQWEKEIDNYTSLIYTLIEESQNQQEKNEQ
 DLLALDKWANLWNWFDISNWLWYIRFIMIVGGLIGLRJVFAVLSIVNRVRKGYSPLSFQTLTPNPR
 GPDRLGRIEEEGGEQDKDRSIRLVNGFLALAWDDLRLNCLFSYHRLRDL LLIVTRIVELLGRGWE
 ALKYLWNLLQYWIQELKNSAVSLLNATAJAVAEGTDRVIEVVQRACRAJLHIPRRJRQGLERALL

>M_mos_3_2 (M_mos_Env_3_2)

MRVKETQMNWPNLWKWGTLLGLVIICSASDNLWVTVYYGVPVWKEATTTLFCASDAKAYDTE
 VHNWVATYACVPTDPNPQEVVLGNVTENFNMWKNNMVEQM HEDIISLWDQSLKPCVRLTPLCV
 TLNCSNANTTNTNSTEEIKNCSFNITTSIRDKVQKEYALFYKLDVVPIDNDNTSYRLISCNTSVITQA
 CPKVSFEPIPIHYCAPAGFAILKCKDKKFNGTGPCNTVSTVQCTHGIRPVVSTQ LLLNGSLAEEEVVI
 RSENFNTNAKTIIVHLNKSVEINCTRPNNNTRKSIHIGPGRAF YATGEIIGDIRQAHCNISRAKWNNT
 LKQIVKLLKEQFNKTIIFNQSSGGDPEITHSFNCGGEFFYCNTSGLFNSTWNSTATQESNNTLNG
 NITLPCR KQIVNMWQEVGKAMYAPPIRGQIRCSSNITGLILTRDGGNNNSTNETFRPGGGDMRDN
 WRSELYKYKVVKIEPLGVAPTAKARRVQREKRAVGTIGAMFLGFLGAAGSTMGAASLT LTVQA
 RLLLSGIVQQQNNLLRAIEAQQHLLQLTVWG IQLQARVLAVERYLKDQQLLGIWGC SGKLICTT
 TVPWNTSWSNKSLNEIWDNMTWMEWEREIDNYTGLIYTLIEESQNQQEKNEQELLELDKWASL
 WNWFDITKWLWYIKIFIMIVGGLVGLRJVFTVLSIVNRVRQGYSP LSFQTHLPAPRGPDREPIEEE
 GGERDRDRSGRLVDGFLAIWVDLRSLCLFSYHQLRDFILIAARTVELLGHSSLKGLRRGWEALKY
 WWNLLQYWSQELKNSAISLLNTTAJVVAEGTDRIEVLQRAGRAJLHIPTRJRQGLERLLL

65/109

Fig. 21 cont'd-2

>M_mos_3_3 (M_mos_Env_3_3)

MRVRGIQRNWPQWWIWGILGFWMLMICNVVGNLWVTVYYGVPVWKEAKTTLFCASDAKAYEK
EVHNVWATHACVPTDPSPQEVVLENTENFNMWKNDMVDQMHEDEVISLWDQSLKPCVKLTHL
CVTLNCTNATNTNYNNSTNVTSSMIGEMKNCSFNITTEIRDKSRKEYALFYRLDIVPLNEQNSSEY
RLINCNTSTITQACPKVSFDPIPIHYCAPAGYAILKCNNKTFNGTGPCNNVSTVQCTHGIKPVVSTQL
LLNGSLAEGEIIIRSENLTDNAKTIIVHLNESVEIVCTRPNNNTRKSVRIGPGQAFYATGDIIGDIRQA
HCNLSRTQWNNTLKQIVTKLREQFGNKTIVFNQSSGGDPEIVMHSFNCGGEFFYCNTTQLFNSTWE
NSNITQPLTLNRTKGPNDTITLPCRKQIINMWQGVGRAMYAPPIEGLIKCSSNITGLLLTRDGGNNS
ETKTTETFRPGGGNMRDNWRNELYKYKVQIEPLGVAPTRAKRRVVEREKRAVGIGAVFLGFLG
TAGSTMGAASITLTVQARQVLSGIVQQQSNLLKAIEAQQHLLKLTWVGKQLQTRVLAIERYLKD
QQLLGLWGCSGKLICTTAVPWNSSWSNKSQTDIWDNMTWMQWDREISNYTDTIYRLLEDSONQ
QEKNEKDLLALDSWKNLWNWFDITNWLWYKIFIIVGGLIGLRJIFAVLSIVNRCRQGYSPSLQTL
IPNPRGPDRLGGEIEEGGEQDRDRSIRLVSGFLALAWDDLRLCLFSYHRLRDFILIVARAVELLGRS
SLRGLQRGWEALKYLGSLVQYWGLELKKSAISLLDTIAJAVAEAGTDRIEVIQRICRAIRNIPRRIRQ
GFEAALL

66/109

Fig. 22

HVI number	Gene name	Nef	Myristylation signal mutated
HV13236	M.con_Nef01_Dmyr.WLV	Group M (2001) consensus	Yes
HV13319	nefM_4.1Dmyr.wlv	Mosaic No. 1	Yes
HV13231	nefM_4.2Dmyr.wlv	Mosaic No. 2	Yes
HV13230	nefM_4.3Dmyr.wlv	Mosaic No. 3	Yes
HV13232	nefM_4.4Dmyr.wlv	Mosaic No. 4	Yes
HV10001	No inser		

All five constructs were cloned into HV10001 (WLV001AM DNA vaccine plasmid).

M.con_Nef01_Dmyr.WLV (657nt.) (hv13236), 67% GC

XhoI

CTCGAGAAGAAAATGCGCGCCAAGTGGTCTGAAGAGCTCCATCGTTCGGGTGGCCGGCGGTCCGGGAG
CGCATCCGGCGGACGCACCCCGCGGCCGAGGGGTCGGCGCGGTCTCGCAAGACCTCGACAAGC
ACGGGGCGATCACGTCGAGCAACACCGCCGCGAACAACCCGACTGCGCGTGGCTGGAGGCCCA
GGAGGAAGAGGAAGAGGTCTGGCTTCCCGGTCCGCCCCGCAAGTGCCGCTCAGGCCGATGACGTAC
AAGGCGGCCCTCGACCTCTCGCACTTCTTGAAGAGAAGGGTGGCCTGGAGGGGCTCATCTACT
CGAAGAAGAGGCAGGAGATCCTCGACCTGTGGGTCTACCACACCCAGGGCTACTTCCCGGACTG
GCAGAACTACACGCCAGGCCCGGGAATCCGCTACCCGCTGACCTTCCGGTGGTGCTTCAAGCTC
GTCCCCGTGACCCGGAGGAGGTCTGAGGAAGCCAACGAGGGCGAGAACACAGCCTCCTGCACC
CGATGTGCCAGCACGGGATGGAGGACGAGGAGCGCGAGGTGCTGATGTGGAAGTTCTGACTCGCG
CCTGGCCCTGCGCCACATCGCCCGGGAGCTCCACCCGGAGTACTACAAGGACTGCTGATAAGCTA
GCGGATCCTGATCA

NheI BamHI BclI

MAAKWSKSSIVGWPAVRERIRRTHPAAEGVGAVSQDLKDHGAITSSNTAANNPDCAWLEAQEEE
EEVGFPVRPQVPLRPMTYKAALDLSHFLKEKGGLEGLIYSKKRQEILDLWVYHTQGYFPDWQNY
TPGPGIRYPLTFGWCFKLVPVDPEEVEEANEGENNSLLHPMCQHGMEDEREVLWVKFDSRLAL
RHIARELHPEYYKDC_

>HV13236 in hv10001, 3953nt.

AAATGGGGGCGCTGAGGTCTGCCTCGTGAAGAAGGTGTTGCTGACTCATAACCAGGCCTGAATCGCCCCATCATCCAG
CCAGAAAGTGAGGGAGCCACGGTTGATGAGAGCTTTGTTGTAGGTGGACCAGTTGGTGATTTTGAACCTTTGCTTTG
CCACGGAACGGTCTGCGTTGTGCGGAAGATGCGTGATCTGATCCTTCAACTCAGCAAAAGTTTCGATTTATTCAACAA
AGCCGCCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAACCAATTAACCAATTCTGCGTTCAAAATGGTA
TGCGTTTGTACACATCCACTATATATCCGTGTGCTTCTGTCCACTCCTGAATCCCATTCAGAAATTCTCTAGCGAT
TCCAGAAGTTTCTCAGAGTCGGAAGTTGACCAGACATTACGAAGTGGCACAGATGGTCATAACCTGAAGGAAGATC
TGATTGCTTAAGTCTTACGTTAAGACCGACGCGCTCGTCTGATAACAGATGCGATGATGACAGCAATCAACATGG
CACCTGCCATTGCTACCTGTACAGTCAAGGATGGTAGAAATGTTGTGCGTCCCTTGACACGAATATTACGCCATTTG
CCTGCATATTCAACAGCTCTTCTACGATAAGGGCACAAATCGCATCGTGGAACGTTTGGGCTTCTACCGATTTAGC
AGTTTGATACACTTTCTCTAAGTATCCACCTGAATCATAAATCGGCAAAATAGAGAAAAATTGACCATGTGTAAGCG
GCCAATCTGATTCACCTGAGATGCATAATCTAGTAGAATCTCTTCGCTATCAAAATTCACCTCCACCTTCCACTCA
CCGGTTGTCCATTTCATGGCTGAACCTCTGCTTCTCTGTGACATGACACACATCATCTCAATATCCGAATACGGACC
ATCAGTCTGACGACCAAGAGAGCCATAAACACCAATAGCCTTAACATCATCCCCATATTTATCCAATATTCGTTCTCT
TAATTTTCATGAACAATCTTCAATTCTTTCTTCTCTAGTCATTATTATTGGTCCGTTTCATAACACCCCTTGTATTACTG
TTTATGTAAGCAGACAGTTTTTATTGTTTCATGATGATATATTTTTATCTTGTGCAATGTAACATCAGAGATTTTGAGA
CACAACGTGGCTTTCCCGGCCCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTA
GAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAACCCGCT

Mosaic and Group M nef_Dmyr-Patent.doc

67/109

Fig. 22 cont'd-1

ACCAGCGGTGGTTTGGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCAGA
TACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACCTTCAAGAAGTCTGTAGCACCAGCTACATACCTC
GCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTTCGTGTCTTACCAGGTTGGACTCAAGACGATA
GTTACCAGGATAAAGGCGCAGCGGTGCGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACA
CCGAACCTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCG
GTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCTGCTATCTTTATAGTCTGT
CGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCA
GCAACGCGGCCCTTTTTACGGTTTCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTCTGCTTATCCCCTGAT
TCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGACGCCGAACGACCGAGCGCAGCTC
AGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTCTCCTTACGCATCTGTGCGGTATTTACACACGCATAT
GGTGCACCTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTC
GCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAG
GGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATAGCCGCGGCATCGATGATATCCATTGCATACGTT
GTATCTATATCATAATATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGTTGACATTGATTATTGACTA
GTTATTAATAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCATATATGGAGTTCGCGGTTACATAACTTACGGTA
AATGGCCCGCTGGCTGACCGCCCAACGACCCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCC
AATAGGGAAGTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAACTGCCCACTTGGCAGTACATCAAGTGTATC
ATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTACATGACCTTA
CGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACAT
CAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCATTGACGTCAATGGGAGTTTGTGTTT
GGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCCATTGACGCAATGGGCGGTAGGCGTGTA
CGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGA
CCTCCATAGAAGACACCGGGACCGATCCAGCCTCCGCGGGCGCGCTCGACAGGGCGACGCGAAACTTGGGCCCACTC
GAGAAGAAAATGGCGGCCAAGTGGTTCGAAGAGCTCCATCGTCGGGTGGCCGGCGGTCCGGGAGCGCATCCGGCGGAC
GCACCCCGCGGGCGAGGGGGTTCGGCGCGGTCTCGCAAGACCTCGACAAGCACGGGGCGATCACGTCGAGCAACACCG
CCGCGAACAACCCCGACTGCGCGTGGCTGGAGGCCAGGAGGAAGAGGTGCGGCTTCCCGGTCCGCCCCGCAA
GTGCCGCTCAGGCCGATGACGTACAAGGCGGCCCTCGACCTCTCGCACTTCTGAAAGAGAAGGGTGGCCTGGAGGG
GCTCATCTACTCGAAGAAGAGGCGAGGATCCTCGACCTGTGGGTCTACCACACCCAGGGCTACTTCCCGGACTGGC
AGAAGTACACGCCAGGCCCGGGAATCCGCTACCCGCTGACCTTCGGGTGGTGCTTCAAGCTCGTCCCGCTCGACCCG
GAGGAGGTCGAGGAAGCCAACGAGGGCGAGAACACAGCCTCCTGCACCCGATGTGCCAGCACGGGATGGAGGACGA
GGAGCGCGAGGTGCTGATGTGGAAGTTCGACTCGCGCCTGGCCCTGCGCCACATCGCCCGGGAGCTCCACCCGGAGT
ACTACAAGGACTGCTGATAAGCTAGCGGATCCTGATCAGCCTCGACTGTGCCCTTCTAGTTGCCAGCCATCTGTTGTT
TGCCCCCTCCCCCGTGCTTCTTTGACCCTGGAAGGTGCCACTCCCACTGTCTCTTCTTAATAAAATGAGGAAATTGC
ATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAG
ACAATAGCAGGCATGCTGGGGAATTT

Thursday, August 2, 2007

Sequence 0 Length : 3953

AatII GACGTC

4 Sites

2510 2563 2646

2832

AccI GTMKAC

3 Sites

3050 3433 3535

AflIII ACRYGT

1 Site

1897

AluI AGCT

13 Sites

109 633 1340

1597 1643 1733

1959

2184 2951 3112

3524 3683 3717

AlwNI CAGNNCTG

2 Sites

1488 2129

AosII GRCGYC

5 Sites

2507 2560 2643

2829 2983

ApaI GGGCCC

1 Site

3075

ApaLI GTGCAC

2 Sites

1583 2081

AvaI CYCGRG

3 Sites

3078 3482 3676

BamHI GGATCC

1 Site

3723

BamI GGYRCC

3 Sites

538 2850 3807

BamII GRGCTC

5 Sites

2953 3075 3114

3391 3685

BclI TGATCA

1 Site

3729

BclI CCSGG

12 Sites

1173 1521 3021

3139 3298 3457

3484

3485 3541 3678

3679 3691

BglI

GCCNNNNNGGC

4 Sites

2475 2597 2668

3320

BglII

AGATCT

1 Site

458

Bsp1286

GDGCHC

8 Sites

652 1587 2085

2953 3075 3114

3391

3685

BspHI

TCATGA

2 Sites

1007 1105

BspNI

CCWGG

13 Sites

52 1738 1751

1872 2475 2668

2977

3270 3381 3444

3478 3657 3801

BseHII

GCGCGC

1 Site

3045

Fig. 22 cont'd-2

Mosaic and Group M: nef_Dmyr-Patent.doc

69/109

Fig. 22 cont'd-3

2282 3040 3166
SalI **GTCGAC**
 2 Sites
 3049 3534
Sau3A **GATC**
 13 Sites
 188 193 458
 1237 1245 1256
 1331
 2970 3026 3214
 3415 3723 3729
Sau96A **GGNCC**
 17 Sites
 123 587 920
 1049 1174 2266
 2468
 2661 3022 3071
 3072 3134 3266
 3299
 3341 3480 3659
ScaI **AGTACT**
 1 Site
 55
StyI **CCWWGG**
 1 Site
 2745
TaqI **TCGA**
 15 Sites
 216 1799 2287
 3050 3079 3105
 3199
 3222 3346 3399
 3421 3535 3550
 3646
 3738
XhoI **CTCGAG**
 1 Site
 3078
XhoII **RGATCY**
 5 Sites
 458 1245 1256
 3415 3723
XmaI **CCCGGG**
 2 Sites
 3482 3676
XmaIII **CGGCCG**
 1 Site
 3166
XmnI **GAANNNTTC**
 1 Site
 811
 Following enzymes have no
 sites
AccIII **AflII** **Asp718**
AsuII **AvrII** **BalI**
BbeI **BspMII** **BstEII**
BstXI **DraI**
Eco47III
EcoO109 **EcoRI** **EspI**
FspI **HindIII** **HpaI**
KpnI **MluI** **MstI**
NarI **NotI** **OxaNI**
PflMI **PpuMI** **PssI**
PstI **PvuI** **PvuII**
SfiI **SplI** **Tth111I**
XbaI **XcaI**

1 Site
 3696
ScrFI **CCNGG**
 25 Sites
 52 1172 1520
 1738 1751 1872
 2475
 2668 2977 3020
 3138 3270 3297
 3381
 3444 3456 3478
 3483 3484 3540
 3657
 3677 3678 3690
 3801
SdnI **GDGCHC**
 8 Sites
 652 1587 2085
 2953 3075 3114
 3391
 3685

SinI **GGWCC**
 7 Sites
 123 587 920
 1049 3022 3134
 3299
SmaI **CCCGGG**
 2 Sites
 3484 3678
SnaBI **TACGTA**
 1 Site
 2725
SpeI **ACTAGT**
 1 Site
 2384
SphI **GCATGC**
 1 Site
 3942
SspI **AATATT**
 2 Sites
 603 991
StuI **AGGCCT**

70/109

Fig. 22 cont'd-4

nefM_4.1Dmyr (hvl3225 in), (663nt.), GC=67%

CTCGAGAAGAAAATGCGGCCAAGTGGTCTGAAGAGCTCCATCGTCCGGTGGCCGGCGGTCCGGGAGAGGA
TGCGGAGGGCGGAGCCGGCGGCCGACGGGGTCTCGGGCGGTCTCGCGGGACCTGGAGAAGCACGGGGCGAT
CACGTCGAGCAACACCGCCGCGACGAACGCGGACTGCGCGTGGCTGGAGGCCCAGGAGGAGGACTCGGAG
GTCGGCTTCCCCGTCCGGCCGACGCTCCCGCTCCGGCCGATGACGTACAAGGCGGCCGTCGACCTCTCCC
ACTTCCTGAAAGAGCAGGGGGGCTGGAGGGGCTCATCTACTCGCAGAAGAGGCAGGACATCCTCGACCT
GTGGGTCTACCACACCCAGGGCTACTTCCCGGACTGGCAGAATAACGCCAGGCCCGGAATCCGCTAC
CCGCTGACCTTCGGGTGGTGTCTCAAGCTCGTCCCCGTGGACCCGCGGGAGGTGAGGAAGCCAACAAGG
GGGAGAACAACCTGCCTCCTGCACCCGATGTGCGAGCACGGGATGGACGACCCGGAGAAGGAGGTGCTGAT
GTGGAAGTTCGACTCGCGGTGGCGTTCACCACATGGCCCGGAGAAGCACCCGGAGTTCTACAAGGAC
TGCTGATAAGCTAGCTGATCAGGATCCACGCGT

MAAKWSKSSIVGWPAVRERMRAEPAADGVGAVSRDLEKHGAITSSNTAATNADCAWLEAQEEDSEVGFP
VRPHVPLRPMTYKAAVDLSHFLKEQGGLLEGLIYSQKRQDILDLWVYHTQGYFPDWQNYTPGPGRYPLTF
GWCFLKLPVDPREVEEANKGENNCLLHMPMSQHGMDDPEKEVLMWKFD SRLAFHHMAREKHPEFYKDC_

HV13319 (nefM_4.1Dmyr.wlv), 3918nt.

AAATGGGGGCGCTGAGGTCTGCCTCGTGAAGAAGGTGTTGCTGACTCATACCAGGCCTGAATCGCCCCAT
CATCCAGCCAGAAAGTGAGGGAGCCACGGTTGATGAGAGCTTTGTTGTAGGTGGACCAGTTGGTGATTTT
GAACTTTTGCTTTGCCACGGAACGGTCTGCGTTGTGGGAAGATGCGTGATCTGATCCTTCAACTCAGCA
AAAGTTCGATTTATTCAACAAAGCCGCGCTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAACCA
ATTAACCAATTCTGCGTTCAAAATGGTATGCGTTTTGACACATCCACTATATATCCGTGTCTGTTCTGTCC
ACTCCTGAATCCCATTCCAGAAATCTCTAGCGATTCCAGAAGTTTCTCAGAGTCGGAAAGTTGACCAGA
CATTACGAACTGGCACAGATGGTCATAACCTGAAGGAAGATCTGATTGCTTAACTGCTTCAGTTAAGACC
GACGCGCTCGTCGTATAACAGATGCGATGATGCAGACCAATCAACATGGCACCTGCCATTGCTACCTGTA
CAGTCAAGGATGGTAGAAATGTTGTCGGTCTTGCACACGAATATTACGCCATTTGCCCTGCATATTCAA
CAGCTCTTCTACGATAAGGGCACAATCGCATCGTGGAACGTTTGGGCTTCTACCGATTTAGCAGTTTGA
TACACTTTCTCTAAGTATCCACCTGAATCATAAATCGGCAAAATAGAGAAAAATTGACCATGTGTAAGCG
GCCAATCTGATTCCACCTGAGATGCATAATCTAGTAGAATCTCTTCGCTATCAAAATTCACCTCCACCTT
CCACTCACCGGTTGTCCATTCTATGGCTGAACTCTGCTTCTCTGTTGACATGACACACATCATCTCAATA
TCCGAATACGGACCATCAGTCTGACGACCAAGAGAGCCATAAACACCAATAGCCTTAACATCATCCCCAT
ATTTATCCAATATTCGTTCTTAATTTTCATGAACAATCTTCATTCTTTCTTCTCTAGTCATTATTATTGG
TCCGTTTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTTCATGATGATATATTT
TTATCTTGTGCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTCCCCGGCCCATGACCAAAATC
CCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATC
CTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAACACCGCTACCAGCGGTGGTTTGTGGCC
GGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCAGATACCAAATACTGTT
CTTCTAGTGTAGCCGTAGTTAGGCCACCACTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGC
TAATCCTGTTACCAAGTGGCTGCTGCCAGTGGCGATAAGTCTGTCTTACCGGGTTGGACTCAAGACGATA
GTTACCGGATAAGGCGCAGCGGTCCGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAAGC
ACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGG
CGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAACGC
CTGGTATCTTTATAGTCTGTGCGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCA
GGGGGGCGGAGCCTATGGA AAAACGCCAGCAACGCGGCCTTTTACGGTTCTTGGCCTTTTGCTGGCCTT
TTGCTCACATGTTCTTTCTGCGTTATCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGC
TGATACCGCTCGCCGACGCCAAGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTG
ATGCGGTATTTCTCTTACGCATCTGTGCGGTATTTACACCGCATATGGTGCACTCTCAGTACAATCT
GCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTGCTGAGTAGTGCGC
GAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGG
CGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATAGCCGCGCATCGATGATATCCATTGCATACGTT
GTATCTATATCATAAATATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGTTGACATTGATTA

Mosaic and CON nef_Dmyr_final.doc

71/109

Fig. 22 cont'd-5

TTGACTAGTTATTAATAGTAATCAATTACGGGGTTCATTAGTTCATAGCCCATATATGGAGTTCGCGGTTA
 CATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGAC
 GTATGTTCCCATAGTAACGCCAATAGGGACTTTCATTGACGTCAATGGGTGGAGTATTTACGGTAAACT
 GCCCCTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAAT
 GGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATT
 AGTCATCGCTATTACCATGGTGTATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCA
 CGGGGATTTCCAAGTCTCCACCCCATGACGTCAATGGGAGTTTGTGTTTGGCACCAAATCAACGGGACT
 TTCCAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTA
 TATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCAT
 AGAAGACACCGGGACCGATCCAGCCTCCGCGGGCGCGCGTCGAGAAGAAAATGGCGGCCAAGTGGTCGAA
 GAGCTCCATCGTCGGGTGGCCGGCGGTCCGGGAGAGGATGCGGAGGGCGGAGCCGGCGGCCGACGGGGTTC
 GGGGCGGTCTCGCGGGACCTGGAGAAGCACGGGGCGATCACGTCGAGCAACACCGCCGCGACGAACGCGG
 ACTGCGCGTGGCTGGAGGCCAGGAGGAGGACTCGGAGGTGCGCTTCCCGGTCCGGCCGACGTCCCGCT
 CCGGCCGATGACGTACAAGGCGGCCGTCGACCTCTCCACTTCTGAAAGAGCAGGGGGGCTGGAGGGG
 CTCATCTACTCGCAGAAGAGGCAGGACATCCTCGACCTGTGGGTCTACCACACCCAGGGCTACTTCCCGG
 ACTGGCAGAACTACACGCCAGGCCCGGAATCCGCTACCCGCTGACCTTCGGGTGGTGTCTCAAGCTCGT
 CCCCCTGGACCCGCGGGAGGTTCGAGGAAGCCAACAAGGGGGAGAACAACCTGCCTCCTGCACCCGATGTCTG
 CAGCACGGGATGGACGACCCGGAGAAGGAGGTGCTGATGTGGAAGTTGACTCGCGGCTGGCGTTCCACC
 ACATGGCCCGGAGAAGCACCCGGAGTTCTACAAGGACTGCTGATAAGCTAGCTGATCAGCCTCGACTGT
 GCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCGTGCCTTCTTGACCTGGAAGGTGCCACT
 CCCACTGTCTTCTTAATAAAATGAGGAAATGCATCGCATTTGTCTGAGTAGGTGTATTCTATTCTGG
 GGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGAATTT

Thursday, August 2, 2007

Sequence 0 Length : 3918

AatII GACGTC
 4 Sites
 2510 2563 2646
 2832
 AccI GTMKAC
 2 Sites
 3317 3404
 AflIII ACRYGT
 1 Site
 1897
 AluI AGCT
 13 Sites
 109 633 1340
 1597 1643 1733
 1959
 2184 2951 3083
 3495 3688 3692
 AlwNI CAGNNCTG
 2 Sites
 1488 2129
 AseII GRCGYC
 5 Sites
 2507 2560 2643
 2829 2983
 ApaI GTGCAC
 2 Sites
 1563 2081
 Aval CYCGRG
 1 Site
 3453
 BanI GGYRCC
 3 Sites
 538 2850 3772
 BanII GRGCTC
 3 Sites

2953 3085 3362
 BclI TGATCA
 1 Site
 3694
 BcnI CCSGG
 10 Sites
 1173 1521 3021
 3110 3269 3428
 3455
 3456 3590 3662
 BglI GCCNNNNNGGC
 3 Sites
 2475 2597 2668
 BglII AGATCT
 1 Site
 458
 Bsp1286 GDGCHC
 6 Sites
 652 1587 2085
 2953 3085 3362
 BspBI TCATGA
 2 Sites
 1007 1105
 BspNI CCWGG
 13 Sites
 52 1738 1751
 1872 2475 2668
 2977
 3169 3241 3352
 3415 3449 3766
 BspHII GCGCGC
 1 Site
 3045
 BstNI CCWGG
 13 Sites
 52 1738 1751
 1872 2475 2668
 2977

3169 3241 3352
 3415 3449 3766
 Cfr10I RCCGGY
 3 Sites
 847 3099 3132
 CfrI YGGCCR
 7 Sites
 769 3065 3097
 3137 3274 3292
 3311
 ClaI ATCGAT
 1 Site
 2267
 DdeI CTNAG
 11 Sites
 12 204 397
 711 787 1214 1623
 2088 2158 2229
 3826
 DpnI GATC
 11 Sites
 190 195 460
 1239 1247 1258
 1333
 2972 3028 3187
 3696
 DraIII CACNNNGTG
 1 Site
 1161
 Eco47I GGWCC
 9 Sites
 122 586 919
 1048 3021 3104
 3164
 3269 3506
 EcoO109 RGGNCCY
 2 Sites
 3165 3348
 EcoRII CCWGG

Mosaic and CON nef_Dmyr_final.doc

72/109

Fig. 22 cont'd-6

13 Sites
50 1736 1749
1870 2473 2666
2975
3167 3239 3350
3413 3447 3764
EcoRV GATATC
1 Site
2294
Fnu4HI GCNGC
22 Sites
234 769 1283
1489 1492 1557
1700
1855 1973 1976
1994 2110 2250
2279
2282 3065 3137
3206 3277 3311
3571
3625
FnuDII CGCG
17 Sites
494 1273 1854
2169 2257 2281
2445
3039 3045 3047
3162 3208 3217
3226
3513 3624 3650
HaeII RGCGCY
3 Sites
12 1657 2027
HaeIII GGCC
20 Sites
55 771 1175
1423 1857 1875
1886
2268 2469 2662
3067 3099 3139
3238
3276 3294 3313
3350 3452 3646
HgiAI GWGCWC
4 Sites
1587 2085 2953
3085
HhaI GCGC
14 Sites
11 496 1273
1382 1556 1656
1723
1993 2026 2169
2249 3045 3047
3226
HincII GTYRAC
4 Sites
413 886 2369
3318
HinfI GATTC
16 Sites
43 59 357
383 401 725 779
807 1527 1923
1998 2222 2795
3250
3458 3619
HinPI GCGC
14 Sites

9 494 1271
1380 1554 1654
1721
1991 2024 2167
2247 3043 3045
3224
HpaII CCGG
17 Sites
848 1172 1329
1519 1545 1692
3019
3100 3108 3133
3268 3273 3291
3427
3454 3589 3661
MaeI CTAG
7 Sites
378 801 1034
1404 2385 3689
3716
MaeII ACGT
13 Sites
669 1160 1196
2306 2507 2519
2560
2643 2724 2829
3190 3281 3301
MaeIII GTNAC
8 Sites
270 1134 1361
1477 1540 2446
2533
2882
MvaI CCNGG
23 Sites
52 1172 1520
1738 1751 1872
2475
2668 2977 3020
3109 3169 3241
3268
3352 3415 3427
3449 3454 3455
3589
3661 3766
NaeI GCCGGC
2 Sites
3101 3134
NciI CCSGG
10 Sites
1172 1520 3020
3109 3268 3427
3454
3455 3589 3661
NcoI CCATGG
1 Site
2745
NdeI CATATG
2 Sites
2076 2619
NheI GCTAGC
1 Site
3688
NlaIII CATG
15 Sites
538 762 864
892 1011 1109 1181
1901 2219 2349
2367 2689 2749
3645
3907

NlaIV GGNNCC
11 Sites
92 540 1830
1869 2852 3023
3131
3166 3349 3509
3774
NruI TCGCGA
1 Site
2257
NsiI ATGCAT
1 Site
796
Nsp7524I RCATGY
2 Sites
1901 3907
NspBII CMGCKG
6 Sites
1314 1559 2281
3039 3471 3513
PflMI CCANNNNNTGC
1 Site
3642
PpuMI RGGWCCY
1 Site
3165
PssI RGGNCCY
2 Sites
3168 3351
RsaI GTAC
10 Sites
559 2093 2263
2330 2604 2684
2717
2768 2925 3304
RsrII CGGWCCG
2 Sites
3105 3270
SacI GAGCTC
2 Sites
2953 3085
SacII CCGCGG
3 Sites
2282 3040 3514
SalI GTCGAC
1 Site
3316
Sau3A GATC
11 Sites
188 193 458
1237 1245 1256
1331
2970 3026 3185
3694
Sau96A GGNNCC
17 Sites
123 587 920
1049 1174 2266
2468
2661 3022 3105
3165 3237 3270
3348
3451 3507 3645
ScrFI CCNGG
23 Sites
52 1172 1520
1738 1751 1872
2475
2668 2977 3020
3109 3169 3241
3268

Mosaic and CON nef_Dmyr_final.doc

73/109

Fig. 22 cont'd-7

3352	3415	3427	3907		3453		
3449	3454	3455	SspI	AATATT	XmaIII	CGGCCG	
3589			2 Sites		4 Sites		
3661	3766		603	991	3137	3274	3292
SdnI	GDGCHC		StuI	AGGCCT	3311		
6 Sites			1 Site		XmnI	GAANNNTTC	
652	1587	2085	55		1 Site		
2953	3085	3362	StyI	CCWGGG	811		
SinI	GGWCC		1 Site				
9 Sites			2745		Following enzymes have no sites		
123	587	920	TaqI	TCGA	AccIII	AflII	ApaI
1049	3022	3105	11 Sites		Asp718	AsuII	AvrII
3165			216	1799	2287		
3270	3507		3050	3076	3193	BalI	BamHI
SmaI	CCCGGG		3317			BspMII	BstEII
1 Site			3392	3521	3617		BstXI
3455			3703			DraI	Eco47III
SnaBI	TACGTA		Tth111I	GACNNNGTC		EspI	EcoRI
1 Site			1 Site			HpaI	FspI
2725			3145			HpaI	KpnI
SpeI	ACTAGT		XhoII	RGATCY		MstI	MluI
1 Site			3 Sites			OxaNI	NarI
2384			458	1245	1256	PstI	NotI
SphI	GCATGC		XmaI	CCCGGG		PvuII	PvuI
1 Site			1 Site			ScaI	SfiI
						SplI	XbaI
						XhoI	XcaI

Following enzymes have no sites

AccIII	AflII	Apal
Asp718	AsuII	AvrII
BalI	BamHI	BbeI
BspMII	BstEII	BstXI
DraI	Eco47III	EcoRI
EspI	FspI	HindIII
HpaI	KpnI	MluI
MstI	NarI	NotI
OxaNI	PstI	PvuI
PvuII	ScaI	SfiI
SplI	XbaI	XcaI
XhoI		

nefM_4.2Dmyr (654nt.) hv13231, GC=66%

ctcgcagAAGAAAATGCGGCCAAGTGGTTCGAAGCGGTCCATCGTCGGGTGGCCGGCGATCCGGGAGAGGA
 TCGCGAGGACGGAGCCGGCGGCCGAGGGGGTTCGGGGCGGCGTCGCAGGACCTCGACAAATACGGGGCGCT
 CACGTCGAGCAACACCGCCAGACGAACCCGACTGCGCGTGGCTGGAGGCCAGGAGGAGGAAGAGGTC
 GGCTTCCCGGTCCGGCCGCAAGTCCCAGTCAGGCCGATGACGTACAAGGGCGCCCTGGACCTCTCCCACT
 TCCTGAAAGAGAAGGGGGGCTGGACGGGCTCATCTACTCGCGGAAGAGGGCAGGAGATCCTCGACCTGTG
 GGTCTACAACACCCAGGGCTACTTCCCCGACTGGCAGAACTACACGCCAGGCCCGGAGTGCGGTACCCG
 CTGACCTTTCGGGTGGTGCTACAAGCTCGTCCCCGTGGACCCGGAGGAGGTTCGAGAAGGCCAACGAGGGGG
 AGAACAACCTCCCTCCTGCACCCGATGTGCTCCACGGGATGGAGGACCCCGAGAGGGAGGTGCTGAAGTG
 GAAGTTCGACTCGCGGTGGCGCTCAAGCACCGGGCCCGGGAGCTCCACCCGGAGTTCTACAAGGACTGC
 TGATAAGCTAGCGGATCCTGATCA

MAAKWSKRSIVGWPAIRERMRRTEPAAEGVGAASQDLDKYGALTSSNTAQTNPDCAWLEAQEEEEVGFPV
 RPQVPVRPMTYKGALDLSHFLKEKGGLDGLIYSRKRQEILDWVYNTQGYFPDWQNYTPGPGVRYPLTFG
 WCYKLVVDPEEVEKANEGENNSLLHPMSLHGMEDEPEREVLKWKFDSSLALKHRARELHPEFYKDC_

>HV13231 in hv10001 (nefM_4.2Dmyr.wlv), 3950nt.

AAATGGGGGCGCTGAGGTCTGCCTCGTGAAGAAGGTGTTGCTGACTCATACCAGGCCTGAATCGCCCCAT
 CATCCAGCCAGAAAGTGAGGGAGCCACGGTTGATGAGAGCTTTGTTGTAGGTGGACCAGTTGGTGATTTT
 GAACTTTTGCTTTGCCACGGAACGGTCTGCGTTGTGGGAAGATGCGTGATCTGATCCTTCAACTCAGCA
 AAAGTTCGATTTATTCAACAAAGCCGCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAACCA
 ATTAACCAATTCTGCGTTCAAAATGGTATGCGTTTGTACACATCCACTATATATCCGTGTGCTTCTGTCC
 ACTCCTGAATCCCATTCCAGAAATTCTCTAGCGATTCCAGAAGTTTCTCAGAGTCGGAAAGTTGACCAGA
 CATTACGAACTGGCACAGATGGTCATAACCTGAAGGAAGATCTGATTGCTTAACTGCTTCAGTTAAGACC
 GACGCGCTCGTCGTATAACAGATGCGATGATGCAGACCAATCAACATGGCACCTGCCATTGCTACCTGTA
 CAGTCAAGGATGGTAGAAATGTTGTGCGTCTTGCACACGAATATTACGCCATTTGCCTGCATATTCAA
 CAGCTCTTCTACGATAAGGGCACAAATCGCATCGTGGAACGTTTGGGCTTCTACCGATTTAGCAGTTTGA
 TACACTTTCTCTAAGTATCCACCTGAATCATAAATCGGCAAAATAGAGAAAAATTGACCATGTGTAAGCG
 GCCAATCTGATTCCACCTGAGATGCATAATCTAGTAGAATCTCTTCGCTATCAAAATTCACCTCCACCTT
 CCACTCACCGGTTGTCCATTCATGGCTGAACTCTGCTTCTCTGTGACATGACACACATCATCTCAATA
 TCCGAATACGGACCATCAGTCTGACGACCAAGAGAGCCATAAACACCAATAGCCTTAACATCATCCCCAT

Mosaic and CON nef_Dmyr_final.doc

74/109

Fig. 22 cont'd-8

ATTTATCCAATATTCGTTCTTAATTTTCATGAACAATCTTCATTCTTTCTTCTCTAGTCATTATTATTGG
 TCCGTTTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTTCATGATGATATATTT
 TTATCTTGTGCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTCCCCGGCCCCATGACCAAAATC
 CCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATC
 CTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAACACCGCTACCAGCGGTGGTTTGTGTTGCC
 GGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACCTGGCTTCAGCAGAGCGCAGATACCAATACTGTT
 CTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGC
 TAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATA
 GTTACCGGATAAGGCGCAGCGGTGGGGCTGAACGGGGGGTTCGTGCACACAGCCAGCTTGGAGCGAACG
 ACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGG
 CGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGCGCACGAGGGAGCTTCCAGGGGGAAACGC
 CTGGTATCTTTATAGTCTGTGCGGTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCA
 GGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTT
 TTGCTCACATGTTCTTCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGC
 TGATACCGCTCGCCGAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTG
 ATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTTACACCCGCATATGGTGCACCTCTCAGTACAATCT
 GCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGC
 GAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGG
 CGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATAGCCGCGGCATCGATGATATCCATTGCATACGTT
 GTATCTATATCATAATATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGTTGACATTGATTA
 TTGACTAGTTATTAATAGTAATCAATTACGGGGTTCATTAGTTTCATAGCCCATATATGGAGTTCGCGTTA
 CATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCCGCCATTGACGTCAATAATGAC
 GTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACT
 GCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAAT
 GGCCCGCCTGGCATTATGCCAGTACATGACCTTACGGGACTTTCTACTTGGCAGTACATCTACGTATT
 AGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCA
 CGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTTGTTTTGGCACCAAAATCAACGGGACT
 TTCCAAAATGTCGTAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTA
 TATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCAT
 AGAAGACACCGGGACCGATCCAGCCTCCGCGGGCGCGCTCGACGGGCGACGCGAAACTTGGGCCCACTC
 GAGAAGAAAAATGGCGGCCAAGTGGTTCGAAGCGGTCCATCGTGGGTGGCCGGCGATCCGGGAGAGGATGC
 GGAGGACGGAGCCGGCGGCCGAGGGGGTTCGGGGCGGCGTTCGAGGACCTCGACAAATACGGGGCGCTCAC
 GTCGAGCAACACCGCCAGACGAACCCGGACTGCGCGTGGCTGGAGGCCAGGAGGAGGAAGAGGTGCGC
 TTCCCGGTCCGGCCGCAAGTCCAGTCAGGCGGATGACGTACAAGGGCGCCCTGGACCTCTCCCACTTCC
 TGAAAGAGAAGGGGGGCTGGACGGGCTCATCTACTCGCGGAAGAGGCAGGAGATCTCGACCTGTGGGT
 CTACAACACCCAGGGCTACTTCCCCGACTGGCAGAACTACACGCCAGGCCCGGAGTGCGGTACCCGCTG
 ACCTTCGGGTGGTGCTACAAGCTCGTCCCCGTGGACCCGGAGGAGGTCGAGAAGGCCAACGAGGGGGAGA
 ACAACTCCCTCCTGCACCCGATGTGCTCCACGGGATGGAGGACCCCGAGAGGGAGGTGCTGAAGTGGAA
 GTTCGACTCGCGGCTGGCGCTCAAGCACCGGGCCCGGGAGCTCCACCCGGAGTTCTACAAGGACTGCTGA
 TAAGCTAGCGGATCCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCC
 CCGTGCCTTCTTGACCCCTGGAAGGTGCCACTCCCCTGTCCTTTCTTAATAAAATGAGGAAATTGCATC
 GCATTGTCTGAGTAGGTGTCTATTCTATTTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGG
 GAAGACAATAGCAGGCATGCTGGGGAATTT

Thursday, August 2, 2007

Sequence 0 Length : 3950

AatII GACGTC

4 Sites

2510 2563 2646

2832

AccI GTMKAC

2 Sites

3050 3430

AflIII ACRYGT

1 Site

1897

AluI AGCT

12 Sites

109 633 1340

1597 1643 1733

1959

2184 2951 3521

3680 3714

AlwNI CAGNNCTG

2 Sites

1488

AosII GRCGYC

7 Sites

2507

2829

3337

ApaI GGGCCC

2 Sites

3075

ApaLI GTGCAC

2129

2560

2643

2983

3186

Mosaic and CON nef_Dmyr_final.doc

Fig. 22 cont'd-9

SUBSTITUTE SHEET (RULE 26)

76/109

Fig. 22 cont'd-10

52 1172 1520
 1738 1751 1872
 2475
 2668 2977 3020
 3138 3246 3270
 3294
 3342 3378 3441
 3475 3481 3537
 3669
 3674 3675 3687
 3798
 NaeI GCCGGC
 2 Sites
 3130 3163
 NarI GGCGCC
 1 Site
 3337
 NciI CCSGG
 12 Sites
 1172 1520 3020
 3138 3246 3294
 3481
 3537 3669 3674
 3675 3687
 NcoI CCATGG
 1 Site
 2745
 NdeI CATATG
 2 Sites
 2076 2619
 NheI GCTAGC
 1 Site
 3714
 NlaIII CATG
 14 Sites
 538 762 864
 892 1011 1109 1181
 1901 2219 2349
 2367 2689 2749
 3939
 NlaIV GGNNCC
 17 Sites
 92 540 1830
 1869 2852 3023
 3073
 3160 3338 3375
 3479 3492 3535
 3613
 3672 3722 3806
 NruI TCGCGA
 1 Site
 2257
 NsiI ATGCAT
 1 Site
 796
 Nsp7524I RCATGY
 2 Sites
 1901 3939
 NspBII CMGCKG
 5 Sites
 1314 1559 2281
 3039 3497
 PflMI CCANNNNTGG
 1 Site
 3605

PpuMI RGGWCCY
 2 Sites
 3194 3611
 PssI RGGNCCY
 4 Sites
 3197 3377 3480
 3614
 RsaI GTAC
 11 Sites
 559 2093 2263
 2330 2604 2684
 2717
 2768 2925 3330
 3492
 RsrII CGGWCCG
 1 Site
 3296
 SacI GAGCTC
 2 Sites
 2953 3682
 SacII CCGCGG
 2 Sites
 2282 3040
 SalI GTCGAC
 1 Site
 3049
 Sau3A GATC
 13 Sites
 188 193 458
 1237 1245 1256
 1331
 2970 3026 3133
 3412 3720 3726
 Sau96A GGNNCC
 22 Sites
 123 587 920
 1049 1174 2266
 2468
 2661 3022 3071
 3072 3112 3194
 3266
 3296 3344 3374
 3477 3533 3611
 3670
 3671
 ScrFI CCNGG
 25 Sites
 52 1172 1520
 1738 1751 1872
 2475
 2668 2977 3020
 3138 3246 3270
 3294
 3342 3378 3441
 3475 3481 3537
 3669
 3674 3675 3687
 3798
 SdnI GDGCHC
 8 Sites
 652 1587 2085
 2953 3075 3388
 3674
 3682
 SinI GGWCC

11 Sites
 123 587 920
 1049 3022 3112
 3194
 3296 3344 3533
 3611
 SmaI CCCGGG
 1 Site
 3675
 SnaBI TACGTA
 1 Site
 2725
 SpeI ACTAGT
 1 Site
 2384
 SphI GCATGC
 1 Site
 3939
 SspI AATATT
 2 Sites
 603 991
 StuI AGGCCT
 1 Site
 55
 StyI CCWWGG
 1 Site
 2745
 TaqI TCGA
 12 Sites
 216 1799 2287
 3050 3079 3105
 3199
 3222 3418 3547
 3643 3735
 XhoI CTCGAG
 1 Site
 3078
 XhoII RGATCY
 5 Sites
 458 1245 1256
 3412 3720
 XmaI CCCGGG
 1 Site
 3673
 XmaIII CGGCCG
 2 Sites
 3166 3300
 XmnI GAANNNTTC
 1 Site
 811

Following enzymes have no sites

AccIII	AflII	AsuII
AvrII	BalI	BspMI
BstEII	BstXI	DraI
Eco47III	EcoRI	EspI
FspI	HindIII	HpaI
MluI	MstI	NotI
OxaNI	PstI	PvuI
PvuII	ScaI	SfiI
SplI	Tth111I	XbaI
XcaI		

>nefM_4.3Dmyr(654nt.), hv13230, GC=66%

ctcgagAAGAAAATGGCGGCCAACTGGTCAAGAGCTCCATCGTCCGGTGGCCGGAGATCCGGGAGCGCA
 TCCGGAGGACGACCCGGCGGCCGAGGGGGTCCGGGCGGCCTCGCGGGACCTGGAGCGCCACGGGGCGAT

Mosaic and CON nef_Dmyr_final.doc

77/109

Fig. 22 cont'd-11

CACGTCGAGCAACACCGCCACCACGAACGCGGCCTGCGCGTGGCTGGAGGCCAGGAGGACGAGGAAGTC
GGCTTCCCGGTCCGGCCGCAAGTCCCGCTCAGGCCGATGACGTTCAAGGGGGCCTTCGACCTCGGCTTCT
TCCTGAAAGAGAAGGGGGGCTGGAGGGGCTCATCTACTCGAAGAAGAGGCAGGACATCCTCGACCTGTG
GGTCTACAACACCCAGGGGTTCTTCCCGGACTGGCAGAACTACACGCCAGGCCCGGAACCCGGTTCCTCG
CTGACCTTCGGGTGGTGCTTCGAGCTCGTCCCGTGGACCCGAAGGAGGTCGAGGAAGCCAACGAGGGGG
AGAACAACCTGCCTCCTGCACCCGATCTGCCAGCACGGGATGGACGACGAGGAGCGCGAGGTGCTGATGTG
GAAGTTCGACTCGTCCCTGGCCCGGCGCCACCTCGCCCGGAGAAGCACCCGGAGTACTACAAGGACTGC
TGATAAGCTAGCGGATCCTGATCA

MAANWSKSSIVGWPEIRERIRRTDPAAEGVGAASRDLEHGAITSSNTATTNAACAWLEAQEDEEVGFPV
RPQVPLRPMTFKGAFDLGFLLKEKGGLEGLIYSKKRQDILDWVYNTQGFPPDWQNYTPGPGTRFPLTFG
WCFELVPVDPKEVEEANEGENNCLLHPICQHGMDDEREVLWVKFDSLSLARRHLAREKHPEYVKDC_

>hvl3230 in hvl0001 (nefM_4.3Dmyr.wlv) (3950nt.)

