

## Original article

# A new approach for sequencing virion genome of Chinese HIV-1 strains subtype B and BC from plasma

MENG Zhe-feng, ZHANG Xiao-yan, XIN Ruo-lei, XING Hui, HE Xiang, XU Jian-qing and SHAO Yi-ming

**Keywords:** *human immunodeficiency virus 1; RNA; full-genome; amplification*

**Background** Although it was widely accepted that full-length HIV genome sequences is important in studying virus genetic evolution and variation as well as developing vaccine candidate, to directly sequencing HIV-1 genome of virion RNA remains as a challenge worldwide. Up to date, no published genomic sequences from virion RNA are available for Chinese prevalent HIV-1 strains due to the absence of specialized protocol and appropriate lab equipments. In this study we developed a straightforward approach for amplifying and sequencing HIV virion RNA from plasma by modifying published protocols and further confirmed it is suitable to process Chinese samples.

**Methods** The methods for viral RNA extraction and gene amplification was modified and optimized as could be widely used in most Chinese labs. Gene alignment of Chinese HIV-1 strains was employed for designing specialized primer sets for Thai-B and BC recombinant strains. Based on comprehensively consideration of high variable gene region and recombinant breakpoints in BC recombinant strains, a three-amplicon strategy (including 4.3-kb *gag-pol*, 2.9-kb *pol-env* and 2.7-kb *env-nef*) was developed. In addition, one amplicon (9 kb near full-length genome) was also used in 32 samples with varied viral loads. All amplicons were directly sequenced by DNA automated sequencer.

**Results** Twenty-five percent (8/32) amplification efficiency was achieved by the one-amplicon strategy and 65.6% (21/32) by three-amplicon strategy. For one amplicon strategy, none of complete near full-length genome sequences was obtained by DNA sequencing. For three-amplicon strategy, 75% sequences were achieved in DNA sequencing. Amplification efficiency but not sequencing efficiency was closely associated with viral loads.

**Conclusion** Three-amplicon strategy covering all encoding regions of HIV-1 is suitable for Thai-B and BC recombinant strains and could be potentially employed in less-well equipped Chinese labs.

*Chin Med J 2011;124(2):304-308*

Up to now, the majority of HIV genome sequences were generated by a three-step approach, amplifying HIV sequence from peripheral blood mononuclear cells (PBMCs) DNA or cultured cell DNA, cloning the amplified fragment into plasmid and sequencing the cloned fragment.<sup>1-5</sup> Among 2223 sequences published in HIV database, only 109 sequences are identified to be derived from plasma and the majority of them were generated by cloning strategy; less than 10 sequences are explicitly generated from direct PCR amplification of virion RNA and subsequent sequencing.<sup>4</sup> So far, only 45 HIV-1 sequences >8.5 kb in length from Chinese dominantly circulated strains are identified in HIV database, none of them was directly derived from virion RNA,<sup>2,4,5</sup> which suggest that direct RT-PCR with virion RNA as the template and sequencing large fragment has been relatively challenging.

Currently, near full-length genome sequence has becoming more widely used for the analysis of HIV-1 gene variation and genetic evolution.<sup>6-9</sup> In China, HIV-1 infection and transmission remains to be one of the most challenging public health problems. Although most researchers realized the importance to generate HIV-1 near full-length genome sequences, the absence of specialized protocol for Chinese prevalence strains and appropriate lab equipments have restrained for further

exploration of genetic evolution and recombinant of circulating HIV-1 strains. In recent years several international protocols have been developed to generate near full-length HIV-1 RNA genomes directly from plasma by using prolonged RT-PCR.<sup>9,10</sup> However, those protocols may be confined to certain HIV-1 prevalent

DOI: 10.3760/cma.j.issn.0366-6999.2011.02.028

Shanghai Public Health Clinical Center, Key Laboratory of Medical Molecular Virology of Ministry of Health, Shanghai Medical College, Institutes of Biomedical Sciences, Fudan University, Shanghai 200433, China (Meng ZF, Zhang XY and Xu JQ)