AAATGGGGGCGCTGAGGTCTGCCTCGTGAAGAAGGTGTTGCTGACTCATACCAGGCCTGAATCGCCCCAT
CATCCAGCCAGAAAGTGAGGGAGCCACGGTTGATGAGAGCTTTGTTGTAGGTGGACCAGTTGGTGATTTT
GAACTTTTGCTTTGCCACGGAACGGTCTGCGTTGTCTGGGAAGATGCGTGATCTGATCCTTCAACTCAGCA
AAAGTTCGATTTATTCAACAAAGCCGCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAACCA
ATTAACCAATTCTGCGTTCAAATGGTATGCGTTTTGACACATCCACTATATATCCGTGTCTGTTCTGTCC
ACTCCTGAATCCCATTCCAGAAATTCTCTAGCGATTCCAGAAGTTTCTCAGAGTCGGAAAGTTGACCAGA
CATTACGAACTGGCACAGATGGTCATAACCTGAAGGAAGATCTGATTGCTTAACTGCTTCAGTTAAGACC
GACGCGCTCGTCGTATAACAGATGCGATGATGCAGACCAATCAACATGGCACCTGCCATTGCTACCTGTA
CAGTCAAGGATGGTAGAAATGTTGTCTGCTCCTTGACACAGCAATATTACGCCATTTGCTGTCATATTCAA
CAGCTCTTCTACGATAAGGGCACAAATCGCATCGTGGAACGTTTGGGCTTCTACCGATTTAGCAGTTTGA
TACACTTTCTCTAAGTATCCACCTGAATCATAAATCGGCAAAATAGAGAAAATTGACCATGTGTAAAGCG
GCCAATCTGATTCCACCTGAGATGCATAATCTAGTAGAATCTCTTCGCTATCAAAATTCACCTCCACCTT
CCACTCACCGGTTGTCCATTCTGCTGAACCTCTGCTTCTCTGTTGACATGACACACATCATCTCAATA
TCCGAATACGGACCATCAGTCTGACGACCAAGAGAGCCATAAACACCAATAGCCTTAACATCATCCCCAT
ATTTATCCAATATTCGTTCTTAAATTTTATGAACAATCTTCATTCTTTCTTCTCTAGTCATTATTATTGG
TCCGTTTATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTTCATGATGATATATTT
TTATCTTGTGCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTCCCGGGCCCATGACCAAATC
CCTTAACGTGAGTTTTCTGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATC
CTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAAACAAAAAACACCGCTACCAGCGGTGGTTTGTGTTGCC
GGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACCTGGCTTCAGCAGAGCGCAGATACCAAATACTGTT
CTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCAGCTACATACCTCGCTCTGC
TAATCCTGTTACCAAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATA
GTTACCGGATAAGGCGCAGCGGTGCGGCTGAACGGGGGGTTCTGTGCACACAGCCAGCTTGGAGCGAACG
ACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGG
CGGACAGGTATCCGGTAAGCGGCAGGGTTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGC
CTGGTATCTTTATAGTCTGTGCGGTTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCA
GGGGGGCGGAGCCTATGGAACAAACGCCAGCAACGCGGCCTTTTACGGTTCTTGCCCTTTTGCTGGCCTT
TTGCTCACATGTTCTTTCTGCGTTATCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGC
TGATACCGCTCGCCGACCGCAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTG
ATGCGGTATTTTCTCTTACGCATCTGTGCGGTATTTACACCCGCATATGGTGCACTCTCAGTACAATCT
GCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGC
GAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGG
CGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATAGCCGCGCATCGATGATATCCATTGCATACGTT
GTATCTATATCATAATATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGTTGACATTGATTA
TTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCATATATGGAGTTCCGCGTTA
CATAACTTACGGTAATGGCCCGCTGGCTGACCGCCCAACGACCCCGCCCATTTGACGTCAATAATGAC
GTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAACT

Mosaic and CON nef_Dmyr_final.doc

78/109

Fig. 22 cont'd-12

GCCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAAT
 GGCCCGCCTGGCATTATGCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATT
 AGTCATCGCTATTACCATGGTGATGCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCA
 CGGGGATTTCCAAGTCTCCACCCCATTTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACT
 TTCCAAAATGTCGTAACAACCTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTA
 TATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCAT
 AGAAGACACCGGGACCGATCCAGCCTCCGCGGGCGCGCTCGACGGGCGACGCGAAACTTGGGCCCCACTC
 GAGAAGAAAAATGGCGGCCAACTGGTGAAGAGCTCCATCGTGGGTGGCCGGAGATCCGGGAGCGCATCC
 GGAGGACGGACCCGGCGGCCGAGGGGGTTCGGGGCGGCCTCGCGGGACCTGGAGCGCCACGGGGCGATCAC
 GTCGAGCAACACCGCCACCACGAACGCGGCCTGCGCTGGCTGGAGGCCAGGAGGACGAGGAAGTCGGC
 TTCCCGGTCCGGCCGCAAGTCCCGCTCAGGCCGATGACGTTCAAGGGGGCCTTCGACCTCGGCTTCTTCC
 TGAAAGAGAAGGGGGGCTGGAGGGGCTCATCTACTCGAAGAAGAGGCAGGACATCCTCGACCTGTGGGT
 CTACAACACCCAGGGGTTCTTCCCGGACTGGCAGAACTACACGCCAGGCCCGGGAACCCGGTTCCCGCTG
 ACCTTCGGGTGGTGCTTCGAGCTCGTCCCGTGGACCCGAAGGAGGTGAGGAAGCCAACGAGGGGGAGA
 ACAACTGCCTCCTGCACCCGATCTGCCAGCACGGGATGGACGACGAGGAGCGCGAGGTGCTGATGTGGAA
 GTTCGACTCGTCCCTGGCCCCGGCGCCACCTCGCCCGGGAGAAGCACCCGGAGTACTACAAGGACTGCTGA
 TAAGCTAGCGGATCCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCC
 CCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCCTGTCTTTTCTAATAAAATGAGGAAATTGCATC
 GCATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGG
 GAAGACAATAGCAGGCATGCTGGGGAATTT

Thursday, August 2, 2007

Sequence 0 Length : 3950

AatII GACGTC
 4 Sites
 2510 2563 2646
 2832
 AccI GTMKAC
 2 Sites
 3050 3430
 AccIII TCCGGA
 1 Site
 3148
 AflIII ACRYGT
 1 Site
 1897
 AluI AGCT
 12 Sites
 109 633 1340
 1597 1643 1733
 1959
 2184 2951 3112
 3521 3714
 AlwNI CAGNNNCTG
 2 Sites
 1488 2129
 AseII GRCGYC
 6 Sites
 2507 2560 2643
 2829 2983 3662
 ApaI GGGCCC
 1 Site
 3075
 ApaLI GTGCAC
 2 Sites
 1583 2081
 AvaI CYCGRG
 3 Sites
 3078 3479 3673
 BamHI GGATCC
 1 Site
 3720

BanI GGYRCC
 4 Sites
 538 2850 3661
 3804
 BanII GRGCTC
 5 Sites
 2953 3075 3114
 3388 3523
 BbeI GGGGCC
 1 Site
 3665
 BclI TGATCA
 1 Site
 3726
 BcnI CCSGG
 14 Sites
 1173 1521 3021
 3139 3163 3295
 3454
 3481 3482 3489
 3660 3675 3676
 3688
 BglI GCCNNNNNGGC
 3 Sites
 2475 2597 2668
 BglII AGATCT
 1 Site
 458
 Bsp1286 GDGCHC
 8 Sites
 652 1587 2085
 2953 3075 3114
 3388
 3523
 BspHI TCATGA
 2 Sites
 1007 1105
 BspMII TCCGGA
 1 Site
 3148
 BspNI CCWGG
 14 Sites

52 1738 1751
 1872 2475 2668
 2977
 3198 3270 3378
 3441 3475 3654
 3798
 BshII GCGCGC
 1 Site
 3045
 BstNI CCWGG
 14 Sites
 52 1738 1751
 1872 2475 2668
 2977
 3198 3270 3378
 3441 3475 3654
 3798
 Cfr10I RCCGGY
 1 Site
 847
 CfrI YGGCCR
 5 Sites
 769 3094 3126
 3166 3300
 ClaI ATCGAT
 1 Site
 2287
 DdeI CTNAG
 12 Sites
 12 204 397
 711 787 1214 1623
 2088 2158 2229
 3315 3858
 DpnI GATC
 14 Sites
 190 195 460
 1239 1247 1258
 1333
 2972 3028 3135
 3216 3591 3722
 3728
 DraIII CACNNNGTG
 1 Site

Mosaic and CON nef_Dmyr_final.doc

79/109

Fig. 22 cont'd-13

1161				4 Sites				3480	3481	3488
Eco47I	GGWCC			413	886	2369		3659	3674	3675
9 Sites				3051				3687		
122	586	919		HinfI	GANTC			NcoI	CCATGG	
1048	3021	3157		14 Sites				1 Site		
3193				43	59	357		2745		
3295	3532			383	401	725	779	NdeI	CATATG	
EcoO109	RGGNCCY			807	1527	1923		2 Sites		
3 Sites				1998	2222	2795		2076	2619	
3194	3337	3374		3645				NheI	GCTAGC	
EcoRII	CCWGG			HinPI	GCGC			1 Site		
14 Sites				18 Sites				3714		
50	1736	1749		9	494	1271		NlaIII	CATG	
1870	2473	2666		1380	1554	1654		14 Sites		
2975				1721				538	762	864
3196	3268	3376		1991	2024	2167		892	1011	1109
3439	3473	3652		2247	3043	3045		1901	2219	2349
3796				3143				2367	2689	2749
EcoRV	GATATC			3203	3253	3620		3939		
1 Site				3662				NlaIV	GGNNCC	
2294				HpaII	CCGG			17 Sites		
Fnu4HI	GCNGC			19 Sites				92	540	1830
20 Sites				848	1172	1329		1869	2852	3023
234	769	1283		1519	1545	1692		3073		
1489	1492	1557		3019				3160	3195	3338
1700				3129	3137	3149		3375	3485	3492
1855	1973	1976		3162	3294	3299		3535		
1994	2110	2250		3453				3663	3722	3806
2279				3480	3488	3659		NruI	TCGCGA	
2282	3094	3166		3674	3687			1 Site		
3184	3247	3303		MaeI	CTAG			2257		
FnuDII	CGCG			7 Sites				NsiI	ATGCAT	
15 Sites				378	801	1034		1 Site		
494	1273	1854		1404	2385	3715		796		
2169	2257	2281		3748				Nsp7524I	RCATGY	
2445				MaeII	ACGT			2 Sites		
3039	3045	3047		12 Sites				1901	3939	
3062	3191	3246		669	1160	1196		NspBII	CMGCKG	
3255				2306	2507	2519		5 Sites		
3622				2560				1314	1559	2281
HaeII	RGCGCY			2643	2724	2829		3039	3497	
5 Sites				3219	3327			PpuMI	RGGWCCY	
12	1657	2027		MaeIII	GTNAC			1 Site		
3206	3665			8 Sites				3194		
HaeIII	GGCC			270	1134	1361		PssI	RGGNCCY	
23 Sites				1477	1540	2446		3 Sites		
55	771	1175		2533				3197	3340	3377
1423	1857	1875		2882				RsaI	GTAC	
1886				MvaI	CCNGG			10 Sites		
2268	2469	2662		28 Sites				559	2093	2263
3073	3096	3128		52	1172	1520		2330	2604	2684
3168				1738	1751	1872		2717		
3186	3249	3267		2475				2768	2925	3693
3302	3320	3339		2668	2977	3020		RsrII	CGGWCCG	
3376				3138	3162	3198		1 Site		
3478	3657			3270				3296		
HgiAI	GWGCWC			3294	3378	3441		SacI	GAGCTC	
5 Sites				3453	3475	3480		3 Sites		
1587	2085	2953		3481				2953	3114	3523
3114	3523			3488	3654	3659		SacII	CCGCGG	
HhaI	GCGC			3674	3675	3687		2 Sites		
18 Sites				3798				2282	3040	
11	496	1273		NarI	GGCGCC			SalI	GTCGAC	
1382	1556	1656		1 Site				1 Site		
1723				3662				3049		
1993	2026	2169		NciI	CCSGG			Sau3A	GATC	
2249	3045	3047		14 Sites				14 Sites		
3145				1172	1520	3020		188	193	458
3205	3255	3622		3138	3162	3294		1237	1245	1256
3664				3453				1331		
HincII	GTYRAC									
Mosaic and CON nef_Dmyr_final.doc										

80/109

Fig. 22 cont'd-14

2970 3026 3133
 3214 3589 3720
 3726
 Sau96A GGNCC
 20 Sites
 123 587 920
 1049 1174 2266
 2468
 2661 3022 3071
 3072 3158 3194
 3266
 3296 3337 3374
 3477 3533 3656
 ScaI AGTACT
 1 Site
 3693
 ScrFI CCNCG
 28 Sites
 52 1172 1520
 1738 1751 1872
 2475
 2668 2977 3020
 3138 3162 3198
 3270
 3294 3378 3441
 3453 3475 3480
 3481
 3488 3654 3659
 3674 3675 3687
 3798
 SdnI GDGCHC
 8 Sites
 652 1587 2085
 2953 3075 3114
 3388

3523
 SinI GGWCC
 9 Sites
 123 587 920
 1049 3022 3158
 3194
 3296 3533
 SmaI CCCGGG
 2 Sites
 3481 3675
 SnaBI TACGTA
 1 Site
 2725
 SpeI ACTAGT
 1 Site
 2384
 SphI GCATGC
 1 Site
 3939
 SapI AATATT
 2 Sites
 603 991
 StuI AGGCCT
 1 Site
 55
 StyI CCWWGG
 1 Site
 2745
 TaqI TCGA
 14 Sites
 216 1799 2287
 3050 3079 3105
 3222

3343 3396 3418
 3517 3547 3643
 3735
 XhoI CTCGAG
 1 Site
 3078
 XhoII RGATCY
 5 Sites
 458 1245 1256
 3133 3720
 XmaI CCCGGG
 2 Sites
 3479 3673
 XmaIII CGGCCG
 2 Sites
 3166 3300
 XmnI GAANNNTTC
 1 Site
 811
 Following enzymes have no sites
 AflII Asp718 AsuII
 AvrII BalI BstEII
 BstXI DraI
 Eco47III EcoRI EspI
 FspI
 HindIII HpaI KpnI
 MluI MstI NaeI
 NotI OxaNI PflMI
 PstI PvuI PvuII
 SfiI SplI Tth111I
 XbaI XcaI

nefM_4.4Dmyr (657nt.) hv13232, GC=66%

ctcgagAAGAAATGCGCCTCCAAGTGGTTCGAAGAGCTCCATCGTCGGGTGGCCGCAAGTCCGGGAGCGCA
 TCCGGCAGACGCCCCCGCGCCGAGGGGGTTCGGCGCGGTCTCGCAGGACCTCGACAAGCACGGGGCGGT
 CACGTCGAGCAACACCGCCGCGAACAACGCGGACTGCGCGTGGCTGCAGGCCAGGAGGAAGAGGAAGAG
 GTCGGCTTCCCGGTCAAGCCGCAAGTGCCGCTCAGGCCGATGACGTACAAGGGGGCCTTCGACCTCTCGT
 TCTTCCTGAAAGAGAAGGGTGGCCTGGACGGGCTCATCTACTCGAAGAAGAGGCAGGAGATCCTCGACCT
 GTGGGTCTACCACACCCAGGGGTCTTCCCGGACTGGCACAACCTACACGCCAGGCCCGGGAACCCGCTAC
 CCGCTGTGCTTCGGGTGGTGCTTCAAGCTCGTCCCATGGACCCGGCCGAGGTTCGAGGAAGCCACCGAGG
 GCGAGAACAACCTCCCTCCTGCACCCGATCTGCCAGCACGGGATGGAGGACGAGGACCGCGAGGTGCTGGT
 CTGGCGGTTCGACTCGTCCCTGGCCCCGGCGCCACATCGCCCCGGGAGCTCCACCCGGAGTACTACAAGGAC
 TGCTGATAAGCTAGCGGATCCTGATCA

MASKWSKSSIVGWPQVRERIRQTPPAEAGVAVSQDLDKHGAVTSSNTAANNADCAWLQAQEEEEEVGFP
 VKPQVPLRPMTYKGAFDLSFFLKEKGGLDGLIYSKKRQEILDWVYHTQGFFPDWHNYTPGPGTRYPLCF
 GWCFLVPMDDPAEVEEATEGENNSLLHPICQHGMEDEDREVLVWRFDSSLARRHIARELHPEYYKDC_

>HV13232 in hv10001 (nefM_4.4Dmyr.wlv), 3953nt.

AAATGGGGGCGCTGAGGTCTGCCTCGTGAAGAAGGTGTTGCTGACTCATACCAGGCCTGAATCGCCCCAT
 CATCCAGCCAGAAAGTGAGGGAGCCACGGTTGATGAGAGCTTTGTTGTAGGTGGACCAGTTGCTGATTTT
 GAACTTTTGCTTTGCCACGGAACGGTCTGCGTTGTGCGGAAGATGCGTGATCTGATCCTTCAACTCAGCA
 AAAGTTCGATTTATTCAACAAAGCCGCCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAACCA
 ATTAACCAATTCTGCGTTCAAAATGGTATGCGTTTGTACACATCCACTATATATCCGTGTGCTTCTGTCC
 ACTCCTGAATCCCATTCCAGAAATTCTCTAGCGATTCAGAAAGTTTCTCAGAGTCGGAAAGTTGACCAGA
 CATTACGAAGTGGCACAGATGGTCATAACCTGAAGGAAGATCTGATTGCTTAACTGCTTCAGTTAAGACC

Mosaic and CON nef_Dmyr_final.doc

81/109

Fig. 22 cont'd-15

GACGCGCTCGTCGTATAACAGATGCGATGATGCAGACCAATCAACATGGCACCTGCCATTGCTACCTGTA
CAGTCAAGGATGGTAGAAATGTTGTGGTCCTTGACACGAATATTACGCCATTTGCCTGCATATTCAAA
CAGCTCTTCTACGATAAGGGCACAAATCGCATCGTGGAACGTTTGGGCTTCTACCGATTTAGCAGTTTGA
TACACTTTCTCTAAGTATCCACCTGAATCATAAATCGGCAAAATAGAGAAAAATTGACCATGTGTAAGCG
GCCAATCTGATTCCACCTGAGATGCATAATCTAGTAGAATCTCTTCGCTATCAAAATTCACCTCCACCTT
CCACTCACCGGTTGTCCATTCATGGCTGAACTCTGCTTCCTCTGTTGACATGACACACATCATCTCAATA
TCCGAATACGGACCATCAGTCTGACGACCAAGAGAGCCATAAACACCAATAGCCTTAACATCATCCCCAT
ATTTATCCAATATTTCGTTCCCTTAATTTTCATGAACAATCTTCATTCTTTCTTCTCTAGTCATTATTATTGG
TCCGTTTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTTCATGATGATATATTT
TTATCTTGTGCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTCCTCCGGCCCATGACCAAAATC
CCTTAACGTGAGTTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATC
CTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAACCCGCTACCAGCGGTGGTTTGTGTTGCC
GGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCAGATACCAAATACTGTT
CTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGC
TAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCTGTCTTACCAGGTTGGACTCAAGACGATA
GTTACCAGGATAAGGCGCAGCGGTGGGCTGAACGGGGGTTCTGTGCACACAGCCAGCTTGGAGCGAACG
ACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGG
CGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAACGC
CTGGTATCTTTATAGTCCTGTGCGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCA
GGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCCTTTTTACGGTTCTTGGCCTTTTGTGCGCCTT
TTGCTCACATGTTCTTTCTGCGTTATCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGC
TGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTG
ATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTACACCCGCATATGGTGCACTCTCAGTACAATCT
GCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGC
GAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGG
CGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATAGCCGCGGCATCGATGATATCCATTGCATACGTT
GTATCTATATCATAATATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGTTGACATTGATTA
TTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCATATATGGAGTTCCGCGTTA
CATAACTTACGGTAAATGGCCCCCTGGCTGACCGCCCAACGACCCCGCCCATTGACGTCAATAATGAC
GTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACT
GCCCCTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAAT
GGCCCCCTGGCATTATGCCAGTACATGACCTTACGGGACTTTCTACTTGGCAGTACATCTACGTATT
AGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGATAGCGTTTTGACTCA
CGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACT
TTCCAAAATGTCGTAACAACCTCCGCCCCATTGACGCAATGGGCGGTAGGCGTGACGGTGGGAGGTCTA
TATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCAT
AGAAGACACCGGGACCGATCCAGCCTCCGCGGGCGCGGTGACGGGCGACGCGAACTTGGGCCCACTC
GAGAAGAAAATGGCCTCCAAGTGGTTCGAAGAGCTCCATCGTGGGTGGCCGCAAGTCCGGGAGCGCATCC
GGCAGACGCCCCCGCGGCCGAGGGGTCGGCGCGGTCTCGCAGGACCTCGACAAGCACGGGGCGGTCCAC
GTCGAGCAACACCGCCGCGAACAACGCGGACTGCGCGTGGCTGCAGGCCAGGAGGAAGAGGAAGAGGTC
GGCTTCCCGGTCAAGCCGCAAGTGCCGCTCAGGCCGATGACGTACAAGGGGGCCTTCGACCTCTCGTTCT
TCCTGAAAGAGAAGGGTGGCCTGGACGGGCTCATCTACTCGAAGAAGAGGCAGGAGATCCTCGACCTGTG
GGTCTACCACACCCAGGGTTCTTCCCGGACTGGCACAACCTACACGCCAGGCCCGGGAACCCGCTACCCG
CTGTGCTTCCGGTGGTGCTTCAAGCTCGTCCCATGGACCCGCGGAGGTGAGGAAGCCACCGAGGGCG
AGAACAACCTCCCTCTGCACCCGATCTGCCAGCACGGGATGGAGGACGAGGACCGCGAGGTGCTGGTCTG
GCGGTTGCACTCGTCCCTGGCCCCGGCGCCACATCGCCGGGAGCTCCACCCGGAGTACTACAAGGACTGC
TGATAAGCTAGCGGATCCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCT
CCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCACTGTCTTTTCTAATAAAATGAGGAAATGTC
ATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGAT
TGGGAAGACAATAGCAGGCATGCTGGGGAATTT

THURSDAY, AUGUST 2, 2007

SEQUENCE 0 LENGTH : 3953

Mosaic and CON nef_Dmyr_final.doc

82/109

Fig. 22 cont'd-16

AATII	GACGTC			3685				1855	1973	1976
4 SITES				BSPHI	TCATGA			1994	2110	2250
2510	2563	2646		2 SITES				2279		
2832				1007	1105			2282	3129	3166
ACCI	GTMKAC			BSPNI	CCWGG			3235	3261	3306
2 SITES				13 SITES				3315		
3050	3433			52	1738	1751		FNUDII	CGCG	
APLIII	ACRYGT			1872	2475	2668		17 SITES		
1 SITE				2977				494	1273	1854
1897				3270	3381	3444		2169	2257	2281
ALUI	AGCT			3478	3657	3801		2445		
13 SITES				BSSHII	GCGCGC			3039	3045	3047
109	633	1340		1 SITE				3062	3165	3183
1597	1643	1733		3045				3237		
1959				BSTNI	CCWGG			3246	3255	3625
2184	2951	3112		13 SITES				HAEII	RGCGCY	
3524	3683	3717		52	1738	1751		4 SITES		
ALWNI	CAGNNNCTG			1872	2475	2668		12	1657	2027
2 SITES				2977				3668		
1488	2129			3270	3381	3444		HAEIII	GGCC	
AOSII	GRCGYC			3478	3657	3801		21 SITES		
7 SITES				CFR10I	RCCGGY			55	771	1175
2507	2560	2643		1 SITE				1423	1857	1875
2829	2983	3156		847				1886		
3665				CFRI	YGGCCR			2268	2469	2662
APAI	GGGCCC			4 SITES				3073	3093	3128
1 SITE				769	3126	3166		3168		
3075				3541				3267	3323	3342
APALI	GTGCAC			CLAI	ATCGAT			3379	3481	3543
2 SITES				1 SITE				3660		
1583	2081			2287				HGIAI	GWGCWC	
AVAI	CYCGRG			DDEI	CTNAG			5 SITES		
3 SITES				12 SITES				1587	2085	2953
3078	3482	3676		12	204	397		3114	3685	
BAMHI	GGATCC			711	787	1214	1623	HEAI	GCGC	
1 SITE				2088	2158	2229		17 SITES		
3723				3318	3861			11	496	1273
BANI	GGYRCC			DPNI	GATC			1382	1556	1656
4 SITES				13 SITES				1723		
538	2850	3664		190	195	460		1993	2026	2169
3807				1239	1247	1258		2249	3045	3047
BANII	GRGCYC			1333				3145		
5 SITES				2972	3028	3417		3183	3255	3667
2953	3075	3114		3594	3725	3731		HINCII	GTYRAC	
3391	3685			DRAIII	CACNNNGTG			4 SITES		
BBEI	GGCGCC			1 SITE				413	886	2369
1 SITE				1161				3051		
3668				ECO47I	GGWCC			HINFI	GANTC	
BCLI	TGATCA			8 SITES				14 SITES		
1 SITE				122	586	919		43	59	357
3729				1048	3021	3193		383	401	725
BCNI	CCSGG			3535				807	1527	1923
13 SITES				3619				1998	2222	2795
1173	1521	3021		ECOO109	RGGNCCY			3648		
3139	3298	3457		2 SITES				HINPI	GCGC	
3484				3194	3340			17 SITES		
3485	3541	3663		ECORII	CCWGG			9	494	1271
3678	3679	3691		13 SITES				1380	1554	1654
BGLI	GCCNNNNNGGC			50	1736	1749		1721		
4 SITES				1870	2473	2666		1991	2024	2167
2475	2597	2668		2975				2247	3043	3045
3320				3268	3379	3442		3143		
BGLII	AGATCT			3476	3655	3799		3181	3253	3665
1 SITE				ECORV	GATATC			HPAII	CCGG	
458				1 SITE				16 SITES		
BSP1286	GDGCHC			2294				848	1172	1329
8 SITES				FNU4HI	GCGNC			1519	1545	1692
652	1587	2085		21 SITES				3019		
2953	3075	3114		234	769	1283		3137	3149	3297
3391				1489	1492	1557		3456	3483	3540
				1700				3662		