Key Laboratory of Infectious Disease Prevention and Control, National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 100050, China (Zhang XY, Xin RL, Xing H, He X, Xu JQ and Shao YM)

Correspondence to: Dr. XU Jian-qing, Shanghai Public Health Clinical Center, Key Laboratory of Medical Molecular Virology, Shanghai Medical College, Institutes of Biomedical Sciences, Fudan University, Shanghai 200433, China (Tel: 86-21-37990333 ext. 7335. Fax: 86-21-57247094. Email: jianqingxu2008@gmail.com); Dr. SHAO Yi-ming, Key Laboratory of Infectious Disease Prevention and Control, National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 100050, China (Tel: 86-10-58900981. Fax: 86-10-63154638. Email: yshao@bnn.cn)

This study was supported by the grants from the Natural Science Foundation of China (No. 30901257), the Chinese National Grand Program on Major New Drugs (No. 2009ZX09303-006), the China Postdoctoral Science Foundation (No. 20090460585), and the Shanghai Postdoctoral Program (No. 10R21410600).

**Table 1.** Primers for the amplification of nearly full-length HIV-1 genome

| Primers     | Position  | Sequences (5'→3')              | Length (kb) |
|-------------|-----------|--------------------------------|-------------|
| 1.5 (+)     | 552-581   | AGTAGTGTGTGCCCGTCTGTTGTGTGACTC | 30          |
| POLMR1 (-)  | 4983-4956 | GTATTACTACTGCCCTTCACCTTCCA     | 28          |
| MSF12 (+)   | 623-649   | AAATCTCTAGCAGTGGCGCCGCAACAG    | 27          |
| POLMR2 (-)  | 4925-4897 | GTCTCTGCTGTCTGTGAATAAACCCGAA   | 29          |
| POLMF3 (+)  | 4618-4643 | CTACAATCCCCAAAGTCAGGGAGTAG     | 26          |
| ED12MR1 (-) | 7810-7785 | GTGCTTCCTGCTRCTCCAAGAACCC      | 26          |
| POLMF4 (+)  | 4743-4770 | CTTAAGACAGCAGTACAAATGGCAGTAT   | 28          |
| ED8MR2 (-)  | 7668-7641 | CACCTTCTCCAATTGTCCCTCATATCTCC  | 28          |
| ED5MF3 (+)  | 6557-6582 | ATGGGATCAAAGCCTAAAGCCATGTG     | 26          |
| 1.3MR (-)   | 9643-9618 | CACTACTGAAGCACTCAAGGCAAGC      | 26          |
| E7MF4 (+)   | 6996-7023 | CAACTACTGTAAATGGTAGCCTGGCAG    | 28          |
| MSR12 (-)   | 9606-9635 | GAAGCACTCAAGGCAAGCTTTATTGAGGCT | 30          |

“+”: Sense primer. “-”: Antisense primer. Positions according to HXB2 numbering system.

strains and thereby unable to work for Chinese circulating strains. The primer set for 9-kb amplicon strategy developed by Rousseau et al was more suitable for subtype C in the Africa according to the HIV-1 gene alignment; the three-amplicon strategy is not reconciled with the recombination pattern of BC recombinant forms, because its primer binding sites lies on the breakpoints of BC recombinant strains. Therefore, taking the genetic properties of Chinese prevalence strains into consideration, we developed a straightforward approach for amplifying and sequencing HIV virion RNA from plasma by modifying published protocols and further confirmed it is suitable to process Chinese samples in current study.

## METHODS

### Study subjects

Study subjects were recruited from commercial blood donors and intravenous drug users in Anhui province and Sichuan province of China, respectively. All individuals were antiretroviral therapy naïve and signed informed consent to participate in the study. Ten milliliters of blood was drawn from each subject into an EDTA pre-filled tube, the plasma was separated and stored at  $-80^{\circ}\text{C}$  until use. Samples were selected based on the viral loads (VL), ranging from 5000 copies/ml to over 120 000 copies/ml. VL was measured for each study subject with the Amplicor Version 1.5 assay (Roche, Durham, NC, USA).