Mosaic and CON nef_Dmyr_final.doc

83/109

Fig. 22 cont'd-17

3677 3690
 MAEI CTAG
 7 SITES
 378 801 1034
 1404 2385 3718
 3751
 MAEII ACGT
 12 SITES
 669 1160 1196
 2306 2507 2519
 2560
 2643 2724 2829
 3219 3330
 MAEIII GTNAC
 9 SITES
 270 1134 1361
 1477 1540 2446
 2533
 2882 3215
 MVAI CCNGG
 26 SITES
 52 1172 1520
 1738 1751 1872
 2475
 2668 2977 3020
 3138 3270 3297
 3381
 3444 3456 3478
 3483 3484 3540
 3657
 3662 3677 3678
 3690 3801
 NARI GGCGCC
 1 SITE
 3665
 NCII CCSGG
 13 SITES
 1172 1520 3020
 3138 3297 3456
 3483
 3484 3540 3662
 3677 3678 3690
 NCOI CCGATG
 2 SITES
 2745 3532
 NDEI CATATG
 2 SITES
 2076 2619
 NHEI GCTAGC
 1 SITE
 3717
 NLAI III CATG
 15 SITES
 538 762 864
 892 1011 1109 1181
 1901 2219 2349
 2367 2689 2749
 3536
 3942
 NLAI IV GGNNCC
 13 SITES
 92 540 1830
 1869 2852 3023
 3073
 3341 3488 3538
 3666 3725 3809
 NRUI TCGCGA
 1 SITE
 2257
 NSII ATGCAT

1 SITE
 796
 NSP7524I RCATGY
 2 SITES
 1901 3942
 NSPBII CMGCKG
 6 SITES
 1314 1559 2281
 3039 3165 3500
 PPUMI RGGWCCY
 1 SITE
 3194
 PSSI RGGNCCY
 2 SITES
 3197 3343
 PSTI CTGCAG
 1 SITE
 3265
 RSAI GTAC
 11 SITES
 559 2093 2263
 2330 2604 2684
 2717
 2768 2925 3333
 3696
 SACI GAGCTC
 3 SITES
 2953 3114 3685
 SACII CCGCGG
 3 SITES
 2282 3040 3166
 SALI GTCGAC
 1 SITE
 3049
 SAU3A GATC
 13 SITES
 188 193 458
 1237 1245 1256
 1331
 2970 3026 3415
 3592 3723 3729
 SAU96A GGNCC
 18 SITES
 123 587 920
 1049 1174 2266
 2468
 2661 3022 3071
 3072 3194 3266
 3340
 3480 3536 3620
 3659
 SCAI AGTACT
 1 SITE
 3696
 SCRPI CCNGG
 26 SITES
 52 1172 1520
 1738 1751 1872
 2475
 2668 2977 3020
 3138 3270 3297
 3381
 3444 3456 3478
 3483 3484 3540
 3657
 3662 3677 3678
 3690 3801
 SDNI GDGCHC
 8 SITES

652 1587 2085
 2953 3075 3114
 3391
 3685
 SINI GGWCC
 8 SITES
 123 587 920
 1049 3022 3194
 3536
 3620
 SMAI CCCGGG
 2 SITES
 3484 3678
 SNABI TACGTA
 1 SITE
 2725
 SPEI ACTAGT
 1 SITE
 2384
 SPHI GCATGC
 1 SITE
 3942
 SSPI AATATT
 2 SITES
 603 991
 STUI AGGCCT
 1 SITE
 55
 STYI CCWGGG
 2 SITES
 2745 3532
 TAQI TCGA
 14 SITES
 216 1799 2287
 3050 3079 3105
 3199
 3222 3346 3399
 3421 3550 3646
 3738
 KHOI CTCGAG
 1 SITE
 3078
 KHOII RGATCY
 5 SITES
 458 1245 1256
 3415 3723
 XMAI CCCGGG
 2 SITES
 3482 3676
 XMAIII CGGCCG
 2 SITES
 3166 3541
 XMNI GAANNNTTC
 1 SITE
 811
 FOLLOWING ENZYMES HAVE NO
 SITES
 ACCIII AFLII ASP718
 ASUII AVRII BALI
 BSPMII BSTEII BSTXI
 DRAI ECO47III ECORI
 ESPI FSPI HINDIII
 HPAI KPNI MLUI
 MSTI NAEI NOTI
 OXANI PFLMI PVUI
 PVUII RSRII SFII
 SPLI TTH111I XBAI
 XCAI

Gag gene constructs:

HVI number	Gene name	Gag	Myristylation signal mutated
		Group M	
HV13234	M.con_Gag01_Dmyr.wlv	(2001)	Yes
HV13309	Gag-M4.1 Dmyr.wlv	Mosaic No. 1	Yes
HV13316	Gag_M4.2 Dmyr.wlv	Mosaic No. 2	Yes
HV13317	Gag_M4.3 Dmyr.wlv	Mosaic No. 3	Yes
HV13318	Gag_M4.4 Dmyr.wlv	Mosaic No. 4	Yes

All five constructs were cloned into HV10001 (WLV001AM DNA vaccine plasmid).

M.con_Gag01_Dmyr

MAARASVLSGGKLDWEKIRLRPGGKKKYRLKHLVWASRELERFALNPGLLETSEGCQQIIGQLQPALQT
GSEELRSLYNTVATLYCVHQRIEVKDTKEALEKIEEEQNKSQOKTQQAADKGNSSKVSQNYPIVQNLQG
QMVHQAI SPRTLNAWVKVIEEKAFSPEVIPMFSALSEGATPQDLNLTMLNTVGGHQAMQMLKDTINEEAA
EWDRLHPVHAGPIPPGQMREPRGSDIAGTTSTLQEQIAWMTSNPPIPVGEIYKRWIILGLNKIVRMYSPV
SILDIRQGPKEPFRDYVDRFFKTLRAEQATQDVKNWMTDTLLVQANPDCKTILKALGPGATLEEMMTAC
QGVGGPSHKARVLAEAMSQVTNAAIMMQRGNFKGQRRRIKCFNCGKEGHIARNCRAPRKKGCWKCGKEGH
QMKDCTERQANFLGKIWPSNKGPRGNFLQSRPEPTAPPAESFGFGEEITPSPKQEPKDKPEPLASLKSIF
GNDPLSQ_

HV13234 (M.con_Gag01_Dmyr.wlv) (1527nt)

XhoI

CTCGAGAAGAAAATGGCGGCTCGCGCCTCGGTCCTCAGCGGGGCAAGTTGGATGCGTGGGAGAA
GATCCGCTTGAGGCCAGGAGGAAGAAGAAGTACCGGCTCAAGCACCTGGTCTGGGCGAGCAGG
GAGCTGGAGCGCTTCGCGCTGAACCCGGGCCTGCTGGAGACATCCGAGGGCTGTGACGATCA
TCGGGCAGCTTCAGCCAGCGCTCCAGACGGGCAGCGAGGAGCTGCGCTCGCTATACAACACGGT
AGCGACCCCTCTACTGCGTGCAACGAGCGGATCGAGGTCAAGGACACGAAGGAGGCTCTTGAGAAG
ATCGAGGAGGAGCAGAACAAGTCGCAGCAGAAGACCCAGCAGGCGGCGGCCGACAAGGGCAACT
CCTCAAAGTATCTCAGAACTACCCGATCGTGCGAGAACCTGCAGGGACAGATGGTCCACCAGGC
CATCAGCCACCGGACGCTTAACGCCTGGGTCAAGGTGATCGAGGAGAAGGCCTTCTCGCCGGAG
GTCATCCCCATGTTCTCGGCACTCTCCGAGGGAGCCACCCCGCAGGACCTGAACACGATGTTGA
ACACGGTCGGCGGGCACCAGGCGGCCATGCAGATGCTCAAGGATACCATCAACGAGGAGGCTGC
GGAGTGGGACCGCCTGCACCCAGTGACGCGGGGCCCATCCCCCGGGCCAGATGAGAGAGCCG
CGGGGATCGGACATCGCGGCACGACCAGCACCTTGAGGAGCAAATCGCCTGGATGACTAGCA
ACCCCCGATCCCGGTGGGGGAGATCTACAAGCGGTGGATCATCCTCGGGTTGAACAAGATCGT
GCGGATGTACAGCCCTGTCTCAATCCTGGACATCCGGCAGGGGCCCAAGGAGCCCTTCCGCGAC
TACGTCGACCGGTTCTTCAAGACTCTCCGGGCGGAGCAGGCGACGCAGGACGTCAAGAACTGGA
TGACGGACACCTTGTTGGTCCAGAACGCTAACCCGGAAGTCAAGACGATCCTGAAGGCTCTCGG
CCCCGGGAGCGACCTTGAGGAGATGATGACCGGTGCCAGGGGGTCGGGGGACCCAGCCACAAG
GCGCGGGTCTTGGCCGAGGCGATGTCCAGGTACGAACGCCGCGATCATGATGCAGCGGGGGA
ACTTCAAGGGCCAGCGCCGGATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCCG
GAACTGCCGGGGCCCCGCGGAAGAAGGGCTGCTGGAAGTGCGGAAAGGAGGGGCACCAGATGAAG
GACTGCACGGAGCGCCAGGCGAATTTCTCGGGAAGATCTGGCCGTCCAACAAGGGGCGGCCAG
GGAACCTTCTGCAATCGCGGCCAGAGCCGACGGCCCCCTCCCGCGGAGAGCTTCGGGTTTCGGCGA

85/109

Fig. 22 cont'd-19

GGAGATCACGCCCAGCCCCAAGCAAGAGCCGAAGGACAAGGAGCCGCCCTTGCGTCCCTCAAG
TCGCTCTTCGGCAACGACCCGCTCTCCCAATGATAAGCTAGCGGATCCTGATCA

NheI BamHI BclI

Cut with XhoI and NheI site for VSV subcloning.

>HV13234 in hv10001 (4,822bp)

AAATGGGGGCGCTGAGGTCTGCCTCGTGAAGAAGGTGTTGCTGACTCATACCAGGCCTGAATCGCCCCAT
CATCCAGCCAGAAAGTGAGGGAGCCACGGTTGATGAGAGCTTTGTTGTAGGTGGACCAAGTTGGTGATTTT
GAACTTTTGCTTTGCCACGGAACGGTCTGCGTTGTGCGGAAGATGCGTGATCTGATCCTTCAACTCAGCA
AAAGTTCGATTTATTCAACAAAGCCGCCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAACCA
ATTAACCAATTCTGCGTTCAAATGGTATGCGTTTTGACACATCCACTATATATCCGTGTGCTTCTGTCC
ACTCCTGAATCCCATTCAGAAATTCTCTAGCGATTCCAGAAGTTTCTCAGAGTCGGAAAGTTGACCAGA
CATTACGAAGTGGCACAGATGGTCATAACCTGAAGGAAGATCTGATTGCTTAACTGCTTCAGTTAAGACC
GACGCGCTCGTCGTATAACAGATGCGATGATGCAGACCAATCAACATGGCACCTGCCATTGCTACCTGTA
CAGTCAAGGATGGTAGAAATGTTGTCGGTCTTGCAACGAATATTACGCCATTTGCTGCATATTCAA
CAGCTCTTCTACGATAAGGGCACAAATCGCATCGTGGAACGTTTGGGCTTCTACCGATTTAGCAGTTGA
TACACTTTCTCTAAGTATCCACCTGAATCATAAATCGGCAAAATAGAGAAAAATTGACCATGTGTAAGCG
GCCAATCTGATTCCACCTGAGATGCATAATCTAGTAGAATCTCTTCGCTATCAAAATTCACCTCCACCTT
CCACTCACCGGTTGTCCATTATGGCTGAACTCTGCTTCCTCTGTTGACATGACACACATCATCTCAATA
TCCGAATACGGACCATCAGTCTGACGACCAAGAGAGCCATAAACACCAATAGCCTTAACATCATCCCCAT
ATTTATCCAATATTCGTTTCCTTAATTTTCATGAACAATCTTCATTCTTCTTCTCTAGTCATTATTATTGG
TCCGTTTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTATTGTTTCATGATGATATATTT
TTATCTTGTGCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTCCCGGCCCATGACCAAAATC
CCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATC
CTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAACACCGCTACCAGCGGTGGTTTGTGTTGCC
GGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACCTGGCTTCAGCAGAGCGCAGATACCAATACTGTT
CTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGC
TAATCCTGTTACCACTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATA
GTTACCGGATAAGGCGCAGCGGTGCGGCTGAACGGGGGGTTCGTGCACACAGCCAGCTTGGAGCGAACG
ACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGG
CGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGC
CTGGTATCTTTATAGTCTGTGCGGTTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCA
GGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTT
TTGCTCACATGTTCTTTCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGC
TGATAACGCTCGCCGAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTG
ATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTACACCGCATATGGTGCACCTCTCAGTACAATCT
GCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGC
GAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGG
CGTTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATAGCCGCGGCATCGATGATATCCATTGCATACGTT
GTATCTATATCATAATATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGTTGACATTGATTA
TTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGTTA
CATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCGCCCATTGACGTCAATAATGAC
GTATGTTCCCATAGTAACGCCAATAGGGACTTTCATTGACGTCAATGGGTGGAGTATTTACGGTAAACT
GCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAAT
GGCCCGCCTGGCATTATGCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATT
AGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGATAGCGGTTTGACTCA
CGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTGTTGGCACCAAAATCAACGGGACT
TTCCAAAATGTGCTAACAACCTCCGCCCATTGACGCAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTA
TATAAGCAGAGCTCGTTTATGTAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCAT
AGAAGACACCGGGACCGATCCAGCCTCCGCGGGCGCGCTCGACGGGCGACGCGAAACTTGGGCCCACTC
GAGAAGAAAATGGCGGCTCGCGCCTCGGTCTCAGCGGGGGCAAGTTGGATGCGTGGGAGAAGATCCGCT
TGAGGGCCAGGAGGGAAGAAGAAGTACCGGCTCAAGCACCTGGTCTGGGCGAGCAGGGAGCTGGAGCGCTT
CGCGCTGAACCCGGGCTGCTGGAGACATCCGAGGGCTCTCAGCAGATCATCGGGCAGCTTCAGCCAGCG
CTCCAGACGGGCAGCGAGGAGCTGCGCTCGCTATACAACACGGTAGCGACCTCTACTGCGTGACCAGC

86/109

Fig. 22 cont'd-20

GGATCGAGGTCAAGGACACGAAGGAGGCTCTTGAGAAGATCGAGGAGGAGCAGAACAAGTCGCAGCAGAA
 GACCCAGCAGGCGGCGGCCGACAAGGGCAACTCCTCAAAAGTATCTCAGAACTACCCGATCGTGCAGAAC
 CTGCAGGGACAGATGGTCCACCAGGCCATCAGCCACGGACGCTTAACGCCTGGGTCAAGGTGATCGAGG
 AGAAGGCCTTCTCGCCGGAGGTCATCCCCATGTTCTCGGCACTCTCCGAGGGAGCCACCCCGCAGGACCT
 GAACACGATGTTGAACACGGTCGGCGGGCACCAGGCGGCCATGCAGATGCTCAAGGATACCATCAACGAG
 GAGGCTGCGGAGTGGGACCGCCTGCACCCAGTGCACGCGGGGCCCATCCCCCGGGCCAGATGAGAGAGC
 CGCGGGGATCGGACATCGCGGGCACGACCAGCACCTTGACAGGAGCAAATCGCCTGGATGACTAGCAACCC
 CCGATCCCCGGTCGGGGAGATCTACAAGCGGTGGATCATCCTCGGGTTGAACAAGATCGTGCGGATGTAC
 AGCCCTGTCTCAATCCTGGACATCCGGCAGGGGCCCAAGGAGCCCTTCCGCGACTACGTGACCGGTTCT
 TCAAGACTCTCCGGGCGGAGCAGGCGACGACGACGTCGAAGAACTGGATGACGGACACCTTGTTGGTCCA
 GAACGCTAACCCGACTGCAAGACGATCCTGAAGGCTCTCGGCCCGGGAGCGACCTTGGAGGAGATGATG
 ACCGCGTGCCAGGGGGTCGGGGGACCCAGCCACAAGGCGCGGGTCTTGGCCGAGGCGATGTCCAGGTCA
 CGAACGCCGCGATCATGATGCAGCGGGGAACTTCAAGGGCCAGCGCCGGATCATCAAGTGCTTCAACTG
 CGGCAAGGAGGGCCACATCGCCCCGGAAGTGCCTGGGCCCCCGCGGAAGAAGGGCTGCTGGAAGTGCGGAAAG
 GAGGGGCACCAGATGAAGGACTGCACGGAGCGCCAGGCGAATTTCTCGGGAAGATCTGGCCGTCCAACA
 AGGGGCGGCCAGGGAATTCCTGCAATCGCGGCCAGAGCCGACGCGCCCTCCGCGGAGAGCTTCGGGTT
 CGGCGAGGAGATCACGCCAGCCCCAAGCAAGAGCCGAAGGACAAGGAGCCGCCCTTGCGTCCCTCAAG
 TCGCTCTTCGGCAACGACCCGCTCTCCCAATGATAAGCTAGCGGATCCTGATCAGCCTCGACTGTGCCTT
 CTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCAC
 TGTCTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGGT
 GGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGAATTT

Thursday, August 2, 2007

Sequence 0 Length : 4823

AatII GACGTC
 5 Sites
 2510 2563 2646
 2832 4027
 AccI GTMKAC
 2 Sites
 3050 3979
 AflIII ACRYGT
 1 Site
 1897
 AluI AGCT
 14 Sites
 109 633 1340
 1597 1643 1733
 1959
 2184 2951 3209
 3278 3311 4471
 4587
 AlwNI CAGNNNCTG
 4 Sites
 1488 2129 3500
 3638
 AseII GRCGYC
 6 Sites
 2507 2560 2643
 2829 2983 4024
 ApaI GGGCCC
 4 Sites
 3075 3754 3955
 4307
 ApaLI GTGCAC
 4 Sites
 1583 2081 3351
 3741
 AvaI CYCGRG
 6 Sites
 3078 3230 3761
 3891 4103 4386
 BamHI GGATCC
 1 Site
 4593

BanI GGYRCC
 5 Sites
 538 2850 3667
 4345 4677
 BanII GRGCTC
 6 Sites
 2953 3075 3754
 3955 3964 4307
 BclI TGATCA
 1 Site
 4599
 BcnI CCSSG
 14 Sites
 1173 1521 3021
 3232 3233 3763
 3764
 3859 4003 4072
 4105 4106 4293
 4303
 BglI GCCNNNNNGGC
 3 Sites
 2475 2597 2668
 BglII AGATCT
 3 Sites
 458 3868 4393
 Bsp1286 GDGCHC
 14 Sites
 652 1587 2085
 2953 3075 3355
 3670
 3745 3754 3804
 3955 3964 4307
 4348
 BspHI TCATGA
 3 Sites
 1007 1105 4213
 BspNI CCWGG
 19 Sites
 52 1738 1751
 1872 2475 2668
 2977
 3157 3189 3522
 3551 3672 3833
 3936

4140 4194 4374
 4420 4671
 BssHII GCGCGC
 1 Site
 3045
 BstNI CCWGG
 19 Sites
 52 1738 1751
 1872 2475 2668
 2977
 3157 3189 3522
 3551 3672 3833
 3936
 4140 4194 4374
 4420 4671
 Cfr10I RCCGGY
 3 Sites
 847 3175 3982
 CfrI YGGCCR
 7 Sites
 769 3445 3676
 4177 4398 4416
 4440
 ClaI ATCGAT
 1 Site
 2287
 DdeI CTNAG
 13 Sites
 12 204 397
 711 787 1214 1623
 2088 2158 2229
 3111 3475 4731
 DpnI GATC
 27 Sites
 190 195 460
 1239 1247 1258
 1333
 2972 3028 3144
 3267 3363 3399
 3489
 3564 3788 3855
 3870 3885 3906
 4086
 4212 4251 4395
 4491 4595 4601

Mosaic and Group M Gag_dmyr-patent.doc

87/109

Fig. 22 cont'd-21

DraIII	CACNNNGTG	4179	4240	4282	52	1172	1520
2 Sites		4305	4400	4418	1738	1751	1872
1161	3740	4442			2475		
Eco47I	GGWCC	4455			2668	2977	3020
11 Sites		HgiAI	GWGCWC		3157	3189	3231
122	586	5 Sites			3232		
1048	3021	1587	2085	2953	3522	3551	3672
3514		3355	3745		3762	3763	3833
3634	3724	HhaI	GCGC		3858		
4151		21 Sites			3936	4002	4071
Eco47III	AGCGCT	11	496	1273	4104	4105	4140
2 Sites		1382	1556	1656	4194		
3216	3289	1723			4292	4302	4374
EcoO109	RGGNCCY	1993	2026	2169	4420	4671	
5 Sites		2249	3045	3047	NciI	CCSGG	
3635	3750	3102			14 Sites		
4152	4304	3217	3224	3290	1172	1520	3020
EcoRII	CCWGG	3316	4169	4246	3231	3232	3762
19 Sites		4372			3763		
50	1736	HincII	GTyrAC		3858	4002	4071
1870	2473	5 Sites			4104	4105	4292
2975		413	886	2369	4302		
3155	3187	3051	3980		NcoI	CCATGG	
3549	3670	HinfI	GANTC		1 Site		
3934		14 Sites			2745		
4138	4192	43	59	357	NdeI	CATATG	
4418	4669	383	401	725	2 Sites		
EcoRV	GATATC	807	1527	1923	2076	2619	
1 Site		1998	2222	2795	NheI	GCTAGC	
2294		3995			1 Site		
Fnu4HI	GCNGC	HinPI	GCGC		4587		
32 Sites		21 Sites			NlaIII	CATG	
234	769	9	494	1271	17 Sites		
1489	1492	1380	1554	1654	538	762	864
1700		1721			892	1011	1109
1855	1973	1991	2024	2167	1901	2219	2349
1994	2110	2247	3043	3045	2367	2689	2749
2279		3100			3602		
2282	3094	3215	3222	3288	3683	4217	4812
3302	3312	3314	4167	4244	NlaIV	GGNNCC	
3442		4370			24 Sites		
3445	3676	BpaII	CCGG		92	540	1830
3780	4207	20 Sites			1869	2852	3023
4271		848	1172	1329	3073		
4322	4416	1519	1545	1692	3623	3669	3726
4530		3019			3751	3752	3952
FnuDII	CGCG	3176	3231	3585	3953		
23 Sites		3762	3858	3944	3961	4153	4154
494	1273	3983			4305	4306	4347
2169	2257	4001	4071	4104	4456		
2445		4247	4292	4301	4528	4595	4679
3039	3045	MaeI	CTAG		NruI	TCGCGA	
3062	3100	8 Sites			1 Site		
3747		378	801	1034	2257		
3782	3798	1404	2385	3841	NsiI	ATGCAT	
4134	4169	4588			1 Site		
4310		4621			796		
4439	4464	MaeII	ACGT		Nsp7524I	RCATGY	
HaeII	RGCGCY	12 Sites			2 Sites		
7 Sites		669	1160	1196	1901	4812	
12	1657	2306	2507	2519	NspBII	CMGCKG	
3218	3291	2560			10 Sites		
4373		2643	2724	2829	1314	1559	2281
HaeIII	GGCC	3976	4024		3039	3115	3359
29 Sites		MaeIII	GTNAC		3782		
55	771	9 Sites			4223	4310	4464
1423	1857	270	1134	1361	PpuMI	RGGWCCY	
1886		1477	1540	2446	2 Sites		
2268	2469	2533			3635	4152	
3073	3155	2882	4196		PsaI	RGGNCCY	
3447		MvaI	CCNGG		5 Sites		
3525	3576	33 Sites			3638	3753	3954
3752	3766				4155	4307	
4102					PstI	CTGCAG	
Mosaic and Group M Gag_dmyr-patent.doc							

88/109

Fig. 22 cont'd-22

1 Site		2668	2977	3020		3975		
3505		3157	3189	3231		XhoI	CTCGAG	
PvuI	CGATCG	3232				1 Site		
1 Site		3522	3551	3672		3078		
3490		3762	3763	3833		XhoII	RGATCY	
RsaI	GTAC	3858				7 Sites		
11 Sites		3936	4002	4071		458	1245	1256
559	2093	4104	4105	4140		3142	3868	4393
2330	2604	4194				4593		
2717		4292	4302	4374		XmaI	CCCGGG	
2768	2925	4420	4671			3 Sites		
3918		SdnI	GDGCHC			3230	3761	4103
SacI	GAGCTC	14 Sites				XmaIII	CGGCCG	
1 Site		652	1587	2085		1 Site		
2953		2953	3075	3355		3445		
SacII	CCGCCG	3670				XmnI	GAANNNTTC	
5 Sites		3745	3754	3804		2 Sites		
2282	3040	3955	3964	4307		811	3576	
4311	4465	4348						
SalI	GTCGAC	SinI	GGWCC			Following enzymes have no		
2 Sites		11 Sites				sites		
3049	3978	123	587	920		AccIII	AflII	Asp718
Sau3A	GATC	1049	3022	3107		AsuII	AvrII	BalI
27 Sites		3515				BbeI	BspMII	BstEII
188	193	3635	3725	4055		BstXI	DraI	EcoRI
1237	1245	4152				EspI	FspI	HindIII
1331		SmaI	CCCGGG			HpaI	KpnI	MluI
2970	3026	3 Sites				MstI	NaeI	NarI
3265	3361	3232	3763	4105		NotI	OxaNI	PflMI
3487		SnaBI	TACGTA			PvuII	RsrII	ScaI
3562	3786	1 Site				SfiI	SplI	XbaI
3868	3883	2725				XcaI		
4084		SpeI	ACTAGT					
4210	4249	1 Site				1 Site		
4489	4593	2384				2953		
Sau96A	GGNCC	SphI	GCATGC			SnaBI	TACGTA	
29 Sites		1 Site				1 Site		
123	587	4812				2725		
1049	1174	SspI	AATATT			SpeI	ACTAGT	
2468		2 Sites				1 Site		
2661	3022	603	991			2384		
3072	3107	StuI	AGGCCT			SphI	GCATGC	
3515		2 Sites				1 Site		
3635	3725	55	3576			4812		
3751	3764	StyI	CCWWGG			Tth111I	GACNNNGTC	
3952		3 Sites				1 Site		
4055	4101	2745	3955	4114		3975		
4238	4280	TaqI	TCGA			XhoI	CTCGAG	
4304		10 Sites				1 Site		
4454		216	1799	2287		3078		
ScrFI	CCNGG	3050	3079	3364		XmaIII	CGGCCG	
33 Sites		3400				1 Site		
52	1172	3565	3979	4608		3445		
1738	1751	Tth111I	GACNNNGTC					
2475	1872	1 Site						

Gag-M4.1dmyr

MAARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQILGQL
 QPSLQTGSEELRSLYNTVAVLYCVHQRIDVKDTKEALEKIEEEQNKSSQKQTQQAKAADGKVSQN
 YPIVQNAQGQMVBHQAI SPRTLNAWVKVIEEKGFSPFVPMFSALEGATPQDLNMLNTIGGHQ
 AAMQMLKDTINDEAAEWDRLHPVQAGPVAPGQIREPRGSDIAGTTSNLQEQIGWMTSNPPIPVG
 DIYKRWIIMGLNKKIVRMYSPTSILDIRQGPKESEFRDYVDRFFRTLRAEQASQEVKNWMTETLLV
 QNSNPDKTILKALGPAATLEEMMTACQGVGGPGHKARVLAEAMSQVQPNIMMQRGNFKGQKR
 IKCFNCGREGHLARNCRAPRKKGCWKCGRGHQMKDCTESKANFLGKIWPSNKGPRGNFLQSRP
 EPSAPPAESFGFGEEITPSQKQEQKDKELYPLASLSLFGNDPLSQ

89/109

Fig. 22 cont'd-23

HV13309 (Gag-M4.1Dmyr.wlv)

CTCGACAAGAAAATGCGGGCGCGCGCCTCGGTCTTAGCGGGGGCGAGTTGGATCGGTG
 GGAAAAGATCCGCTTGAGGCCAGGAGGGAAGAAGAAGTACAAGCTCAAGCACATCGTCT
 GGGCGAGCAGAGAGCTCGAGCGGTTGCGGGTCAACCCGGGCCTGCTTGAGACATCGGAG
 GGCTGTCGGCAAATCCTGGGGCAGCTTCAACCGTCCTTGCAAACGGGCAGCGAGGAGCT
 GCGATCGCTATACAACACGGTAGCGGTCTCTACTGCGTGCACCAGCGGATCGACGTCA
 AGGACACGAAGGAGGCTCTTGAGAAGATTGAGGAAGAGCAGAACAAGTCGCAGCAGAAG
 ACGCAGCAGGCGAAGGCCGCGAGACGGCAAAGTATCTCAGAACTACCCGATCGTGCAGAA
 CGCGCAGGGACAGATGGTCCACCAGGCCATCTCCCCACGGACGCTTAACGCCTGGGTCA
 AAGTAATCGAGGAGAAGGGATTGAGCCCGGAAGTGATCCCCATGTTCTCGGCACTTGCG
 GAGGGAGCCACCCCGCAGGACCTGAACACGATGTTGAACACCATCGGCGGGCACCAGGC
 GGCCATGCAGATGCTTAAGGACACCATCAACGACGAGGCTGCGGAGTGGGACCGGCTTC
 ACCCGGTGCAGGCGGGGGCCCGTTCGCGCCGGGCCAGATCAGAGAGCCGCGGGGATCGGAC
 ATCGCGGGAACCACCAGCAACTTGACAGGAGCAAATCGGGTGGATGACTTCGAACCCCCC
 GATCCCGGTTCGGGGACATCTACAAGAGATGGATCATCATGGGGTTGAACAAGATCGTGA
 GGATGTACAGCCCGACAGCATCCTGGACATCCGACAGGGACCGAAGGAGTCGTTTCAGA
 GACTACGTAGACCGGTTCTTCCGGACTCTCCGGGCGGAGCAGGCGTCGCAGGAGGTCAA
 GAACTGGATGACGGAGACCTTGTGGTCCAGAACTCGAACCCGGACTGCAAGACCATCC
 TGAAGGCTCTCGGCCCCGGCGGCGACGTTGGAAGAGATGATGACGGCGTGCCAGGGAGTC
 GGGGGACCCGGCCACAAGGCGCGGGTCTTGGCCGAGGCGATGAGCCAGGTGCAGCAGCC
 GAACATCATGATGCAGCGGGGCAACTTCAAGGGACAGAAGCGGATCAAGTGTTCAACT
 GTGGCAGGGAGGGACACCTCGCCAGGAACTGCCGGGCCCCGCGGAAGAAGGGCTGCTGG
 AAGTGCGGAAGGGAGGGGCACCAAATGAAGGACTGCACGGAGTCGAAGGCGAATTTCT
 CGGGAAGATCTGGCCGTCCAACAAGGGGCGGCCAGGGAACTTTCTGCAAAGCCGGCCGG
 AGCCGTTCGGCCCCGCGGCGGAGTCCTTTGGGTTTCGGGGAGGAGATCACGCCCTCGCAG
 AAGCAAGAGCAAAAGGACAAGGAGCTCTACCCTCTTGCGTCCCTCAAGTCGCTCTTCGGC
 AACGACCCGCTTTTCGCAATGATAAGCTAGCGGATCCTGATCAGGCGCGCC

NheI

AscI

HV13309 in HV10001, 4836bp

AAATGGGGGCGCTGAGGTCTGCCTCGTGAAGAAGGTGTTGCTGACTCATAACAGGCCTGAATCG
 CCCCATCATCCAGCCAGAAAGTGAGGGAGCCACGGTTGATGAGAGCTTTGTTGTAGGTGGACCA
 GTTGGTGATTTTGAACCTTTTGCTTTGCCACGGAACGGTCTGCGTTGTCGGGAAGATGCGTGATC
 TGATCCTTCAACTCAGCAAAAGTTCGATTTATTCAACAAAGCCCGCTCCCGTCAAGTCAGCGT
 AATGCTCTGCCAGTGTTACAACCAATTAACCAATTCTGCGTTCAAATGGTATGCGTTTTGACA
 CATCCACTATATATCCGTGTGCTTCTGTCCACTCCTGAATCCCATTCCAGAAATCTCTAGCGA
 TTCCAGAAGTTTTCTCAGAGTCGGAAAGTTGACCAGACATTACGAACTGGCACAGATGGTCATAA
 CCTGAAGGAAGATCTGATTGCTTAAGTCTTCAAGTAAAGACCGACGCGCTCGTCGTATAACAGA
 TGCGATGATGCAGACCAATCAACATGGCACCTGCCATTGCTACCTGTACAGTCAAGGATGGTAG
 AAATGTTGTGCGTCTTGCACACGAATATTACGCCATTTGCCTGCATATTCAAACAGCTCTTCT
 ACGATAAGGGCACAAATCGCATCGTGGAACGTTTGGGCTTCTACCGATTTAGCAGTTTGATACA
 CTTTCTCTAAGTATCCACCTGAATCATAAATCGGCAAAATAGAGAAAATTGACCATGTGTAAG
 CGGCCAATCTGATTCCACCTGAGATGCATAATCTAGTAGAATCTCTTCGCTATCAAAATTTCACT
 TCCACCTTCCACTCACCGGTTGTCCATTTCATGGCTGAACTCTGCTTCTCTGTTGACATGACAC
 ACATCATCTCAATATCCGAATACGGACCATCAGTCTGACGACCAAGAGAGCCATACACCAAT
 AGCCTTAACATCATCCCATATTTATCCAATATTCGTTCTTAAATTTTCATGAACAATCTTCATT
 CTTTCTTCTCTAGTCATTATTATGGTCCGTTTCATAACACCCCTTGTAATTACTGTTTATGTAAG

90/109

Fig. 22 cont'd-24

CAGACAGTTTTATTGTTTCATGATGATATATTTTTATCTTGTGCAATGTAACATCAGAGATTTTG
AGACACAACGTGGCTTTCCCCGGCCCATGACCAAAATCCCTTAACGTGAGTTTTCTGTTCCACTG
AGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATC
TGCTGCTTGCAAACAAAAAACCACCGCTACCAGCGGTGGTTTGTGGCCGGATCAAGAGCTAC
CAACTCTTTTTCCGAAGGTAACCTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGT
GTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTA
ATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGAC
GATAGTTACCGGATAAGGCGCAGCGGTCCGGCTGAACGGGGGGTTTCGTGCACACAGCCCAGCTT
GGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTT
CCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGA
GGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTGCGGTTTCGCCACCTCTGACT
TGAGCGTCGATTTTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCG
GCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCC
CTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGAGCCGAAC
GACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTT
ACGCATCTGTGCGGTATTTACACCGCATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCG
CATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAA
AATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGG
CGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATAGCCGCGGCATCGATGATATCCATTGCA
TACGTTGTATCTATATCATAATATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGT
TGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCAT
ATATGGAGTTCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCC
CCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGA
CGTCAATGGGTGGAGTATTTACGGTAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGC
CAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTTATGCCCAGTACAT
GACCTTACGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTG
ATGCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTTGACTCACGGGGATTTCCAAGTC
TCCACCCCATTTGACGTCAATGGGAGTTTGTGTTTGGCACCAAAATCAACGGGACTTTCCAAAATG
TCGTAACAACTCCGCCCCATTGACGCAATGGGCGGTAGGCGGTGACGGTGGGAGGTCTATATA
AGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCC
ATAGAAGACACCGGGACCGATCCAGCCTCCGCGGGCGCGCGTTCGACGGGCGACGCGAACTTGG
GCCCACTCGACAAGAAAATGCGCGCGCGCCTCGGTCTTAGCGGGGGCGAGTTGGAT
CGGTGGGAAAAGATCCGCTTGAGGCCAGGAGGGAAGAAGAAGTACAAGCTCAAGCACAT
CGTCTGGGCGAGCAGAGAGCTCGAGCGGTTCCGCGTCAACCCGGGCCTGCTTGAGACAT
CGGAGGGCTGTGCGCAAATCCTGGGGCAGCTTCAACCGTCTTGCAAACGGGCAGCGAG
GAGCTGCGATCGCTATACAACACGGTAGCGGTCCTCTACTGCGTGCACCAGCGGATCGA
CGTCAAGGACACGAAGGAGGCTCTTGAGAAGATTGAGGAAGAGCAGAACAAGTCGCAGC
AGAAGACGCAGCAGGCGAAGGCCGACAGCGGCAAAGTATCTCAGAACTACCCGATCGTG
CAGAACGCGCAGGGACAGATGGTCCACCAGGCCATCTCCCCACGGACGCTTAACGCCTG
GGTCAAAGTAATCGAGGAGAAGGGATTAGCCCGGAAGTGATCCCCATGTTCTCGGCAC
TTGCGGAGGGAGCCACCCCGCAGGACCTGAACACGATGTTGAACACCATCGGCGGGCAC
CAGGCGGCCATGCAGATGCTTAAGGACACCATCAACGACGAGGCTGCGGAGTGGGACCG
GCTTCACCCGGTGCAGGCGGGGCCGTCGCGCCGGGCCAGATCAGAGAGCCGCGGGGAT
CGGACATCGCGGGAACCAACAGCAACTTGCAGGAGCAAATCGGGTGGATGACTTCGAAC
CCCCGATCCCGGTGCGGGGACATCTACAAGAGATGGATCATCATGGGGTTGAACAAGAT
CGTGAGGATGTACAGCCCGACCGCATCCTGGACATCCGACAGGGACCGAAGGAGTCGT
TCAGAGACTACGTAGACCGGTTCTTCCGGACTCTCCGGGCGGAGCAGGCGTCGCAGGAG

91/109

Fig. 22 cont'd-25

GTCAAGAACTGGATGACGGAGACCTTGTGTTGGTCCAGAACTCGAACCCTGGACTGCAAGAC
 CATCCTGAAGGCTCTCGGCCCGGCGGCGACGTTGGAAGAGATGATGACGGCGTGCCAGG
 GAGTCGGGGGACCCGGCCACAAGGCGCGGGTCTTGCCGAGGCGATGAGCCAGGTGCAG
 CAGCCGAACATCATGATGCAGCGGGGCAACTTCAAGGGACAGAAGCGGATCAAGTGTTT
 CAACTGTGGCAGGGAGGGACACCTCGCCAGGAACTGCCGGGCCCCGCGGAAGAAGGGCT
 GCTGGAAGTGCGGAAGGGAGGGGCACCAAATGAAGGACTGCACGGAGTCGAAGGCGAAT
 TTCCTCGGGAAGATCTGGCCGTCCAACAAGGGGCGGCCAGGGAACCTTTCTGCAAAGCCG
 GCCGGAGCCGTCGGCCCCGCGGCGGAGTCCTTTGGGTTCGGGGAGGAGATCACGCCCT
 CGCAGAAGCAAGAGCAAAAGGACAAGGAGCTCTACCCTCTTGCGTCCCTCAAGTCGCTC
 TTCGGCAACGACCCGCTTTCGCAATGATAAGCTAGCGGATCCTGATCAGGCGCGCCGAGCTCGC
 TGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCCGTGCCTT
 CCTTGACCCTGGAAGGTGCCACTCCCCTGTCTTTTCTTAATAAAATGAGGAAATTGCATCGCA
 TTGTCTGAGTAGGTGTCACTTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGAT
 TGGGAAGACAATAGCAGGCATGCTGGGGAATTT

Thursday, August 2, 2007

Sequence 0 Length : 4836				1 Site 4584		4185	4280	4408
				BanI	GGYRCC	4684		
				5 Sites		BssHII	GCGCGC	
				538	2850 3658	4 Sites		
AatII	GACGTC			4333	4690	3045	3098	3100
5 Sites				BanII	GRGCTC	4599		
2510 2563 2646				7 Sites		BstNI	CCWGG	
2832 3370				2953	3075 3211	18 Sites		
AccI	GTMKAC			3745	4295 4519	52	1738	1751
2 Sites				4608		1872	2475	2668
3050 3970				BclI	TGATCA	2977		
AccIII	TCCGGA			2 Sites		3157	3270	3513
1 Site				4590	4612	3542	3663	3927
3982				BcnI	CCSGG	4131		
AflII	CTTAAG			14 Sites		4185	4280	4408
1 Site				1173	1521 3021	4684		
3681				3232	3233 3577	BstXI	CCANNNNNTGG	
AflIII	ACRYGT			3730		1 Site		
1 Site				3755	3850 3994	3926		
1897				4063	4096 4148	Cfr10I	RCCGGY	
AluI	AGCT			4291		5 Sites		
16 Sites				BglI	GCCNNNNNGGC	847	3718	3973
109 633 1340				4 Sites		4426	4448	
1597 1643 1733				2475	2597 2668	CfrI	YGGCCR	
1959				4449		7 Sites		
2184 2951 3179				BglII	AGATCT	769	3667	4148
3209 3278 3311				2 Sites		4168	4386	4404
4517				458	4381	4428		
4578 4606				Bsp1286	GDGCHC	Clal	ATCGAT	
AlwNI	CAGNNNCTG			13 Sites		1 Site		
4 Sites				652	1587 2085	2287		
1488 2129 3629				2953	3075 3211	DdeI	CTNAG	
4285				3355		13 Sites		
AosII	GRCGYC			3661	3745 4295	12	204	397
7 Sites				4336	4519 4608	711	787	1214 1623
2507 2560 2643				BspHI	TCATGA	2088	2158	2229
2829 2983 3367				3 Sites		3111	3466	4744
4005				1007	1105 4204	DpnI	GATC	
ApaI	GGGCCC			BspMII	TCCGGA	26 Sites		
3 Sites				1 Site		190	195	460
3075 3745 4295				3982		1239	1247	1258
ApaLI	GTGCAC			BspNI	CCWGG	1333		
3 Sites				18 Sites		2972	3028	3130
1583 2081 3351				52	1738 1751	3144	3317	3363
AsuII	TTCGAA			1872	2475 2668	3480		
1 Site				2977		3585	3762	3779
3834				3157	3270 3513	3846	3876	3897
AvaI	CYCGRG			3542	3663 3927	4242		
3 Sites				4131		4383	4479	4586
3210 3230 4374						4592	4614	
BamHI	GGATCC					DraIII	CACNNNCTG	

Mosaic and Group M Gag_dmyr-patent.doc

92/109

Fig. 22 cont'd-26

2 Sites
1161 3731
Eco47I GGWCC
13 Sites
122 586 919
1048 3021 3106
3337
3505 3625 3715
3941 4045 4142
EcoO109 RGGNCCY
4 Sites
3626 3741 4143
4292
EcoRII CCWGG
18 Sites
50 1736 1749
1870 2473 2666
2975
3155 3268 3511
3540 3661 3925
4129
4183 4278 4406
4682
EcoRV GATATC
1 Site
2294
Fnu4HI GCNGC
31 Sites
234 769 1283
1489 1492 1557
1700
1855 1973 1976
1994 2110 2250
2279
2282 3094 3276
3302 3312 3423
3435
3448 3667 3706
3771 4099 4191
4194
4212 4310 4404
FnuDII CGCG
21 Sites
494 1273 1854
2169 2257 2281
2445
3039 3045 3047
3062 3098 3100
3222
3492 3750 3773
3789 4160 4298
4599
HaeII RGCGCY
3 Sites
12 1657 2027
HaeIII GGCC
26 Sites
55 771 1175
1423 1857 1875
1886
2268 2469 2662
3073 3155 3235
3447
3516 3669 3743
3757 4093 4150
4170
4293 4388 4406
4430 4443
HgiAI GWGCWC
7 Sites
1587 2085 2953
3211 3355 4519
4608
HhaI GCGC
Mosaic and Group M Gag_dmyr-patent.doc

21 Sites
11 496 1273
1382 1556 1656
1723
1993 2026 2169
2249 3045 3047
3098
3100 3102 3494
3752 4160 4599
4601
HincII GTYRAC
5 Sites
413 886 2369
3051 3227
HinfI GATC
19 Sites
43 59 357
383 401 725 779
807 1527 1923
1998 2222 2795
3568
3951 3986 4135
4356 4455
HinfI GCGC
21 Sites
9 494 1271
1380 1554 1654
1721
1991 2024 2167
2247 3043 3045
3096
3098 3100 3492
3750 4158 4597
4599
HpaII CCGG
23 Sites
848 1172 1329
1519 1545 1692
3019
3231 3576 3719
3729 3753 3849
3974
3983 3992 4062
4095 4147 4289
4427
4431 4449
MaeI CTAG
7 Sites
378 801 1034
1404 2385 4579
4634
MaeII ACGT
13 Sites
669 1160 1196
2306 2507 2519
2560
2643 2724 2829
3367 3967 4104
MaeIII GTNAC
8 Sites
270 1134 1361
1477 1540 2446
2533
2882
MvaI CCNGG
32 Sites
52 1172 1520
1738 1751 1872
2475
2668 2977 3020
3157 3231 3232
3270

3513 3542 3576
3663 3729 3754
3849
3927 3993 4062
4095 4131 4147
4185
4280 4290 4408
4684
NaeI GCCGGC
2 Sites
4428 4450
NciI CCSGG
14 Sites
1172 1520 3020
3231 3232 3576
3729
3754 3849 3993
4062 4095 4147
4290
NcoI CCATGG
1 Site
2745
NdeI CATATG
2 Sites
2076 2619
NheI GCTAGC
1 Site
4578
NlaIII CATG
18 Sites
538 762 864
892 1011 1109 1181
1901 2219 2349
2367 2689 2749
3593
3674 3884 4208
4825
NlaIV GGNNCC
23 Sites
92 540 1830
1869 2852 3023
3073
3614 3660 3717
3742 3743 3794
3943
4144 4145 4293
4294 4335 4435
4444
4586 4692
NruI TCGCGA
1 Site
2257
NsiI ATGCAT
1 Site
796
Nsp7524I RCATGY
2 Sites
1901 4825
NspBII CMGCKG
8 Sites
1314 1559 2281
3039 3359 3773
4214
4298
PpuMI RGGWCCY
2 Sites
3626 4143
PseI RGGNCCY
4 Sites
3629 3744 4146
4295
PvuI CGATCG
2 Sites
3318 3481