### RNA extraction

The spin column from QIAmp DNA Blood Midi Kit (QIAGEN, Germany) and reagents from the QIAmp viral RNA Mini kit was combined for viral RNA extraction.<sup>9</sup> The ratio of plasma volume, AVL buffer and wash buffers was used as instructed in the protocol of QIAmp viral RNA Mini kit for QIAmp DNA Blood Midi Kit column. HIV-1 positive plasma specimen 1 ml was used for RNA extraction and the final elution volume was 120  $\mu\text{l}$ . Extracted RNA was either reverse transcribed into cDNA immediately or stored at  $-80^{\circ}\text{C}$  until use.

### Reverse transcription

SuperScript™ III First-Strand Synthesis Supermix (Invitrogen, USA) was used for cDNA synthesis.<sup>9-12</sup> The

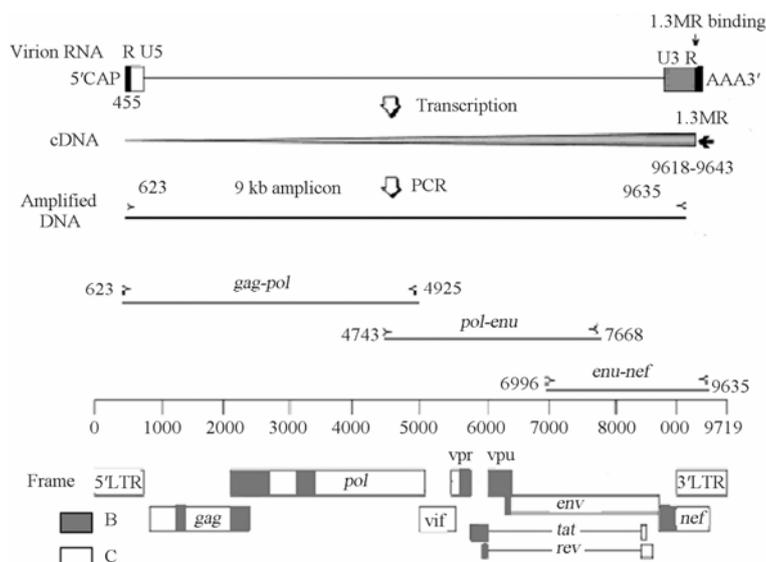
mixture was heated to  $65^{\circ}\text{C}$  for 5 minutes and rapidly cooled to  $0^{\circ}\text{C}$ . Reverse transcription reaction was performed as previously described.<sup>9,10,13-15</sup> Synthesized cDNA was stored at  $-20^{\circ}\text{C}$ .

### PCR amplification

The Expand High Fidelity PCR System (Roche Molecular Biochemicals, Mannheim, Germany) was used to amplify HIV-1 virion genome from the synthesized cDNA.<sup>5,9,10</sup> For the primer design (amplification primers are listed in Table 1), the most conserved regions were chosen as the primer binding sites using the full-length genome alignment of group M HIV-1 diverse strains withdrawn from the HIV sequence database (Figure).<sup>2</sup> For example, 7810-7785 and 6557-6582 (HXB<sub>2</sub> nts) in *env* were conserved for almost all M group HIV-1 strains in addition to the primer binding sites in *gag* and *pol*. Moreover, the amplicon length was designed to be associated with conservative or variation of targeted gene, the largest *gag-pol* amplicon represents the most conserved gene region in HIV-1; the high variable *env* gene region was split into two fragments for amplification (*pol-env* and *env-nef*). The selected sets of primers were further examined to match the Chinese prevalent strains Thai-B and BC recombinant strains.<sup>2,5,16-18</sup> Specially, recombinant pattern of BC recombinant forms were taken into consideration as another important factor in designing primers. The primers binding sites were designed to avoid the recombinant breakpoints of BC recombinant strains (Figure). In addition, degenerate bases within the primers were also used to cover the genetic diversity present at the primer binding sites.

For one amplicon strategy, the primers were designed to obtain 9.0 kb near full-length sequences as follows: primer 1.5(+) and 1.3MR as outer primer, primer MSF12 and MSR12 as inner primer. The amplification reaction was performed as previously described.<sup>2,5</sup>

Three-amplicon strategy was performed as following. The first amplicon was the *gag-pol* amplicon (nts 623-4925 HXB<sub>2</sub>), the second spans *pol* to *env* (nts 4743-7668 HXB<sub>2</sub>) and the third from *env* to *nef* (nts 6996-9635 HXB<sub>2</sub>) (amplicon described as Figure and primers are listed in Table 1). The cycling conditions for these three



**Figure.** Nearly full-length RT-PCR design. Viral RNA was reverse-transcribed to cDNA with primer 1.3MR using SuperScript™ III RNase H2 RT. Two different strategies were employed to amplify the nearly full-length genome: one amplified 9.0-kb, and the other amplified three overlapping fragments of 4.3-kb (*gag-pol*), 2.9-kb (*pol-env*) and 2.7-kb (*env-nef*), respectively.

regions were identical to the described one for 9 kb amplicon, except the incubation time at 68°C in each cycle was 5 minutes or 3 minutes rather than 10 minutes. A hot-start using DynaWax (Finnzymes, Espoo, Finland) was employed in each PCR, with two layers separated by wax. In the bottom layer were the dNTPs and template cDNA while the top layer contained the PCR buffer, DNA polymerase and primers.

### Sequencing

Eight samples achieved in success of 4 amplicons (including 9 kb amplicon and three amplicons simultaneously) were selected to perform sequencing reaction using a primer-walking approach.<sup>2,5</sup> To ensure the single template used for sequencing reaction, 4 dilutions (1:5, 1:10, 1:20, 1:50) of synthesized cDNA was amplified and the resulted positive PCR products with the highest dilution were used in direct automated fluorescent sequencing (BigDyeH Terminator v3.1 Cycle Sequencing Assay, Applied Biosystems, USA) in an automated 3730xl sequencer (Applied Biosystems). The nearly full-length genome sequence was assembled by overlapping the sequences of the three amplicons and merging them into one sequence as long as the two overlapping sequences were greater than 99% homologous.<sup>19,20</sup>

### Amplification and sequencing of C2V3 of *env*

The synthesized cDNAs from 22 samples were amplified using primer E7 (5'-CTGTTAAATGGCAGTCTAGC-3') and E8 (5'-CACTTCTCCAATTGTCCCTCA-3') and purified PCR products were used for sequencing *env* C2V3 region. In order to identify the subtype or CRF, phylogenetic analysis were performed for the generated sequences as described previously.<sup>2,5,19-22</sup>

### Statistical analysis

Chi-square test was used for the statistical analysis with SPSS 11.0 (USA). *P* value less than 0.05 was considered statistically significant.

## RESULTS

### Efficacy of PCR amplification

The amplification efficacies for four different fragments obtained from 32 samples are shown in Table 2. For the 9-kb amplification, 8 positive PCR reactions were obtained from 32 samples. For the three-amplicons strategy, 65.6% success (21/32) was achieved in all PCR reactions. In details, the positive amplifications of *gag-pol* fragment were 75% (24/32), 65.6% (21/32) for *pol-env* and 68.75% (22/32) for *env-nef*, respectively. The significant difference of PCR efficiency between 9-kb and three-amplicon strategy was detected by chi-squared test ( $P=0.003$ ).

**Table 2.** Amplification efficacy of HIV-1 RNA fragments Sensitivities of by different detections with varied HIV-1 viral loads in plasma samples

| PCR regions                | Viral load (copies/ml)   |                          |
|----------------------------|--------------------------|--------------------------|
|                            | >80 000 (PCR efficiency) | <80 000 (PCR efficiency) |
| Thai-B                     |                          |                          |
| Sample number ( <i>n</i> ) | 3                        | 7                        |
| <i>gag-nef</i> (9 kb)      | 1/3                      | 1/7                      |
| <i>gag-pol</i> (4.3 kb)    | 3/3                      | 3/7                      |
| <i>pol-env</i> (2.9 kb)    | 3/3                      | 3/7                      |
| <i>env-nef</i> (2.7 kb)    | 3/3                      | 3/7                      |
| BC recombinant strain      |                          |                          |
| Sample number ( <i>n</i> ) | 10                       | 12                       |
| <i>gag-nef</i> (9 kb)      | 4/10                     | 2/12                     |
| <i>gag-pol</i> (4.3 kb)    | 10/10                    | 7/12                     |
| <i>pol-env</i> (2.9 kb)    | 10/10                    | 5/12                     |
| <i>env-nef</i> (2.7 kb)    | 10/10                    | 6/12                     |

No significant difference between 10 Thai-B and 22 BC recombinant strains was observed in the efficacy of long RT-PCR. Three amplicons for Thai-B was 60% (6/10) and for BC recombinant strains was 68% (15/22). Positive 9 kb amplicons for Thai-B and for BC recombinant strains were 20% (2/10) and 27% (6/22), respectively.

### Sensitivities for successful PCR amplification and sequencing among different groups of HIV-1 VL

For both Thai-B and BC recombinant strains, the positive

amplifications of four different fragments obtained in the 32 samples were closely associated with VL (Table 2). The success for generating all three amplicons were 100% in samples with a VL greater than 80 000 copies/ml, and around 42% in samples with VL less than 80 000 copies/ml ( $P=0.0015$ ). The amplification from virion RNA failed in the sample with VL less than 20 000 copies/ml. However, for the amplicons present in strong bands, it remains as a challenge to sequence the nested PCR products even VL is high.

For the three-amplicon strategy, 6 full-length sequences of 8 samples were achieved, another two failed in *pol-env* fragment sequencing. Four sequences were derived from high viral load samples (VL greater than 120 000 copies/ml), 2 sequences with VL less than 80 000. Amplicons from other two high VL samples resulted in unavailable chromatography pictures in the sequencing reaction. The resulting sequences were around 9 kb (nts 623–9635 HXB<sub>2</sub>) and contained the complete HIV-1 coding region. All of the reading frames were open without apparent insertion, deletion, or rearrangement.

For the 9-kb PCR products, none of complete and clear full-length sequences was achieved in sequencing chromatography pictures; most assembled fragments were less than 7.5 kb and failed in sequencing *env* gene.

## DISCUSSION

Thai-B, BC recombinant strain and CRF01\_AE are currently the dominant circulating strains of HIV-1 in China; the former two accounts for more than 70% infection in population.<sup>23-28</sup> The unique constitution of HIV-1 prevalence strains in China and different genetic characteristics of viral gene may disable the published international protocols for generating full-length genome. Therefore, specialized primers for Chinese prevalence strains are prerequisite for success to generate near full-length genome from plasma. The relatively close evolution relationship between Thai-B and BC recombinant strain render it possible to design a set of primers suitable for both subtypes strains; indeed, we designed a set of primers and successfully tested them in the present study. Although the primers were almost conserved among the diversity strains for entire HIV-1 M group, the one amplicon strategy (9 kb amplicon) was proved unable to reconcile with directly sequencing virion RNA. On the contrary, the three-amplicon strategy specialized for particular strains showed high efficiency in amplifying and sequencing.

The primers listed in Table 1 were designed by using the full-length genome alignment of group M HIV-1 diverse strains from the HIV sequence database and particularly matched the genetic variation characteristics of Thai-B and BC recombinant strains currently circulating in China. Moreover, not only the most conserved regions but also no breakpoints were considered as the binding sites for

primers.<sup>2,5</sup> As described in Figure and Table 1, there exists no breakpoint within 500 bp at the binding sites of primers, which helps to avoid PCR recombination and unsuccessful assembled sequences. Through this rational design, a similar efficacy of long RT-PCR was achieved for Thai-B and BC recombinant strains.

The gene variation would undoubtedly affect the efficacy of prolonged RT-PCR. The presented strategy of three amplicons was designed to tackle with this issue and to ensure a similar efficacy of different viral gene amplification. Conserved *gag-pol* amplicon was the longest fragment (4.3 kb), while a high variable gene region (such as *env*) was separately amplified with 2.7 kb and 2.9 kb. As described in Table 2, three amplicons showed similar efficiency for RT-PCR and sequencing.

The optimization of extraction of RNA and cDNA synthesis was also essential for the accurately direct sequencing of full-length HIV RNA. Considering ultracentrifugation is unavailable in most Chinese labs, RNA extraction approach recommended by Rousseau et al<sup>9</sup> was explored. In fact, most HIV-1 infected individuals have a viral RNA load above 20 000 copies/ml in plasma if untreated or if they are on a non-suppressive antiretroviral therapy,<sup>6</sup> the presented method here without ultracentrifugation in this study could be applicable for a number of Chinese labs and was able to obtain viral genome directly from plasma with relatively high efficacy. For cDNA synthesis, priming with Oligo dT and 1.3MR resulted in much better yields for synthesizing full-length cDNA than random hexamer, and whereas 1.3MR tended to generate more consistent results than Oligo dT. Therefore, 1.3MR was used for all reverse transcription reactions in our experiments.

The accuracy of direct sequencing of full-length virion RNA from HIV-1 relies on the abundant and homogeneous DNA copies. For the amplicons present in strong bands, the quality of chromatography picture in sequencing is independent on viral loads. Multiple sequencing templates due to the presence of quasispecies viruses in one fluorescent sequencing reaction were the major challenge. Quasispecies viruses with equal or unequal high virus loads, which showed strong band in RT-PCR, could result in an obviously overlapping peak in chromatography pictures. The highest diluted cDNA was used in amplification, which could partially decrease the overlapping peak in chromatography picture. However, under certain circumstance, even with series dilution, it remains impossible to sequence amplicons directly from virion RNA, which may suggest that the high diversified quasispecies *in vivo* existed.

In summary, the modified and optimized protocol in this study was primarily proved to be suitable for amplifying and sequencing nearly full-length genome of HIV-1 Thai-B and BC recombinant strains. In addition, the three-amplicon strategy without ultracentrifugation

renders this protocol feasible in laboratory of molecular epidemiology for HIV/AIDS.

#### REFERENCES

- Salminen MO, Koch C, Sanders-Buell E, Ehrenberg PK, Michael NL, Carr JK, et al. Recovery of virtually full-length HIV-1 provirus of diverse subtypes from primary virus cultures using the polymerase chain reaction. *Virology* 1995; 213: 80-86.
- Meng Z, Xing H, He X, Ma L, Xu W, Shao Y. Genetic characterization of three newly isolated CRF07\_BC near full-length genomes in China. *AIDS Res Hum Retrovir* 2007; 23: 1045-1050.
- Ratner L, Haseltine W, Patarca R, Livak KJ, Starcich B, Josephs SF, et al. Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature* 1985; 313: 277-284.
- HIV sequence database. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, 2001. <http://www.hiv.lanl.gov>.
- Meng Z, He X, Xing H, Xin R, Sun J, Yi F, et al. Construction and Characterization of an Infectious Molecular Clone of HIV-1 CRF07\_BC. *AIDS Res Hum Retrovir* 2008; 24: 259-264.
- Mellors JW, Kingsley LA, Rinaldo CR Jr, Todd JA, Hoo BS, Kokka RP, et al. Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion. *Ann Intern Med* 1995; 122: 573-579.
- Kousiappa I, Van De Vijver DA, Kostrikis LG. Near full-length genetic analysis of HIV sequences derived from Cyprus: evidence of a highly polyphyletic and evolving infection. *AIDS Res Hum Retrovir* 2009; 25: 727-740.
- Dash PK, Siddappa NB, Mangaiarkarasi A, Mahendarkar AV, Roshan P, Anand KK, et al. Exceptional molecular and coreceptor-requirement properties of molecular clones isolated from an human immunodeficiency virus type-1 subtype C infection. *Retrovirology* 2008; 5: 25.
- Rousseau CM, Birditt BA, McKay AR, Stoddard JN, Lee TC, McLaughlin S, et al. Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes. *J Virol Methods* 2006; 136: 118-125.
- Nadai Y, Eyzaguirre LM, Constantine NT, Sill AM, Cleghorn F, Blattner WA, et al. Protocol for nearly full-length sequencing of HIV-1 RNA from plasma. *PloS one* 2008; 3: e1420-1425.
- Villahermosa ML, Thomson M, Vázquez de Parga E, Cuevas MT, Contreras G, Pérez-Alvarez L, et al. Improved conditions for extraction and amplification of human immunodeficiency virus type 1 RNA from plasma samples with low viral load. *J Hum Virol* 2000; 3: 27-34.
- Klarman GJ, Schaub CA, Preston BD. Template-directed pausing of DNA synthesis by HIV-1 reverse transcriptase during polymerization of HIV-1 sequences *in vitro*. *J Biol Chem* 1993; 268: 9793-9802.
- Harrison GP, Mayo MS, Hunter E, Lever AM. Pausing of reverse transcriptase on retroviral RNA templates is influenced by secondary structures both 5' and 3' of the catalytic site. *Nucleic Acids Res* 1998; 26: 3433-3442.
- Hawkins PR, Jin P, Fu GK. Full-length cDNA synthesis for long-distance RT-PCR of large mRNA transcripts. *Biotechniques* 2003; 34: 768-770, 772-763.
- Das M, Harvey I, Chu LL, Sinha M, Pelletier J. Full-length cDNAs: more than just reaching the ends. *Physiol Genomics* 2001; 6: 57-80.
- Rozak DA, Bryan PN. Offset recombinant PCR: a simple but effective method for shuffling compact heterologous domains. *Nucleic Acids Res* 2005; 33: e82-e85.
- Malet I, Belnard M, Agut H, Cahour A. From RNA to quasispecies: a DNA polymerase with proofreading activity is highly recommended for accurate assessment of viral diversity. *J Virol Methods* 2003; 109: 161-170.
- Robertson DL, Sharp PM, McCutchan FE, Hahn BH. Recombination in HIV-1. *Nature* 1995; 374: 124-126.
- Primalign HIV sequence database tool. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, 2006. <http://hiv-web.lanl.gov/content/hiv-db/PRIMALIGN/PRIME.html>.
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; 22: 4673-4680.
- Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987; 4: 406-425.
- Ray SC. Simplot for Windows, version 2.5. Johns Hopkins Medical Institutions, Baltimore, USA, 1999. <http://www.welch.jhu.edu/~sray/download>.
- McCutchan FE. Global epidemiology of HIV. *J Med Virol* 2006; 78 Suppl 1: s7-s12.
- Wang JJ, Reilly KH, Luo J, Zang CP, Wang N. Dynamic mathematical models of HIV/AIDS transmission in China. *Chin Med J* 2010; 123: 2120-2127.
- Sheng L, Cao WK. HIV/AIDS epidemiology and prevention in China. *Chin Med J* 2008; 121: 1230-1236.
- Xing H, Chen Z, Liang H. The evolution and rapid spread of B'/C recombinant HIV-1 strains in western China. XIV International AIDS Congress, Barcelona, Spain, 2002.
- Cui Y, Liao A, Wu ZY. An overview of the history of epidemic of and response to HIV/AIDS in China: achievements and challenges. *Chin Med J* 2009; 122: 2251-2257.
- Qiu C, Xu JQ. HIV-1/AIDS vaccine development: are we in the darkness. *Chin Med J* 2008; 121: 939-945.

(Received July 9, 2010)

Edited by SUN Jing