93/109

Fig. 22 cont'd-27

RsaI	GTAC					ScrFI	CCNGG				603	991
11 Sites						32 Sites					StuI	AGGCCT
559	2093	2263				52	1172	1520			1 Site	
2330	2604	2684				1738	1751	1872			55	
2717						2475					StyI	CCWWGG
2768	2925	3174				2668	2977	3020			1 Site	
3909						3157	3231	3232			2745	
SacI	GAGCTC					3270					TaqI	TCGA
4 Sites						3513	3542	3576			12 Sites	
2953	3211	4519				3663	3729	3754			216	1799
4608						3849					3050	3211
SacII	CCGCGG					3927	3993	4062			3364	
4 Sites						4095	4131	4147			3556	3834
2282	3040	3774				4185					4359	4056
4299						4280	4290	4408			4621	
SalI	GTCGAC					4684					XhoI	CTCGAG
1 Site						SdnI	GDGCHC				1 Site	
3049						13 Sites					3210	
Sau3A	GATC					652	1587	2085			XhoII	RGATCY
26 Sites						2953	3075	3211			6 Sites	
188	193	458				3355					458	1245
1237	1245	1256				3661	3745	4295			3142	4381
1331						4336	4519	4608			XmaI	CCCCGG
2970	3026	3128				SinI	GGWCC				1 Site	
3142	3315	3361				13 Sites					3230	
3478						123	587	920			XmaIII	CGGCCG
3583	3760	3777				1049	3022	3107			1 Site	
3844	3874	3895				3338					4428	
4240						3506	3626	3716			XmnI	GAANNNTTC
4381	4477	4584				3942	4046	4143			2 Sites	
4590	4612					SmaI	CCCCGG				811	3567
Sau96A	GGNCC					1 Site					Following enzymes have no sites	
27 Sites						3232					Asp718	AvrII
123	587	920				SnaBI	TACGTA				Bali	
1049	1174	2266				2 Sites					BbeI	BstEII
2468						2725	3968				Eco47III	EcoRI
2661	3022	3071				SpeI	ACTAGT				FspI	HindIII
3072	3107	3233				1 Site					KpnI	MluI
3338						2384					NarI	NotI
3506	3626	3716				SphI	GCATGC				PflMI	PstI
3741	3742	3755				1 Site					RsrII	ScaI
3942						4825					SplI	Tth111I
4046	4092	4143				SspI	AATATT				XcaI	XbaI
4291	4292	4442				2 Sites						

Gag_M4.2 Dmyr

MAARASVLSGGKLDWEKIRLRPGGKKKYRLKHLVWASRELDRLFALNPSLLETAEGCKQIMKQL
 QPALKTGTEELKSLYNTVATLYCVHEKIDVRDTKEALDKIEEEQNKIQQKTQQAKEADGKVSQN
 YPIVQNIQGQMVHQPISPRTLNAWVKVVEEKAFSPEVIPMFTALSDGATPQDLNSMLNAVGGHQ
 AAMQILKDTINEEAADWDRLEHPVHAGPVAPGQMREPRGSDIAGTTSTLQEIQGWMTNPPPIPVG
 EIYKRWIIILGLNKIVRMYSFVSILDIKQGPKESEFRDYVDRFFKVLRAEQATQDVKNWMTDTLLI
 QNANPDCKSILRALGPGATLEEMMTACQGVGGPSHKARILAEAMSQVTNSATIMMQRGNFRNQR
 KTVKCFNCGKEGHLARNCKAPRKRGCWKCGKEGHQMKECTERQANFLGKIWPSSSKGRPGNFPQS
 RPEPTAPPEESFRFGEETTTTPSQKQEPIDKELYPLASLRSLFGNDPSSQAS

HV13316 (Gag_M4.2 Dmyr.wlv) cloned in to XhoI

GTCGAGAAGAAAATGGCGGCTCGCGCCTCGGTCCTTAGCGGGGGCAAGTTGGATGCGTGGGAAA
 AGATCCGCTTGAGGCCAGGAGGGAAGAAGAAGTACCGGCTCAAGCACCTTGTCTGGGCGAGCAG
 AGAGCTCGACCGGTTTCGCGCTGAACCCGTCGCTGCTTGAGACAGCCGAGGGCTGCAAGCAAATC
 ATGAAGCAGCTTCAACCGGCGTTGAAGACGGGCACGGAGGAGCTGAAGTCGCTATACAACACGG
 TAGCGACGCTCTACTGCGTGACGAGAAGATCGACGTCAGGGACACGAAGGAGGCTCTTGACAA

Mosaic and Group M Gag_dmyr-patent.doc

94/109

Fig. 22 cont'd-28

GATTGAGGAAGAGCAGAACAAAGATCCAGCAGAAGACGCAGCAGGCGAAGGAGGCAGACGGCAAA
GTATCTCAGAACTACCCGATCGTGCAGAACATCCAGGGACAGATGGTCCACCAGCCGATCTCCC
CACGGACGCTTAACGCCTGGGTCAAAGTAGTCGAGGAGAAGGCATTTCAGCCCGGAAGTGATCCC
CATGTTCACTGCACTTAGCGACGGAGCCACCCCGCAGGACCTGAACAGCATGTTGAACGCCGTC
GGCGGGCACCAGGCGGCCATGCAGATCCTTAAGGACACCATCAACGAGGAGGCTGCGGACTGGG
ACCGGCTTCACCCGGTGCACGCGGGGCCCGTCCGCGCCGGGCCAGATGAGAGAGCCGCGGGGATC
GGACATCGCGGGAACCACCAGCACCTTGCAGGAGCAAATCGGGTGGATGACTAACAACCCCCCG
ATCCCGGTGCGGGAGATCTACAAGAGATGGATCATCCTGGGGTTGAACAAGATCGTGAGGATGT
ACAGCCAGTCAGCATCCTGGACATCAAGCAGGGACCGAAGGAGTCGTTTCAGAGACTACGTCGA
CCGGTTCTTCAAAGTCCTCCGGGCGGAGCAGGCGACGCAGGACGTCAAGAACTGGATGACGGAC
ACCTTGTGATCCAGAACGCGAACCCGGACTGCAAGTCGATCCTGCGGGCTCTCGGCCCGGGAG
CGACGTTGGAAGAGATGATGACGGCGTGCCAGGGAGTCGGGGGACCTCGCACAAAGGCGCGGAT
CTTGCCCGAGGCGATGTCACAAGTGACTAACAGCGCGACCATCATGATGCAGCGGGGCAACTTC
CGGAACCAGCGGAAGACGGTGAAGTGTTTCAACTGTGGCAAGGAGGGACACCTCGCCAGGAAC
GCAAGGCCCCGCGGAAGCGGGGCTGCTGGAAGTGCGGAAAGGAGGGGCACCAAATGAAGGAGTG
CACGGAGCGGCAGGCGAATTTCTCGGGAAGATCTGGCCGTCTCTGAAGGGGCGGCCAGGGAAC
TTTCCGCAAAGCCGGCCGGAGCCGACGGCCCCGCCGAGGAGTCCTTTGCGTTGCGGGAGGAGA
CGACCACGCCCTCGCAGAAGCAAGAGCCGATCGACAAGGAGCTCTACCCTCTTGCGTCCCTCCG
GTCGCTCTTCGGCAACGACCCGTGCTCGCAAGCGTCGTGATAAGCTAGCGGATCCGGCGCGCC

NheI

AscI

>HV13316 in HV10001 4816bp

AAATGGGGGCGCTGAGGTCTGCCTCGTGAAGAAGGTGTTGCTGACTCATAACCAGGCCTGAATCG
CCCCATCATCCAGCCAGAAAGTGAGGGAGCCACGGTTGATGAGAGCTTTGTTGTAGGTGGACCA
GTTGGTGATTTTGAACCTTTGCTTTGCCACGGAACGGTCTGCGTTGTGCGGAAGATGCGTGATC
TGATCCTTCAACTCAGCAAAAGTTCGATTTATTCAACAAAGCCGCGTCCCGTCAAGTCAGCGT
AATGCTCTGCCAGTGTTACAACCAATTAACCAATTCTGCGTTCAAAATGGTATGCGTTTTGACA
CATCCACTATATATCCGTGTCTGTTCTGTCCACTCCTGAATCCCATTTCCAGAAATTCTCTAGCGA
TTCCAGAAGTTTCTCAGAGTCGGAAAGTTGACCAGACATTACGAACTGGCACAGATGGTCATAA
CCTGAAGGAAGATCTGATTGCTTAAGTCTTCAAGTAAAGACCGACGCGCTCGTCGTATAACAGA
TGCGATGATGCAGACCAATCAACATGGCACCTGCCATTGCTACCTGTACAGTCAAGGATGGTAG
AAATGTTGTGCGGTCTTGCACACGAATATTACGCCATTGCTGCTGATATTCAAACAGCTCTTCT
ACGATAAGGGGCACAAATCGCATCGTGGAACGTTTGGGCTTCTACCGATTTAGCAGTTTGATACA
CTTTCTCTAAGTATCCACCTGAATCATAAATCGGCAAAATAGAGAAAAATTGACCATGTGTAAG
CGGCCAATCTGATTCCACCTGAGATGCATAATCTAGTAGAATCTCTTCGCTATCAAAATTTCACT
TCCACCTTCCACTCACCGGTTGTCCATTTCATGGCTGAACTCTGCTTCTCTGTTGACATGACAC
ACATCATCTCAATATCCGAATACGGACCATCAGTCTGACGACCAAGAGAGCCATAAACACCAAT
AGCCTTAACATCATCCCATATTTATCCAATATTCGTTCCCTTAATTTTCATGAACAATCTTCATT
CTTTCTTCTCTAGTCATTATTATTGGTCCGTTTCATAACACCCCTTGTATTACTGTTTATGTAAG
CAGACAGTTTTATTGTTTCATGATGATATATTTTTATCTTGTCGAATGTAACATCAGAGATTTTG
AGACACAACGTGGCTTTCCCCGGCCCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTG
AGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATC
TGCTGCTTGCAAACAAAAAACACCGCTACCAGCGGTGGTTTGTGTTGCCGGATCAAGAGCTAC
CAACTCTTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCAGATACCAATACTGTTCTTCTAGT
GTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTA
ATCCTGTTACCAAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGAC
GATAGTTACCGGATAAGGCGCAGCGGTGCGGCTGAACGGGGGGTTCGTGCACACAGCCAGCTT
GGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTT
CCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGA
GGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTGCGGTTTCGCCACCTCTGACT

95/109

Fig. 22 cont'd-29

TGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACCGCG
GCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCC
CTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAAC
GACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTT
ACGCATCTGTGCGGTATTTACACCCGCATATGGTGCACCTCTCAGTACAATCTGCTCTGATGCCG
CATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTGCTGAGTAGTGCGCGAGCAA
AATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGG
CGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATAGCCGCGGCATCGATGATATCCATTGCA
TACGTTGTATCTATATCATAATATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGT
TGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCAT
ATATGGAGTTCGCGGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCC
CCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGA
CGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGC
CAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACAT
GACCTTACGGGACTTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTG
ATGCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTC
TCCACCCCATTTGACGTCAATGGGAGTTTGTTTTGGACCAAATCAACGGGACTTTCCAAAATG
TCGTAACAACCTCCGCCCCATTGACGCAAATGGGCGGTAGGCGGTACGGTGGGAGGTCTATATA
AGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCC
ATAGAAGACACCGGGACCGATCCAGCCTCCGCGGGCGCGCGTCGAGAAGAAAATGGCGGCTCGC
GCCTCGGTCCCTTAGCGGGGGCAAGTTGGATGCGTGGGGAAAAGATCCGCTTGAGGGCCAGGAGGGA
AGAAGAAGTACCGGCTCAAGCACCTTGTCTGGGCGAGCAGAGAGCTCGACCGGTTTCGCGCTGAA
CCCGTCGCTGCTTGAGACAGCCGAGGGCTGCAAGCAAATCATGAAGCAGCTTCAACCGGCGTTG
AAGACGGGCACCGAGGAGCTGAAGTCGCTATACAACACGGTAGCGACGCTCTACTGCGTGCACG
AGAAGATCGACGTCAGGGACACGAAGGAGGCTCTTGACAAGATTGAGGAAGAGCAGAACAAGAT
CCAGCAGAAGACGCAGCAGGCGAAGGAGGCAGACGGCAAAGTATCTCAGAACTACCCGATCGTG
CAGAACATCCAGGGACAGATGGTCCACCAGCCGATCTCCCCACGGACGCTTAACCGCTGGGTCA
AAGTAGTCGAGGAGAAGGCATTGAGCCCGGAAGTGATCCCCATGTTCACTGCACTTAGCGACGG
AGCCACCCCGCAGGACCTGAACAGCATGTTGAACGCCGTCGGCGGGCACCAGGCGGCCATGCAG
ATCCTTAAGGACACCATCAACGAGGAGGCTGCGGACTGGGACCGGCTTCAACCGGTGCACGCGG
GGCCCGTCGCGCCGGGCCAGATGAGAGAGCCGCGGGGATCGGACATCGCGGGAACCACCAGCAC
CTTGACAGGAGCAAATCGGGTGGATGACTAACAACCCCCGATCCCGGTGCGGGGAGATCTACAAG
AGATGGATCATCCTGGGGTTGAACAAGATCGTGAGGATGTACAGCCCAGTCAGCATCCTGGACA
TCAAGCAGGGACCGAAGGAGTCGTTGAGAGACTACGTCGACCGGTTCTTCAAAGTCCTCCGGGC
GGAGCAGGCGACGCAGGACGTCAAGAACTGGATGACGGACACCTTGTGATCCAGAACGCGAAC
CCGGAAGTCAAGTCGATCCTGCGGGCTCTCGGCCCGGAGCGACGTTGGAAGAGATGATGACGG
CGTGCCAGGGAGTCGGGGGACCTCGCACAAGGCGCGGATCTTGGCCGAGGCGATGTCAAGT
GACTAACAGCGCGACCATCATGATGCAGCGGGCAACTTCCGGAACCAGCGGAAGACGGTGAAG
TGTTTTCAACTGTGGCAAGGAGGGACACCTCGCCAGGAAGTCAAGGCCCGCGGAAGCGGGGCT
GCTGGAAGTGCAGAAAGGAGGGGCACCAAATGAAGGAGTGACGGAGCGGCAGGCGAATTTCTT
CGGGAAGATCTGGCCGTCCTCGAAGGGGCGGCCAGGGAAGTTCGCAAGCCGGCCGGAGCCG
ACGGCCCCCGCGGAGGAGTCCTTTTCGGTTTCGGGGAGGAGACGACCACGCCCTCGCAGAAGCAAG
AGCCGATCGACAAGGAGCTCTACCCTCTTGCGTCCCTCCGGTCGCTCTTCGGCAACGACCCGTC
GTCGCAAGCGTCGTGATAAGCTAGCGGATCCGGCGCGCCGAGCTCGCTGATCAGCCTCGACTGT
GCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCCGTGCTTCTTGAACCTGGAAGGT
GCCACTCCCAGTGTCTTTTCTAATAAAATGAGGAAATTGCATCGCATTTGTCTGAGTAGGTGTC
ATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAG
GCATGCTGGGGAATTT

Thursday, August 2, 2007

Sequence 1 Length : 4816

AatII GACGTC
6 Sites

Mosaic and Group M Gag_dmyr-patent.doc

96/109

Fig. 22 cont'd-30

2510 2563 2646
 2832 3341 3989
 AccI GTMKAC
 1 Site
 3941
 AccIII TCCGGA
 1 Site
 4199
 AflII CTTAAG
 1 Site
 3652
 AflIII ACRYGT
 1 Site
 1897
 AluI AGCT
 15 Sites
 109 633 1340
 1597 1643 1733
 1959
 2184 2951 3180
 3249 3282 4497
 4564
 4586
 AlwNI CAGNNNCTG
 4 Sites
 1488 2129 3600
 4262
 AasII GRGCGC
 7 Sites
 2507 2560 2643
 2829 2983 3338
 3986
 ApaI GGGCCC
 1 Site
 3716
 ApaLI GTGCAC
 5 Sites
 1583 2081 3322
 3703 4326
 AvaI CYCGRG
 2 Sites
 4065 4351
 BamHI GGATCC
 1 Site
 4570
 BanI GGYRCC
 5 Sites
 538 2850 3629
 4310 4670
 BanII GRGCGC
 6 Sites
 2953 3182 3716
 4059 4499 4588
 BclI TGATCA
 1 Site
 4592
 BcnI CCSGG
 11 Sites
 1173 1521 3021
 3548 3701 3726
 3821
 3965 4034 4067
 4068
 BglI GCCNNNNNGGC
 3 Sites
 2475 2597 2668
 BglII AGATCT
 3 Sites
 458 3830 4358
 Bsp1286 GDGCHC
 15 Sites
 652 1587 2085
 2953 3182 3274
 3326

Mosaic and Group M Gag_dmyr-patent.doc

3632 3707 3716
 4059 4313 4330
 4499
 4588
 BspHI TCATGA
 4 Sites
 1007 1105 3239
 4178
 BspMII TCCGGA
 1 Site
 4199
 BspNI CCWGG
 17 Sites
 52 1738 1751
 1872 2475 2668
 2977
 3128 3466 3513
 3634 3853 3898
 4102
 4257 4385 4664
 BssHII GCGCGC
 2 Sites
 3045 4579
 BstNI CCWGG
 17 Sites
 52 1738 1751
 1872 2475 2668
 2977
 3128 3466 3513
 3634 3853 3898
 4102
 4257 4385 4664
 Cfr10I RCCGGY
 7 Sites
 847 3146 3185
 3255 3689 3944
 4403
 CfrI YGGCCR
 6 Sites
 769 3638 4139
 4363 4381 4405
 ClaI ATCGAT
 1 Site
 2287
 DdeI CTNAG
 14 Sites
 12 204 397
 711 787 1214 1623
 2088 2158 2229
 3082 3437 3574
 4724
 DpnI GATC
 28 Sites
 190 195 460
 1239 1247 1258
 1333
 2972 3028 3115
 3334 3391 3451
 3490
 3556 3649 3750
 3817 3832 3847
 3868
 4018 4048 4135
 4360 4486 4572
 4594
 DraIII CACNNNGTG
 2 Sites
 1161 3702
 Eco47I GGWCC
 11 Sites
 122 586 919
 1048 3021 3077
 3476

3596 3686 3912
 4113
 EcoO109 RGGNCCY
 4 Sites
 3597 3712 4114
 4269
 EcoRII CCWGG
 17 Sites
 50 1736 1749
 1870 2473 2666
 2975
 3126 3464 3511
 3632 3851 3896
 4100
 4255 4383 4662
 EcoRV GATATC
 1 Site
 2294
 Fnu4HI GCNGC
 27 Sites
 234 769 1283
 1489 1492 1557
 1700
 1855 1973 1976
 1994 2110 2250
 2279
 2282 3065 3208
 3228 3247 3406
 3638
 3677 3742 4186
 4287 4336 4381
 FnuDII CGCG
 21 Sites
 494 1273 1854
 2169 2257 2281
 2445
 3039 3045 3047
 3071 3193 3709
 3721
 3744 3760 4027
 4131 4171 4275
 4579
 HaeII RGCGCY
 3 Sites
 12 1657 2027
 HaeIII GGCC
 21 Sites
 55 771 1175
 1423 1857 1875
 1886
 2268 2469 2662
 3126 3640 3714
 3728
 4064 4141 4270
 4365 4383 4407
 4420
 HgiAI GWGCWC
 9 Sites
 1587 2085 2953
 3182 3326 3707
 4330
 4499 4588
 HhaI GCGC
 20 Sites
 11 496 1273
 1382 1556 1656
 1723
 1993 2026 2169
 2249 3045 3047
 3073
 3195 3723 4131
 4171 4579 4581
 HincII GTYRAC
 4 Sites

97/109

Fig. 22 cont'd-31

413 886 2369
 3942
 HinfI GATTC
 16 Sites
 43 59 357
 383 401 725 779
 807 1527 1923
 1998 2222 2795
 3922
 4106 4432
 HinfI GCGC
 20 Sites
 9 494 1271
 1380 1554 1654
 1721
 1991 2024 2167
 2247 3043 3045
 3071
 3193 3721 4129
 4169 4577 4579
 HpaII CCGG
 25 Sites
 848 1172 1329
 1519 1545 1692
 3019
 3147 3186 3256
 3547 3690 3700
 3724
 3820 3945 3963
 4033 4066 4200
 4404
 4408 4426 4518
 4574
 MaeI CTAG
 7 Sites
 378 801 1034
 1404 2385 4565
 4614
 MaeII ACGT
 14 Sites
 669 1160 1196
 2306 2507 2519
 2560
 2643 2724 2829
 3338 3938 3986
 4075
 MaeIII GTNAC
 10 Sites
 270 1134 1361
 1477 1540 2446
 2533
 2882 4151 4158
 MvaI CCNGG
 28 Sites
 52 1172 1520
 1738 1751 1872
 2475
 2668 2977 3020
 3128 3466 3513
 3547
 3634 3700 3725
 3820 3853 3898
 3964
 4033 4066 4067
 4102 4257 4385
 4664
 NaeI GCCGGC
 1 Site
 4405
 NciI CCSGG
 11 Sites
 1172 1520 3020
 3547 3700 3725
 3820
 Mosaic and Group M Gag_dmyr-patent.doc

3964 4033 4066
 4067
 NcoI CCATGG
 1 Site
 2745
 NdeI CATATG
 2 Sites
 2076 2619
 NheI GCTAGC
 1 Site
 4564
 NlaIII CATG
 19 Sites
 538 762 864
 892 1011 1109 1181
 1901 2219 2349
 2367 2689 2749
 3243
 3564 3612 3645
 4182 4805
 NlaIV GGNCC
 22 Sites
 92 540 1830
 1869 2852 3023
 3585
 3631 3688 3713
 3714 3765 3914
 4115
 4116 4204 4271
 4312 4412 4421
 4572
 4672
 NruI TCGCGA
 1 Site
 2257
 NsiI ATGCAT
 1 Site
 796
 Nsp7524I RCATGY
 3 Sites
 1901 3612 4805
 NspBII CMGCKG
 8 Sites
 1314 1559 2281
 3039 3744 4188
 4209
 4275
 PpuMI RGGWCCY
 2 Sites
 3597 4114
 PsaI RGGNCCY
 4 Sites
 3600 3715 4117
 4272
 PvuI CGATCG
 2 Sites
 3452 4487
 RsaI GTAC
 11 Sites
 559 2093 2263
 2330 2604 2684
 2717
 2768 2925 3145
 3880
 SacI GAGCTC
 4 Sites
 2953 3182 4499
 4588
 SacII CCGCGG
 4 Sites
 2282 3040 3745
 4276
 SalI GTCGAC
 1 Site

3940
 Sau3A GATC
 28 Sites
 188 193 458
 1237 1245 1256
 1331
 2970 3026 3113
 3332 3389 3449
 3488
 3554 3647 3748
 3815 3830 3845
 3866
 4016 4046 4133
 4358 4484 4570
 4592
 Sau96A GGNCC
 21 Sites
 123 587 920
 1049 1174 2266
 2468
 2661 3022 3078
 3477 3597 3687
 3712
 3713 3726 3913
 4063 4114 4269
 4419
 ScrFI CCNGG
 28 Sites
 52 1172 1520
 1738 1751 1872
 2475
 2668 2977 3020
 3128 3466 3513
 3547
 3634 3700 3725
 3820 3853 3898
 3964
 4033 4066 4067
 4102 4257 4385
 4664
 SdnI GDGCHC
 15 Sites
 652 1587 2085
 2953 3182 3274
 3326
 3632 3707 3716
 4059 4313 4330
 4499
 4588
 SniI GWCC
 11 Sites
 123 587 920
 1049 3022 3078
 3477
 3597 3687 3913
 4114
 SmaI CCCGGG
 1 Site
 4067
 SnaBI TACGTA
 1 Site
 2725
 SpeI ACTAGT
 1 Site
 2384
 SphI GCATGC
 1 Site
 4805
 SspI AATATT
 2 Sites
 603 991
 StuI AGGCCT
 1 Site
 55

98/109

Fig. 22 cont'd-32

StyI	CCWWGG	Tth111I	GACNNNGTC	1 Site	
1 Site		1 Site		4065	
2745		3937		XmaIII	CGGCCG
TaqI	TCGA	XhoII	RGATCY	1 Site	
12 Sites		10 Sites		4405	
216	1799	458	1245	XmnI	GAANNNTTC
3050	3182	3113	3389	3 Sites	
3527		3830	3647	811	3538 4225
3941	4045	4133	4358		
4487	4601	XmaI	CCCGGG		

Need re-create XhoI site at the 5' end

Primer:

Gag-M2-4-fg/C:

GGGCGCCTCGAGAAGAAAATGGCGGCTCG

Gag_M4.3 Dmyr

MAARASILRGGKLDKWEKIRLRPGGKKRYMLKHLIWASRELERFALNPGLLETAEGCQQIIEQL
 QSTLKTGSEELKSLFNTVATLYCVHQRIEVKDTKEALDKVEEEQNKSKKKAQQAAADTGNSSQV
 SQNYPPIVQNLQGMVHQALSPRTLNAWVKVIEEKAFSPEIIPMFTALSEGATPSDLNTMLNTVG
 GHQAAMQMLKDTINEEAAEWDRVHPVHAGPIPPGQMREPRGSDIAGTTSSLQEQUIAWMTSNPPV
 PVGEIYKRWIVLGLNKIVRMYSFVSILDIRQGPKEPFRDYVDRFFKTLRAEQASQDVKNWMTET
 LLVQANANPDCKTILRALGPGASLEEMMTACQGVGGPSHKARVLAEAMSQTNSAILMQRSNFKGS
 KRIVKCFNCGKEGHIARNCRAPRKRGCWKCGQEGHQMKDCNERQANFLGKIWPSHKGRPGNFLQ
 NRPEPTAPPEPTAPPAESFRFEETTPAPKQELKDREPLTSLKSLFGSDPLSQAS

HV13317 (Gag_M4.3 Dmyr.wlv)

GTCGAGAAGAAAATGGCGGCTCGCGCCTCGATCCTTCGAGGGGGCAAGTTGGATAAGTGGGAAA
 AGATCCGCTTGAGGCCAGGAGGGAAGAAGAGATACATGCTCAAGCACCTGATCTGGGCGAGCAG
 AGAGCTGGAGCGGTTTCGCGCTGAACCCGGGCTTGCTTGAGACAGCGGAGGGCTGTCAGCAAATC
 ATCGAGCAGCTTCAAAGCACGCTGAAGACGGGCAGCGAGGAGCTGAAGTCGCTATTCAACACGG
 TAGCGACCTCTACTGCGTGACACGCGGATCGAGGTCAAGGACACGAAGGAGGCTCTTGACAA
 GGTGGAGGAAGAGCAGAACAGTCAAGAAGAAGGCGCAGCAGGCGGCGGCCGACACCCGGCAAC
 TCCTCACAAGTATCTCAGAACTACCCGATCGTGACAGAACCTGCAGGGACAGATGGTCCACCAGG
 CCCTCTCCCCACGACGCTTAACGCCTGGGTCAAAGTAATCGAGGAGAAGGCCTTCAGCCCGGA
 AATCATCCCCATGTTACAGCACTTTCGAGGGAGCCACCCGAGCGACCTGAACACGATGTTG
 AACACCGTCGGCGGGCACCAGGCGGCCATGCAGATGCTTAAGGACACCATCAACGAGGAGGCTG
 CGGAGTGGGACCGGGTGACCCGGTGACGCGGGGCCATCCACCGGGCCAGATGAGAGAGCC
 GCGGGGATCGGACATCGCGGGAACCACGAGCTTGAGGAGCAAATCGCCTGGATGACTTCG
 AACCCCCCGGTCCCGGTGGGGAGATCTACAAGAGATGGATCGTCCTCGGGTTGAACAAGATCG
 TGAGGATGTACAGCCCTGTGTCAATCCTGGACATCCGACAGGGACCGAAGGAGCCCTTCAGAGA
 CTACGTGACCGGTTCTTCAAGACTCTCCGGGCGGAGCAGGCGTCGAGGACGTCAAGAACTGG
 ATGACGGAGACCTTGTTGGTCCAGAACGCTAACCCGGACTGCAAGACCATCCTGCGCGCTCTCG
 GCCCGGGAGCGTCTTTGGAAGAGATGATGACGGCGTGCCAGGGAGTCGGGGGACCTTCACACAA
 GGCGCGGGTCTTGCGCGAGGCGATGAGCCAGACGAACTCCGCCATCCTCATGCAGCGGTCCAAC
 TTCAAGGGAAGCAAGCGGATCGTCAAGTGTTTCAACTGTGGCAAGGAGGGACACATCGCCAGGA
 ACTGCCGGGGCCCCGCGGAAGCGAGGCTGCTGGAAGTGCGGACAGGAGGGGCACCAAATGAAGGA
 CTGCAACGAGCGCCAGGCGAATTTCTCGGGAAGATCTGGCCGTCCCACAAGGGGCGGCCAGGG
 AACTTCCTTCAAACCGGCCAGAGCCGACGGCCCCCTCCCGAGCCGACCGCCCCGCGGCGGAGT
 CCTTTCGCTTCGAGGAGACCACGCCCGCCCCCAAGCAAGAGCTCAAGGACCGCGAGCCTCTTAC

99/109

Fig. 22 cont'd-33

CTCCCTCAAGTCGCTCTTCGGCTCCGACCCGCTTTCGCAAGCGTCGTGATAAGCTAGCGGATCC
GGCGCGC

AscI

NheI

Need re-create XhoI site at the 5' end

HV13317 in HV10001 4824bp

AAATGGGGGCGCTGAGGTCTGCCTCGTGAAGAAGGTGTTGCTGACTCATACCAGGCCTGAATCG
CCCCATCATCCAGCCAGAAAGTGAGGGAGCCACGGTTGATGAGAGCTTTGTTGTAGGTGGACCA
GTTGGTGATTTTGAACCTTTGCTTTGCCACGGAACGGTCTGCGTTGTGCGGAAGATGCGTGATC
TGATCCTTCAACTCAGCAAAAGTTCGATTTATTCAACAAAGCCGCCGTCCCGTCLAGTCAGCGT
AATGCTCTGCCAGTGTTACAACCAATTAACCAATTCTGCGTTCAAAATGGTATGCGTTTGTACA
CATCCACTATATATCCGTGTCGTTCTGTCCACTCCTGAATCCCATTTCCAGAAATTCTCTAGCGA
TTCCAGAAGTTTCTCAGAGTCGGAAAGTTGACCAGACATTACGAACTGGCACAGATGGTCATAA
CCTGAAGGAAGATCTGATTGCTTAACTGCTTCAGTTAAGACCGACGCGCTCGTCGTATAACAGA
TGCGATGATGCAGACCAATCAACATGGCACCTGCCATTGCTACCTGTACAGTCAAGGATGGTAG
AAATGTTGTGCGGTCCTTGCACACGAATATTACGCCATTTGCCTGCATATTCAAACAGCTCTTCT
ACGATAAGGGCACAATCGCATCGTGGAACGTTTGGGCTTCTACCGATTTAGCAGTTTGATACA
CTTTCTCTAAGTATCCACCTGAATCATAAATCGGCAAAATAGAGAAAAATTGACCATGTGTAAG
CGGCCAATCTGATTCCACCTGAGATGCATAATCTAGTAGAATCTCTTCGCTATCAAAATTTCACT
TCCACCTTCCACTCACCGTTGTCCATTCATGGCTGAACTCTGCTTCTCTGTTGACATGACAC
ACATCATCTCAATATCCGAATACGGACCATCAGTCTGACGACCAAGAGAGCCATAAACACCAAT
AGCCTTAACATCATCCCCATATTTATCCAATATTTCGTTCCCTTAATTTTCATGAACAATCTTCATT
CTTTCTTCTCTAGTCATTATTATTGGTCCGTTTCATAACACCCCTTGTATTACTGTTTATGTAAG
CAGACAGTTTTATTGTTTCATGATGATAATTTTTATCTTGTGCAATGTAACATCAGAGATTTTG
AGACACAACGTGGCTTTCCCCGGCCCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTG
AGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATC
TGCTGCTTGCAAACAAAAAACACCGCTACCAGCGGTGGTTTGTGTTGCCGGATCAAGAGCTAC
CAACTCTTTTTCCGAAGGTAACCTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGT
GTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTA
ATCCTGTTACCAGTGCGTGTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGAC
GATAGTTACCGGATAAGGCGCAGCGGTGCGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTT
GGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTT
CCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGA
GGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTGCGGTTTCGCCACCTCTGACT
TGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCG
GCCTTTTTACGGTTCTTGGCCTTTTGTGCTGGCCTTTTGTCTCACATGTTCTTTCCTGCGTTATCCC
CTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGAGCCGAAC
GACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTT
ACGCATCTGTGCGGTATTTACACCGCATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCG
CATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTSCGCGAGCAA
AATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGG
CGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATAGCCGCGGCATCGATGATATCCATTGCA
TACGTTGTATCTATATCATAATATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGT
TGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCAT
ATATGGAGTTCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCC
CCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGA
CGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTSTATCATATGC
CAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCAGTACAT
GACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTG

Mosaic and Group M Gag_dmyr-patent.doc

100/109

Fig. 22 cont'd-34

ATGCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGA CTACGCGGGATTTC CAAGTC
 TCCACCCCATTTGACGTCAATGGGAGTTTGT TTTTGGCACCAAAATCAACGGGACTTTCCAAAATG
 TCGTAACAACCTCCGCCCCATTGACGCAAATGGGCGGTAGGCGGTGTACGGTGGGAGGTCTATATA
 AGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCC
 ATAGAAGACACCGGGACCGATCCAGCCTCCGCGGGCGCGCGTCGAGAAGAAAATGGCGGCTCGC
 GCCTCGATCCTTCGAGGGGGCAAGTTGGATAAGTGGGAAAAGATCCGCTTGAGGCCAGGAGGGA
 AGAAGAGATACATGCTCAAGCACCTGATCTGGGCGAGCAGAGAGCTGGAGCGGTTTCGCGCTGAA
 CCCGGGCCTGCTTGAGACAGCGGAGGGCTGTCTAGCAAATCATCGAGCAGCTTCAAAGCACGCTG
 AAGACGGGCAGCGAGGAGCTGAAGTCGCTATTCAACACGGTAGCGACCCTCTACTGCGTGCACC
 AGCGGATCGAGGTCAAGGACACGAAGGAGGCTCTTGACAAGGTGGAGGAAGAGCAGAACAAAGTC
 GAAGAAGAAGGCGCAGCAGGCGGCGGCCGACACCGGCAACTCCTCACAAGTATCTCAGAACTAC
 CCGATCGTGCAGAACCTGCAGGGACAGATGGTCCACCAGGCCCCCTCTCCCCACGGACGCTTAACG
 CCTGGGTCAAAGTAATCGAGGAGAAGGCCTTCAGCCCCGAAATCATCCCCATGTTTACAGCACT
 TTCCGAGGGAGCCACCCCGAGCGACCTGAACACGATGTTGAACACCGTTCGGCGGGCACCAGGCG
 GCCATGCAGATGCTTAAGGACACCATCAACGAGGAGGCTGCGGAGTGGGACCGGGTGCACCCGG
 TGCACGCGGGGCCATCCACCGGGCCAGATGAGAGAGCCGCGGGGATCGGACATCGCGGGAAC
 CACCAGCAGCTTGCAGGAGCAAATCGCCTGGATGACTTCGAACCCCCCGGTCCCCGTCTGGGGAG
 ATCTACAAGAGATGGATCGTCCTCGGGTTGAACAAGATCGTGAGGATGTACAGCCCTGTGTCAA
 TCCTGGACATCCGACAGGGACCGAAGGAGCCCTTCAGAGACTACGTGACCCGTTCTTCAAGAC
 TCTCCGGGCGGAGCAGGCGTCGCAGGACGTCAAGAACTGGATGACGGAGACCTTGTGGTCCAG
 AACGCTAACCCGGA CTGCAAGACCATCCTGCGCGCTCTCGGCCCGGAGCGTCCTTGGAAGAGA
 TGATGACGGCGTGCCAGGGAGTCGGGGGACCCTCACACAAGGCGCGGGTCTTGGCCGAGGCGAT
 GAGCCAGACGAACCTCCGCCATCCTCATGCAGCGGTCCA ACTTCAAGGGAAGCAAGCGGATCGTC
 AAGTGTTCAACTGTGGCAAGGAGGGACACATCGCCAGGAACTGCCGGGCCCCCGCGGAAGCGAG
 GCTGCTGGAAGTGGGACAGGAGGGGGCACCAATGAAGGACTGCAACGAGCGCCAGGCGAATTT
 CCTCGGGAAGATCTGGCCGTCCCACAAGGGGCGGCCAGGGAACTTCCTTCAA AACCGGCCAGAG
 CCGACGGCCCCCTCCCGAGCCGACCGCCCCCGCGCGGAGTCCTTTTCGCTTCGAGGAGACCACGC
 CCGCCCCCAAGCAAGAGCTCAAGGACCGCGAGCCTCTTACCTCCCTCAAGTCGCTCTTCGGCTC
 CGACCCGCTTTTCGAAGCGTCGTGATAAGCTAGCGGATCCGGCGCGCGAGCTCGCTGATCAGCC
 TCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCCGTGCCTTCCTTGACCC
 TGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAG
 TAGGTGTCAATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGAC
 AATAGCAGGCATGCTGGGGAATTT

Thursday, August 2, 2007

		2184	2951	3180	6 Sites		
		3249	3282	3785	3201	3600	3862
Sequence 2	Length : 4824	4497			4074	4354	4429
		4573	4594		BamHI	GGATCC	
AatII	GACGTC	AlwNI	CAGNNCTG		1 Site		
5 Sites		4 Sites			4579		
2510	2563	1488	2129	3471	BanI	GGYRCC	
2832	3998	4265			5 Sites		
AccI	GTMKAC	AosII	GRCGYC		538	2850	3638
1 Site		7 Sites			4313	4678	
3950		2507	2560	2643	BanII	GRGCTC	
AflII	CTTAAG	2829	2983	3985	6 Sites		
1 Site		3995			2953	3725	3935
3661		ApaI	GGGCCC		4275	4499	4596
AflIII	ACRYGT	2 Sites			BclI	TGATCA	
1 Site		3725	4275		1 Site		
1897		ApaLI	GTGCAC		4600		
AluI	AGCT	5 Sites			BcnI	CCSGG	
16 Sites		1583	2081	3322	16 Sites		
109	633	3703	3712		1173	1521	3021
1597	1643	AsuII	TTCGAA		3203	3204	3557
1959		1 Site			3701		
		3814					
		AvaI	CYCGRG				

Mosaic and Group M Gag_dmyr-patent.doc

101/109

Fig. 22 cont'd-35

3710 3735 3824
3830 3974 4043
4076
4077 4271
BglI GCCNNNNNGGC
4 Sites
2475 2597 2668
4447
BglII AGATCT
3 Sites
458 3839 4361
Bsp1286 GDGCHC
14 Sites
652 1587 2085
2953 3326 3641
3707
3716 3725 3935
4275 4316 4499
4596
BspHI TCATGA
2 Sites
1007 1105
BspNI CCWGG
18 Sites
52 1738 1751
1872 2475 2668
2977
3128 3493 3522
3643 3804 3907
4111
4260 4342 4388
4672
BssHII GCGCGC
3 Sites
3045 4064 4588
BstNI CCWGG
18 Sites
52 1738 1751
1872 2475 2668
2977
3128 3493 3522
3643 3804 3907
4111
4260 4342 4388
4672
Cfr10I RCCGGY
5 Sites
847 3424 3953
4406 4446
CfrI YGGCCR
7 Sites
769 3416 3647
4148 4366 4384
4408
ClaI ATCGAT
1 Site
2287
DdeI CTNAG
12 Sites
12 204 397
711 787 1214 1623
2088 2158 2229
3446 4732
DpnI GATC
22 Sites
190 195 460
1239 1247 1258
1333
2972 3028 3079
3115 3163 3334
3460
3759 3841 3856
3877 4219 4363
4581

Mosaic and Group M Gag_dmyr-patent.doc

4602
DraIII CACNNNGTG
2 Sites
1161 3711
Eco47I GGWCC
13 Sites
122 586 919
1048 3021 3485
3695
3824 3921 4025
4122 4192 4502
EcoO109 RGGNCCY
4 Sites
3495 3721 4123
4272
EcoRII CCWGG
18 Sites
50 1736 1749
1870 2473 2666
2975
3126 3491 3520
3641 3802 3905
4109
4258 4340 4386
4670
EcoRV GATATC
1 Site
2294
Fnu4HI GCNGC
28 Sites
234 769 1283
1489 1492 1557
1700
1855 1973 1976
1994 2110 2250
2279
2282 3065 3247
3273 3406 3413
3416
3647 3686 3751
3783 4189 4290
4384
PnuDII CGCG
21 Sites
494 1273 1854
2169 2257 2281
2445
3039 3045 3047
3071 3193 3718
3753
3769 4064 4140
4278 4508 4588
4590
HaeII RCGCGY
4 Sites
12 1657 2027
4341
HaeIII GGCC
25 Sites
55 771 1175
1423 1857 1875
1886
2268 2469 2662
3126 3206 3418
3496
3547 3649 3723
3737 4073 4150
4273
4368 4386 4410
4423
HgiAI GWGCWC
8 Sites

1587 2085 2953
3326 3707 3716
4499
4596
HhaI GCGC
22 Sites
11 496 1273
1382 1556 1656
1723
1993 2026 2169
2249 3045 3047
3073
3195 3405 4064
4066 4140 4340
4588
4590
HincII GTYRAC
4 Sites
413 886 2369
3951
HinfI GATC
16 Sites
43 59 357
383 401 725 779
807 1527 1923
1998 2222 2795
3966
4115 4453
HinPI GCGC
22 Sites
9 494 1271
1380 1554 1654
1721
1991 2024 2167
2247 3043 3045
3071
3193 3403 4062
4064 4138 4338
4586
4588
HpaII CCGG
23 Sites
848 1172 1329
1519 1545 1692
3019
3202 3425 3556
3699 3709 3733
3823
3829 3954 3972
4042 4075 4269
4407
4447 4583
MaeI CTAG
7 Sites
378 801 1034
1404 2385 4574
4622
MaeII ACGT
12 Sites
669 1160 1196
2306 2507 2519
2560
2643 2724 2829
3947 3995
MaeIII GTNAC
8 Sites
270 1134 1361
1477 1540 2446
2533
2882
MvaI CCNGG
34 Sites

102/109

Fig. 22 cont'd-36

52 1172 1520
 1738 1751 1872
 2475
 2668 2977 3020
 3128 3202 3203
 3493
 3522 3556 3643
 3700 3709 3734
 3804
 3823 3829 3907
 3973 4042 4075
 4076
 4111 4260 4270
 4342 4388 4672
 NaeI GCCGGC
 1 Site
 4448
 NciI CCSGG
 16 Sites
 1172 1520 3020
 3202 3203 3556
 3700
 3709 3734 3823
 3829 3973 4042
 4075
 4076 4270
 NcoI CCATGG
 1 Site
 2745
 NdeI CATATG
 2 Sites
 2076 2619
 NheI GCTAGC
 1 Site
 4573
 NlaIII CATG
 18 Sites
 538 762 864
 892 1011 1109 1181
 1901 2219 2349
 2367 2689 2749
 3150
 3573 3654 4188
 4813
 NlaIV GGNNCC
 24 Sites
 92 540 1830
 1869 2852 3023
 3594
 3640 3697 3722
 3723 3774 3827
 3923
 3932 4124 4125
 4273 4274 4315
 4424
 4542 4581 4680
 NruI TCGCGA
 1 Site
 2257
 NsiI ATGCAT
 1 Site
 796
 Nsp7524I RCATGY
 3 Sites
 1901 3150 4813
 NspBII CMGCKG
 9 Sites

1314 1559 2281
 3039 3220 3330
 3753
 4191 4278
 PpuMI RGGWCCY
 1 Site
 4123
 PsaI RGGNCCY
 4 Sites
 1498 3724 4126
 4275
 PstI CTGCAG
 1 Site
 3476
 PvuI CGATCG
 1 Site
 3461
 RsaI GTAC
 10 Sites
 559 2093 2263
 2330 2604 2684
 2717
 2768 2925 3889
 SacI GAGCTC
 3 Sites
 2953 4499 4596
 SacII CCGCGG
 4 Sites
 2282 3040 3754
 4279
 SalI GTCGAC
 1 Site
 3949
 Sau3A GATC
 22 Sites
 188 193 458
 1237 1245 1256
 1331
 2970 3026 3077
 3113 3161 3332
 3458
 3757 3839 3854
 3875 4217 4361
 4579
 4600
 Sau96A GGNNCC
 26 Sites
 123 587 920
 1049 1174 2266
 2468
 2661 3022 3204
 3486 3495 3696
 3721
 3722 3735 3825
 3922 4026 4072
 4123
 4193 4271 4272
 4422 4503
 ScrFI CCNGG
 34 Sites
 52 1172 1520
 1738 1751 1872
 2475
 2668 2977 3020
 3128 3202 3203
 3493
 3522 3556 3643
 3700 3709 3734
 3804

3823 3829 3907
 3973 4042 4075
 4076
 4111 4260 4270
 4342 4388 4672
 SdnI GDGCHC
 14 Sites
 652 1587 2085
 2953 3326 3641
 3707
 3716 3725 3935
 4275 4316 4499
 4596
 SinI GGWCC
 13 Sites
 123 587 920
 1049 3022 3486
 3696
 3825 3922 4026
 4123 4193 4503
 SmaI CCCGGG
 2 Sites
 3203 4076
 SnaBI TACGTA
 1 Site
 2725
 SpeI ACTAGT
 1 Site
 2384
 SphI GCATGC
 1 Site
 4813
 SspI AATATT
 2 Sites
 603 991
 StuI AGGCCT
 2 Sites
 55 3547
 StyI CCWWGG
 2 Sites
 2745 4085
 TaqI TCGA
 14 Sites
 216 1799 2287
 3050 3076 3084
 3242
 3335 3391 3536
 3814 3950 4466
 4609
 Tth111I GACNNNGTC
 1 Site
 3946
 XhoII RGATCY
 7 Sites
 458 1245 1256
 3113 3839 4361
 4579
 XmaI CCCGGG
 2 Sites
 3201 4074
 XmaIII CGGCCG
 1 Site
 3416
 XmnI GAANNNTTC
 2 Sites
 811 3547

103/109

Fig. 22 cont'd-37

Gag_M4.4 Dmyr

MAARASVLRGEKLDKWERIRLRPGGKKHYMLKHLVWASRELEKFALNPGLLETSEGCKQIIKQL
QPALQTGTEELRSLFNTVATLYCVHAGIEVRDTKEALDKIEEIQNKSKQKTQQAAAGTGSSSKV
SQNYPIVQNLQGMVHQPLSPRTLNAWVKVVEEKGFNPEVIPMFSALSEGATPQDLNMLNIVG
GHQAAMQMLKETINEEAAEWDRHLHPVHAGPIAPGQMREPRGSDIAGSTSTLQEQIAWMTGNPPV
PVGDIYKRWIILGLNKIVKMYSPITSILDIKQGPKEPFRDYVDRFYKTLRAEQATQEVKNWMTDT
LLVQANANPDCKSILKALGTGATLEEMMSACQGVGGPAHKARVLAEAMSQANNTNIMMQRSNFKG
PKRIIKCFNCGKEGHIAKNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGRIWPSSSKGRPGNFL
QSRPEPTAPPAEPTAPPAESFKFEETTPAPKQEPKDREPLTSLRSLFGSDPLLQAS

HV13318 (Gag_M4.4 Dmyr.wlv)

GTCGAGAAGAAAATGGCGGCTCGCGCCTCGGTCCTTCGAGGGGAGAAGTTGGATAAAGTGGGAAC
GGATCCGCTTGAGGCAGGAGGGAAGAAGCACTACATGCTCAAGCACCTGGTCTGGGCGAGCAG
AGAGCTGGAGAAGTTTCGCGCTGAACCCGGGCCTGCTTGAGACATCCGAGGGCTGTAAGCAAATC
ATCAAGCAGCTTCAACCAGCGCTCCAGACGGGCACCGAGGAGCTGCGCTCGCTATTCAACACGG
TAGCGACCCCTCTACTGCGTGCACGCCGAATCGAAGTTCGTGACACGAAGGAGGCTCTTGACAA
GATCGAGGAAATCCAGAACAAGTCGAAGCAGAAGACCCAGCAGGCGGCGGCCGGGACCGGCTCG
TCCTCAAAAGTATCTCAGAACTACCCGATCGTGAGAACCTGCAGGGACAGATGGTCCACCAGC
CGCTCTCCCCACGGACGCTTAACGCCTGGGTCAAAGTAGTCGAGGAGAAGGGGTTCAACCCGGA
AGTCATCCCCATGTTCTCGGCACCTTCCGAGGGAGCCACCCCGCAGGACCTGAACATGATGTTG
AACATCGTCGGCGGGCACCAGGCGGCCATGCAGATGCTTAAGGAGACCATCAACGAGGAGGCTG
CGGAGTGGGACCGCCTGCACCCGGTGCACGCGGGGCCCATCGCACCCGGGCCAGATGAGAGAGCC
GCGGGGATCGGACATCGCGGGATCCACCAGCACCTTGAGGAGCAAATCGCCTGGATGACTGGG
AACCCCCCGGTCCCGGTGCGGGACATCTACAAGAGATGGATCATCCTCGGGTTGAACAAGATCG
TGAAGATGTACAGCCCTACGTCAATCCTGGACATCAAGCAGGGCCCCAAGGAGCCCTTCAGAGA
CTACGTGACCGGTTCTACAAGACTCTCCGGGCGGAGCAGGCGACGACGAGGAGGTCAAGAACTGG
ATGACGGACACCTTGTTGGTCCAGAACGCTAACCCGGACTGCAAGAGCATCCTGAAGGCTCTCG
GCACGGGAGCGACCTTGGAAGAGATGATGTCCGCGTGCCAGGGAGTCGGGGGACCCGCGCACAA
GGCGCGGGTCTTGCGGAGGCGATGTCCCAGGCCAACACGAACATCATGATGCAGCGGTCC
AACTTCAAGGGACCGAAGCGGATCATCAAGTGTTCCTCACTGTGGCAAGGAGGGACACATCGCCA
AGAAGTGCAGGGCCCCCGGGAAGAAGGGCTGCTGGAAGTGCAGGAAAGGAGGGGCACCAAATGAA
GGACTGCACGGAGCGCCAGGCGAATTTCTCGGGCGGATCTGGCCGTCTCGAAGGGGCGGCCA
GGGAACCTTCTTCAATCGCGGCCAGAGCCGACGGCCCCCTCCCGCGGAGCCGACCGCCCCGCGG
CGGAGTCTTTTAAAGTTTCGAGGAGACCACGCCCCCCCCAAGCAAGAGCCGAAGGACCGCGAGCC
TCTTACCTCCCTCCGGTCTGCTCTTCGGCTCCGACCCGCTTCTGCAAGCGTCGTGATAAGCTAGC
GGATCCGGCGCGCC

NheI

AscI

HV13318 in in HV10001 4831bp

AAATGGGGGCGCTGAGGTCTGCCTCGTGAAGAAGGTGTTGCTGACTCATAACCAGGCCTGAATCG
CCCCATCATCCAGCCAGAAAGTGAGGGAGCCACGGTTGATGAGAGCTTTGTTGTAGGTGGACCA
GTTGGTGATTTTGAACCTTTGCTTTGCCACGGAACGGTCTGCGTTGTGCGGAAGATGCGTGATC
TGATCCTTCAACTCAGCAAAAGTTGATTTATTCAACAAAGCCGCGTCCCGTCAAGTCAGCGT
AATGCTCTGCCAGTGTTACAACCAATTAACCAATTCTGCGTTCAAAATGGTATGCGTTTTGACA
CATCCACTATATATCCGTGTGCTTCTGTCCACTCCTGAATCCCATTCAGAAATTCTCTAGCGA
TTCCAGAAGTTTCTCAGAGTCGGAAGTTGACCAGACATTACGAACTGGCACAGATGGTCATAA
CCTGAAGGAAGATCTGATTGCTTAACTGCTTCAGTTAAGACCGACGCGCTCGTCGTATAACAGA
TGCGATGATGCAGACCAATCAACATGGCACCTGCCATTGCTACCTGTACAGTCAAGGATGGTAG
AAATGTTGTGCGTCTTGCACACGAATATTACGCCATTGCTGCATATTCAAACAGCTCTTCT

Mosaic and Group M Gag_dmyr-patent.doc

104/109

Fig. 22 cont'd-38

ACGATAAGGGCACAAATCGCATCGTGGAAACGTTTGGGCTTCTACCGATTTAGCAGTTTGATACA
CTTTCTCTAAGTATCCACCTGAATCATAAATCGGCAAAATAGAGAAAAATTGACCATGTGTAAG
CGGCCAATCTGATTCACCTGAGATGCATAATCTAGTAGAATCTCTTCGCTATCAAAATTCACT
TCCACCTTCCACTCACCGGTTGTCCATTTCATGGCTGAACTCTGCTTCCTCTGTTGACATGACAC
ACATCATCTCAATATCCGAATACGGACCATCAGTCTGACGACCAAGAGAGCCATAAACACCAAT
AGCCTTAACATCATCCCCATATTTATCCAATATTCGTTCTTAATTTTCATGAACAATCTTCATT
CTTTCTTCTCTAGTCATTATTATTGGTCCGTTTCATAACACCCCTTGTATTACTGTTTATGTAAG
CAGACAGTTTTATTGTTTCATGATGATATATTTTATCTTGTGCAATGTAACATCAGAGATTTTG
AGACACAACGTGGCTTTCCCCGGGCCATGACCAAAATCCCTTAACGTGAGTTTTTCGTTCCACTG
AGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATC
TGCTGCTTGCAAACAAAAAACACCGCTACCAGCGGTGGTTTGTGTTGCCGGATCAAGAGCTAC
CAACTCTTTTTCCGAAGGTAACCTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGT
GTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTA
ATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGAC
GATAGTTACCGGATAAGGCGCAGCGGTGCGGCTGAACGGGGGGTTTCGTGCACACAGCCCAGCTT
GGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTT
CCCGAAGGGAGAAAGGCGGACAGGTATCCGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGA
GGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTGCGGTTTCGCCACCTCTGACT
TGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCG
GCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCTGCGTTATCCC
CTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAAC
GACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTT
ACGCATCTGTGCGGTATTTACACCGCATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCG
CATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAA
AATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGG
CGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATAGCCGCGGCATCGATGATATCCATTGCA
TACGTTGTATCTATATCATAATATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGT
TGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCAT
ATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCCGCTGGCTGACCGCCCAACGACCC
CCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGA
CGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGC
CAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCCGCTGGCATTATGCCCAGTACAT
GACCTTACGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTG
ATGCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTC
TCCACCCCATTGACGTCAATGGGAGTTTGTGTTTGGCACCAAAATCAACGGGACTTTCCAAAATG
TCGTAACAACCTCCGCCCATTGACGCAAATGGGCGGTAGGCGGTACGGTGGGAGGTCTATATA
AGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCC
ATAGAAGACACCGGGACCGATCCAGCCTCCGCGGGCGCGCTCGAGAAGAAAATGGCGGCTCGC
GCCTCGGTCTTTCGAGGGGAGAAGTTGGATAAGTGGGAACGGATCCGCTTGAGGCCAGGAGGGA
AGAAGCACTACATGCTCAAGCACCTGGTCTGGGCGAGCAGAGAGCTGGAGAAGTTCGCGCTGAA
CCCGGGCCTGCTTGAGACATCCGAGGGCTGTAAGCAAATCATCAAGCAGCTTCAACCAGCGCTC
CAGACGGGCACCGAGGAGCTGCGCTCGCTATTCAACACGGTAGCGACCCCTCTACTGCGTGCACG
CCGGAATCGAAGTTCGTGACACGAAGGAGGCTCTTGACAAGATCGAGGAAATCCAGAACAAGTC
GAAGCAGAAGACCCAGCAGGCGGCGCGGACCGGCTCGTCTCCTCAAAGTATCTCAGAACTAC
CCGATCGTGCAGAACCTGCAGGGACAGATGGTCCACCAGCCGCTCTCCCCACGGACGCTTAACG
CCTGGGTCAAAGTAGTCGAGGAGAAGGGGTTCAACCCGGAAGTCATCCCCATGTTCTCGGCACT
TTCCGAGGGAGCCACCCCGCAGGACCTGAACATGATGTTGAACATCGTCGGCGGGCACCAGGCG
GCCATGCAGATGCTTAAGGAGACCATCAACGAGGAGGCTGCGGAGTGGGACCGCCTGCACCCGG
TGCACGCGGGGCCCATCGCACCGGGCCAGATGAGAGAGCCGCGGGGATCGGACATCGCGGGATC

HV 1001-
R3113

105/109

Fig. 22 cont'd-39

CACCAGCACCTTGCAGGAGCAAATCGCCTGGATGACTGGGAACCCCCCGGTCCCGGTCTGGGGAC
 ATCTACAAGAGATGGATCATCCTCGGGTTGAACAAGATCGTGAAGATGTACAGCCCTACGTCAA
 TCCTGGACATCAAGCAGGGCCCCAAGGAGCCCTTCAGAGACTACGTGACCGGTTCTACAAGAC
 TCTCCGGGCGGAGCAGGCGACGCAGGAGGTCAAGAACTGGATGACGGACACCTTGTTGGTCCAG
 AACGCTAACCCGGACTGCAAGAGCATCCTGAAGGCTCTCGGCACGGGAGCGACCTTGGAAGAGA
 TGATGTCCGCGTGCCAGGGAGTCGGGGGACCCGCGCACAAGGCGCGGTCTTGCCCGAGGGCGAT
 GTCCAGGCCAACACACGAACATCATGATGCAGCGGTCCAACCTCAAGGGACCGAAGCGGATC
 ATCAAGTGTTCAACTGTGGCAAGGAGGGACACATCGCCAAGAACTGCCGGGCCCCGCGGAAGA
 AGGGCTGCTGGAAGTGCGGAAAGGAGGGGCACCAAATGAAGGACTGCACGGAGCGCCAGGCGAA
 TTTCCTCGGGCGGATCTGGCCGTCTCGAAGGGGCGGCCAGGGAACTTCCTTCAATCGCGGCCA
 GAGCCGACGGCCCCCTCCCGCGGAGCCGACCGCCCCGCGGCGGAGTCCTTTAAGTTCGAGGAGA
 CCACGCCCCGCCCAAGCAAGAGCCGAAGGACCGCGAGCCTCTTACCTCCCTCCGGTCGCTCTT
 CGGCTCCGACCCGCTTCTGCAAGCGTCG TGATAAGCTAGCGGATCCGGCGCGCCGAGCTCGCTG
ATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCC
TTGACCCTGGAAGGTGCCACTCCCCTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATT
GTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTG
GGAAGACAATAGCAGGCATGCTGGGGAATTT

Primers:

HV1001-F2892

CCGCCCCATTGACGCAAATGG

HV1001-R3113

GCTGGCAACTAGAAGGCACAG

(+ strand: CTGTGCCTTCTAGTTGCCAGC

Thursday, August 2, 2007

Sequence 4	Length : 4831	3725 ApaI	3925 GTGCAC	4278	BglII	AGATCT
AatII	GACGTC	4 Sites	1583	2081	3322	1 Site
4 Sites		3712				458
2510	2563	2646				BspI286
2832						14 Sites
AccI	GTMKAC	AvaI	CYCGRG			652
1 Site		3 Sites	3201	3862	4357	1587
3950		BamHI	GGATCC			2085
AflII	CTTAAG	3 Sites	3113	3772	4585	3274
1 Site		BanI	GGYRCC			3326
3661		6 Sites	538	2850	3271	3716
AflIII	ACRYGT	3638	4316	4685		3925
1 Site		BanII	GRGCTC			4319
1897		6 Sites	2953	3725	3925	4603
AluI	AGCT	3935	4278	4603		BspHI
14 Sites		BclI	TGATCA			TCATGA
109	633	1340				3 Sites
1597	1643	1733				1007
1959						1105
2184	2951	3180				4184
3249	3282	4579				BspNI
4601						CCWGG
AlwNI	CAGNNCTG	1173	1521	3021		18 Sites
4 Sites		3203	3204	3421		52
1488	2129	3471				1738
3609						1751
AosII	GRCGYC	3557				2668
5 Sites		3710	3735	3824		2977
2507	2560	2643				3128
2829	2983					3160
ApaI	GGGCCC	2475	2597	2668		3522
3 Sites		4453				3804
						3907
						4111
						4165
						4345
						4391
						4679
						BesHII
						GCGCGC
						2 Sites
						3045
						4594
						BstNI
						CCWGG
						18 Sites
						52
						1738
						1751
						2668
						2977

Mosaic and Group M Gag_dmyr-patent.doc

106/109

Fig. 22 cont'd-40

3128 3160 3522
 3643 3804 3907
 4111
 4165 4345 4391
 4679
 Cfr10I RCCGGY
 4 Sites
 847 3424 3953
 4452
 CfrI YGGCCR
 7 Sites
 769 3416 3647
 4148 4369 4387
 4411
 ClaI ATCGAT
 1 Site
 2287
 DdeI CTNAG
 12 Sites
 12 204 397
 711 787 1214 1623
 2088 2158 2229
 3446 4739
 DpnI GATC
 20 Sites
 190 195 460
 1239 1247 1258
 1333
 2972 3028 3115
 3370 3460 3759
 3774
 3856 3877 4222
 4366 4587 4609
 DraIII CACNNNGTG
 2 Sites
 1161 3711
 Eco47I GGWCC
 16 Sites
 122 586 919
 1048 3021 3077
 3421
 3485 3605 3695
 3824 4025 4122
 4195
 4209 4508
 Eco47III AGCGCT
 1 Site
 3260
 EcoO109 RGGNCCY
 6 Sites
 3606 3721 3921
 3922 4123 4275
 EcoRII CCWGG
 18 Sites
 50 1736 1749
 1870 2473 2666
 2975
 3126 3158 3520
 3641 3802 3905
 4109
 4163 4343 4389
 4677
 EcoRV GATATC
 1 Site
 2294
 Pnu4HI GCNGC
 28 Sites
 234 769 1283
 1489 1492 1557
 1700
 1855 1973 1976
 1994 2110 2250
 2279

2282 3065 3247
 3283 3413 3416
 3496
 3647 3686 3751
 4192 4293 4387
 4411
 PnuDII CGCG
 23 Sites
 494 1273 1854
 2169 2257 2281
 2445
 3039 3045 3047
 3071 3193 3718
 3753
 3769 4105 4129
 4140 4281 4410
 4435
 4514 4594
 HaeII RGCGCY
 5 Sites
 12 1657 2027
 3262 4344
 HaeIII GGCC
 24 Sites
 55 771 1175
 1423 1857 1875
 1886
 2268 2469 2662
 3126 3206 3418
 3649
 3723 3737 3923
 4150 4168 4276
 4371
 4389 4413 4426
 HgiAI GWGCWC
 6 Sites
 1587 2085 2953
 3326 3716 4603
 HhaI GCGC
 22 Sites
 11 496 1273
 1382 1556 1656
 1723
 1993 2026 2169
 2249 3045 3047
 3073
 3195 3261 3287
 4131 4140 4343
 4594
 4596
 HincII GTYRAC
 4 Sites
 413 886 2369
 3951
 HinfI GATTC
 17 Sites
 43 59 357
 383 401 725 779
 807 1527 1923
 1998 2222 2795
 3332
 3966 4115 4459
 HinfI GCGC
 22 Sites
 9 494 1271
 1380 1554 1654
 1721
 1991 2024 2167
 2247 3043 3045
 3071
 3193 3259 3285
 4129 4138 4341
 4592
 4594

HpaII CCGG
 23 Sites
 848 1172 1329
 1519 1545 1692
 3019
 3202 3329 3419
 3425 3556 3709
 3733
 3823 3829 3954
 3972 4042 4272
 4453
 4533 4589
 MaeI CTAG
 7 Sites
 378 801 1034
 1404 2385 4580
 4629
 MaeII ACGT
 12 Sites
 669 1160 1196
 2306 2507 2519
 2560
 2643 2724 2829
 3898 3947
 MaeIII GTNAC
 9 Sites
 270 1134 1361
 1477 1540 2446
 2533
 2882 3343
 MvaI CCNGG
 32 Sites
 52 1172 1520
 1738 1751 1872
 2475
 2668 2977 3020
 3128 3160 3202
 3203
 3420 3522 3556
 3643 3709 3734
 3804
 3823 3829 3907
 3973 4042 4111
 4165
 4273 4345 4391
 4679
 NaeI GCCGGC
 1 Site
 4454
 NciI CCSGG
 14 Sites
 1172 1520 3020
 3202 3203 3420
 3556
 3709 3734 3823
 3829 3973 4042
 4273
 NcoI CCATGG
 1 Site
 2745
 NdeI CATATG
 2 Sites
 2076 2619
 NheI GCTAGC
 1 Site
 4579
 NlaIII CATG
 19 Sites
 538 762 864
 892 1011 1109 1181
 1901 2219 2349
 2367 2689 2749
 3150

107/109

Fig. 22 cont'd-41

3573 3618 3654
 4188 4820
 NlaIV GGNCC
 31 Sites
 92 540 1830
 1869 2852 3023
 3115
 3273 3423 3594
 3640 3697 3722
 3723
 3774 3817 3827
 3923 3924 3932
 4124
 4125 4211 4276
 4277 4318 4427
 4439
 4548 4587 4687
 NruI TCGCGA
 1 Site
 2257
 NsiI ATGCAT
 1 Site
 796
 Nsp7524I RCATGY
 3 Sites
 1901 3150 4820
 NspBII CMGCKG
 8 Sites
 1314 1559 2281
 3039 3753 4194
 4281
 4435
 PpuMI RGGWCCY
 2 Sites
 3606 4123
 PssI RGGNCCY
 6 Sites
 3609 3724 3924
 3925 4126 4278
 PstI CTGCAG
 1 Site
 3476
 PvuI CGATCG
 1 Site
 3461
 RsaI GTAC
 10 Sites
 559 2093 2263
 2330 2604 2684
 2717
 2768 2925 3889
 SacI GAGCTC
 2 Sites
 2953 4603
 SacII CCGCGG
 5 Sites
 2282 3040 3754
 4282 4436
 SalI GTCGAC

1 Site
 3949
 Sau3A GATC
 20 Sites
 188 193 458
 1237 1245 1256
 1331
 2970 3026 3113
 3368 3458 3757
 3772
 3854 3875 4220
 4364 4585 4607
 Sau96A GGNCC
 29 Sites
 123 587 920
 1049 1174 2266
 2468
 2661 3022 3078
 3204 3422 3486
 3606
 3696 3721 3722
 3735 3825 3921
 3922
 4026 4123 4196
 4210 4274 4275
 4425
 4509
 ScrFI CCNGG
 32 Sites
 52 1172 1520
 1738 1751 1872
 2475
 2668 2977 3020
 3128 3160 3202
 3203
 3420 3522 3556
 3643 3709 3734
 3804
 3823 3829 3907
 3973 4042 4111
 4165
 4273 4345 4391
 4679
 SdnI GDGCHC
 14 Sites
 652 1587 2085
 2953 3274 3326
 3641
 3716 3725 3925
 3935 4278 4319
 4603
 SinI GGWCC
 16 Sites
 123 587 920
 1049 3022 3078
 3422
 3486 3606 3696
 3825 4026 4123
 4196

4210 4509
 SmaI CCCGGG
 1 Site
 3203
 SnaBI TACGTA
 1 Site
 2725
 SpeI ACTAGT
 1 Site
 2384
 SphI GCATGC
 1 Site
 4820
 SspI AATATT
 2 Sites
 603 991
 StuI AGGCCT
 1 Site
 55
 StyI CCWWGG
 3 Sites
 2745 3926 4085
 TagI TCGA
 13 Sites
 216 1799 2287
 3050 3084 3335
 3371
 3391 3536 3950
 4378 4472 4616
 Tth111I GACNNNGTC
 1 Site
 3946
 XhoII RGATCY
 7 Sites
 458 1245 1256
 3113 3772 4364
 4585
 XmaI CCCGGG
 1 Site
 3201
 XmaIII CGGCCG
 1 Site
 3416
 XmnI GAANNNTTC
 2 Sites
 811 3547
 Following enzymes have no sites
 AccIII Asp718 AsuII
 AvrII BalI BbeI
 BspMII BstEII BstXI
 DraI EcoRI EspI
 FspI HindIII HpaI
 KpnI MluI MstI
 NarI NotI OxaNI
 PflMI PvuII RsrII
 ScaI SfiI SplI
 XbaI XcaI XhoI

Primer below can be used for Gag-M4.1 through 4.4 to generate XhoI site:

Gag-M2-4-fG/C: GGGCGCCTCGAGAAGAAAATGGCGGCTCG

WLV001AM (vector sequence), hv10001

AAATGGGGGCGCTGAGGTCTGCCTCGTGAAGAAGGTGTTGCTGACTCATACCAGGCCTGAATCG
 CCCCATCATCCAGCCAGAAAGTGAGGGAGCCACGGTTGATGAGAGCTTTGTTGTAGGTGGACCA

Mosaic and Group M Gag_dmyr-patent.doc

108/109

Fig. 22 cont'd-42

GTTGGTGATTTTGAAC TTTTGCTTTGCCACGGAACGGTCTGCGTTGT CGGGAAGATGCGTGATC
TGATCCTTCAACTCAGCAAAAGTTCGATTTATTCAACAAAGCCGCCGTCCCCTCAAGTCAGCGT
AATGCTCTGCCAGTGTTACAACCAATTAACCAATTCTGCGTTCAAAATGGTATGCGTTTTGACA
CATCCACTATATATCCGTGTCGTTCTGTCCACTCCTGAATCCCATTCAGAAATTCTCTAGCGA
TTCCAGAAGTTTCTCAGAGTCGGAAAGTTGACCAGACATTACGAACTGGCACAGATGGTCATAA
CCTGAAGGAAGATCTGATTGCTTAACTGCTTCAGTTAAGACCGACGCGCTCGTCGTATAACAGA
TGCGATGATGCAGACCAATCAACATGGCACCTGCCATTGCTACCTGTACAGTCAAGGATGGTAG
AAATGTTGTGCGGTCTTGCACACGAATATTACGCCATTTGCCTGCATATTCAAACAGCTCTTCT
ACGATAAGGGCACAAATCGCATCGTGGAACGTTTGGGCTTCTACCGATTTAGCAGTTTGATACA
CTTTCTCTAAGTATCCACCTGAATCATAAATCGGCAAAATAGAGAAAATTGACCATGTGTAAG
CGGCCAATCTGATTCCACCTGAGATGCATAATCTAGTAGAATCTCTTCGCTATCAAATTACT
TCCACCTTCCACTCACCGGTTGTCCATTGCTGCTGAACTCTGCTTCCTCTGTTGACATGACAC
ACATCATCTCAATATCCGAATACGGACCATCAGTCTGACGACCAAGAGAGCCATAAACACCAAT
AGCCTTAACATCATCCCCATATTTATCCAATATTGCTTCCTTAATTTTCATGAACAATCTTCATT
CTTTCTTCTCTAGTCATTATTATTGGTCCGTTTCATAACACCCCTTGTATTACTGTTTATGTAAG
CAGACAGTTTTATTGTTTCATGATGATATATTTTTATCTTGTGCAATGTAACATCAGAGATTTTG
AGACACAACGTGGCTTTCCCCGGCCCATGACCAAAATCCCTTAACGTGAGTTTTTCGTTCCACTG
AGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATC
TGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTTGTTTGCCGGATCAAGAGCTAC
CAACTCTTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCAGATACCAATACTGTTCTTCTAGT
GTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTA
ATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGAC
GATAGTTACCGGATAAGGCGCAGCGGTGCGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTT
GGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTT
CCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGA
GGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACT
TGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCG
GCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCTGCGTTATCCC
CTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGAGCCGAAC
GACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTT
ACGCATCTGTGCGGTATTTACACCGCATATGGTGCCTCTCAGTACAATCTGCTCTGATGCCG
CATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAA
AATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGG
CGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATAGCCGCGGCATCGATGATATCCATTGCA
TACGTTGTATCTATATCATAATATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGT
TGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCAT
ATATGGAGTTCGCGGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCC
CCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGA
CGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGC
CAAGTCCGCCCCCTATTGACGTCAATGACGTAATGGCCCCGCTGGCATTATGCCCAGTACAT
GACCTTACGGGACTTTCTTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTG
ATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTTGACTCACGGGGATTTCCAAGTC
TCCACCCCATTGACGTCAATGGGAGTTTGTGTTTGGCACCAAAATCAACGGGACTTTCCAAAATG
TCGTAACAACTCCGCCCCATTGACGCAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATA
AGCAGAGCTCGTTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCC
ATAGAAGACACCGGGACCGATCCAGCCTCCGCGGGCGCGCGTCGACGGGCGACGCGAAAATTGG
GCCCCTCGAGAGGGCGCGCCGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCA
TCTGTTGTTTTGCCCTCCCCCGTGCTTCCTTGACCCTGGAAGGTGCCACTCCCCTGTCTTTT

109/109

Fig. 22 cont'd-43

CCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGG
GGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGAATTT