

トピックス

HIV 感染症統合データベースの開発

仲宗根 正¹⁾, 原 敬志^{1),2)}, 染谷 健二^{1),2)}, 池尾 一穂³⁾,
五條堀 孝³⁾, 山本 直樹¹⁾, 本多 三男¹⁾

¹⁾ 国立感染症研究所/エイズ研究センター

²⁾ 元科学技術振興機構

³⁾ 国立遺伝学研究所/生命情報・DDBJ 研究センター

目的: エイズワクチン開発に重要な情報をデータベース(DB)化して公開(一部制限)し、HIV 関連研究者に活用もらうことを目的として HIV 感染症統合データベースを構築しました (URL <https://aids.nih.go.jp>)。

概要: 本 DB では、1988 年より解析中の日本およびタイ国 HIV 感染者からのウイルス分離結果、遺伝子配列、蛋白構造情報などのウイルス遺伝子生物学的情報に加えて、臨床データを時系列的に管理・検索可能な統合 DB を構築し WEB 上で提供します。

内容: DB は、平成 15 年 4 月末現在、DDBJ の HIV 遺伝子 DB に加えて、提供 HIV 感染者 572、ウイルス分離解析数 3217、C2V3 遺伝子解析数 361、V3 部蛋白構造解析数 269 (PDB 形式)、対応臨床データ (生年、性別、CD4 細胞数、ウイルス量、薬剤履歴、その他) からなります。主機能は、DDBJ の HIV 遺伝子データに対する遺伝子相同性検索、遺伝子系統樹解析、genosubtyping、独自の division 作成機能、V3 部蛋白 3 次元構造の閲覧機能、臨床データ検索機能があります。

展望: この統合化された DB の有効活用によりエイズワクチン・薬剤開発の進展が期待できます。

キーワード: HIV 感染症、データベース、HIV 遺伝子、HIV 蛋白 3 次元構造

日本エイズ学会誌 6 : 42-49, 2004

はじめに

国立感染症研究所・エイズ研究センターでは、1988 年から主に日本の HIV-1 についてウイルス遺伝子生物学的検査を行っています。その目的は、日本に蔓延している HIV-1 の特徴を把握し、予防治療法開発に役立てる事であり、旧厚生省や科学技術振興機構、エイズ予防財団のエイズ対策研究の一環として行われてきました。検体数は平成 14 年 3 月末現在 3,600 余に達し、解析された結果は全て電子化され保管されています。得られた成果は論文¹⁻⁹⁾、学会あるいは研究会等で逐一報告されてきましたが、今回、過去 14 年に渡り蓄積してきたデータがさらに幅広く活用されるよう、HIV 感染症統合データベースを開発し、その公開を行いました。

データベース概要

1988 年より解析中の日本およびタイ国 HIV 感染者からのウイルス分離結果、遺伝子配列、蛋白構造情報などのウイルス遺伝子生物学的検査結果に加えて対応する臨床データ

著者連絡先：仲宗根正（〒162-8640 東京都新宿区戸山 1-23-1
国立感染症研究所・エイズ研究センター）
Fax : 03-5285-1183, E-mail : nakabone@nih.go.jp

2003 年 5 月 16 日受付；2003 年 8 月 15 日受理

を時系列的に管理・検索可能となる統合データベース(DB)を構築し WEB 上で提供します。

本システムではユーザに対しインターネットを介してデータ公開を行うため、HTTP ベースで WWW の機構に基づく形式でデータを配信する、いわゆる WEB サーバと WEB ブラウザのクライアント-サーバ形式でシステム構築を行っています。

WEB サーバとしては Apache を用い、通常の設定ではインターネット上を流れるデータは平文であるため、これを暗号化するために公開鍵方式の暗号化を行う SSL のモジュールを Apache に追加しました。また、PostgreSQL を用いたリレーショナル DB として、臨床 DB を構築し、各種臨床データの登録・検索を行えるように設計しました。

開発言語として PHP を用い、各言語での画面表示、DB 接続、ユーザ管理、遺伝子解析プログラム (C 言語) との連携等を実行するようプログラムし、ユーザ (WEB ブラウザ) から URL を指定することにより、前記の Apache 上で実行される各機能を実現しています。

具体的には次のような HIV 感染症統合データベースが構築されて、平成 14 年 10 月 10 日に公開(一部制限)されました(図 1)。

主機能

- DDBJ の HIV 遺伝子データに対する遺伝子相同性検

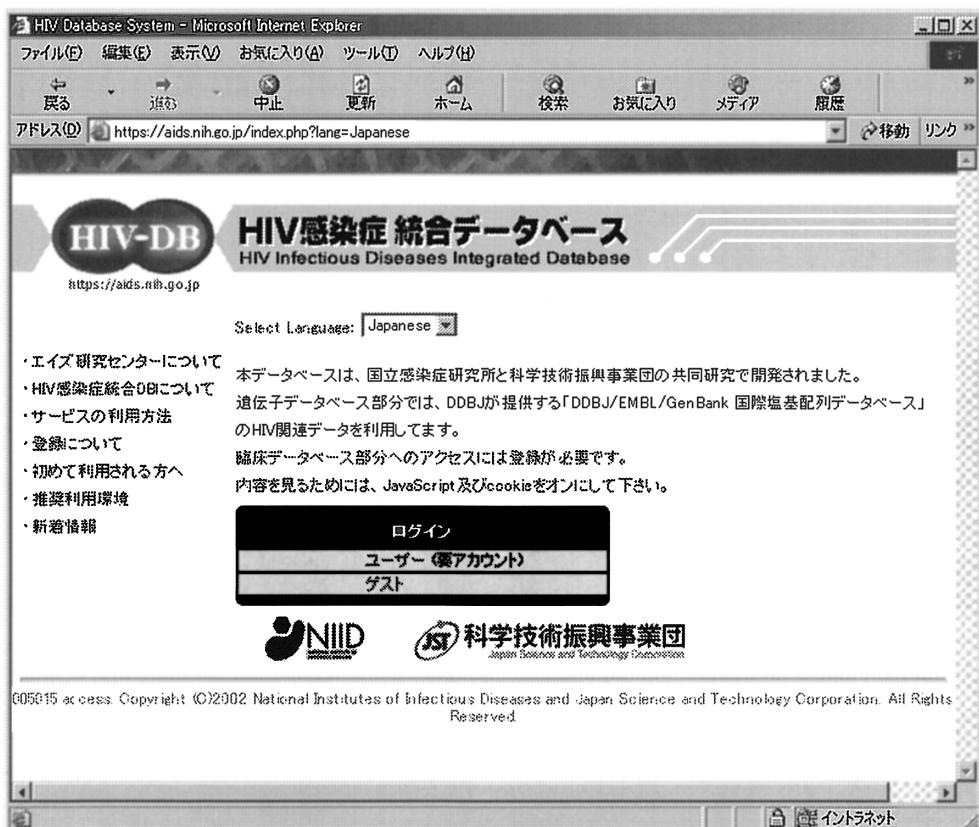


図 1 HIV 感染症統合データベースのトップページ : <https://aids.nih.go.jp>。この画面で使用言語を選択し、ログインします。

索 : DDBJ (DNA Data Bank of Japan) の HIV 遺伝子データに対して新規目的遺伝子の相同性を検索します。目的遺伝子が HIV 遺伝子のどの部分に該当するのか、あるいは過去に報告されている遺伝子のどれに最も近いのか追及できます (図 2, 3)。

2) 遺伝子系統樹解析 : 選択した遺伝子データ群について各々の類縁関係を系統樹形式で解析します。目的遺伝子がどの HIV から派生しているか等が解明でき、例えば HIV の地域的伝播経路を推定するのに役立ちます。

3) HIV-1 遺伝子サブタイピング : 現在 20 数種ある HIV のサブタイプ、サブサブタイプ、組換え体の中で、目的遺伝子がどこに属するのかタイピングします。どのタイプのウイルスが流入あるいは流行しているか判定できます (図 4)。

4) ユーザ独自の division 作成機能* : 既存あるいは新規遺伝子データの中から必要なデータ群を選定抽出する事ができます。これにより詳細なウイルス進化系統樹解析が可能になります。

5) ウィルス蛋白 3 次元構造の閲覧機能 (要 plug in 設定) : 中和抗体の主標的である HIV の表面蛋白 V3 部分についてその 3 次元構造を分子力学的手法により解析しました。その構造をブラウザ上で閲覧できます。閲覧にはユーザ側で、

3D viewer のブラウザへの plug in 設定が必要です (図 5)。

6) 臨床データ検索機能* : サンプル採取時の感染者の臨床データを検索できます。さらに臨床データ側から対応する遺伝子データ、蛋白構造データやウイルス分離データを検索できます。この機能により、目的遺伝子あるいは蛋白構造が感染者のどの病期、どのような免疫状態 (CD4 細胞数が目安となる) あるいはどのような治療状態で出現したのか追究できます (図 6)。

主内容 (数字は平成 15 年 4 月時点)

a) DDBJ の HIV 遺伝子データ (82,447 例) : DDBJ が年 4 回リリースしている「DDBJ/EMBL/GenBank 国際塩基配列データベース」中の HIV 関連遺伝子データを抽出加工して搭載しています。

b) ウィルス分離解析データ* (3217 例) : 国立感染症研究所・エイズ研究センターで 1988 年から解析蓄積してきたデータ。感染者血中の感染性ウイルス力値を半定量的に表しています。一般的に、感染者の病態進行に伴い力値が上昇するため、感染性ウイルス力値は病態の指標 (言い換えればウイルス活動性の指標) となり得ます。しかしながら測定の煩雑さ、非迅速性、低感度といった理由から、現在では遺伝子工学的測定法による血漿ウイルス量のほうが

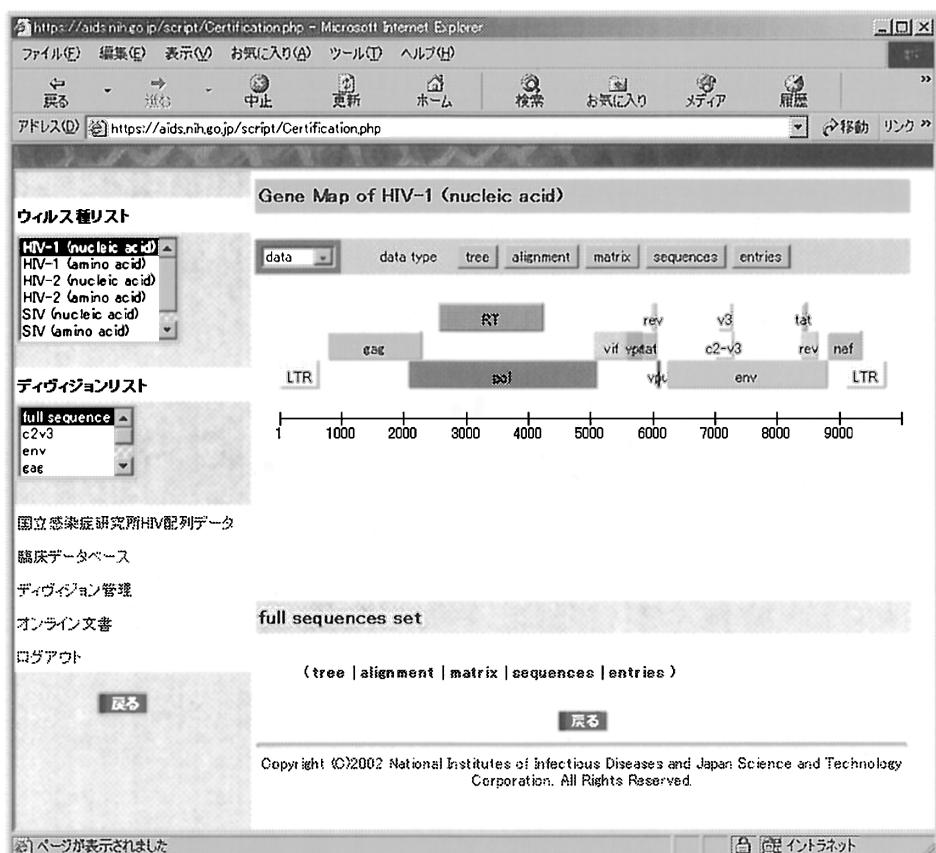


図 2 HIV 遺伝子マップ表示画面。各箇所をクリックすることにより同箇所の全ての HIV 遺伝子情報がダウンロードされ表示されます。

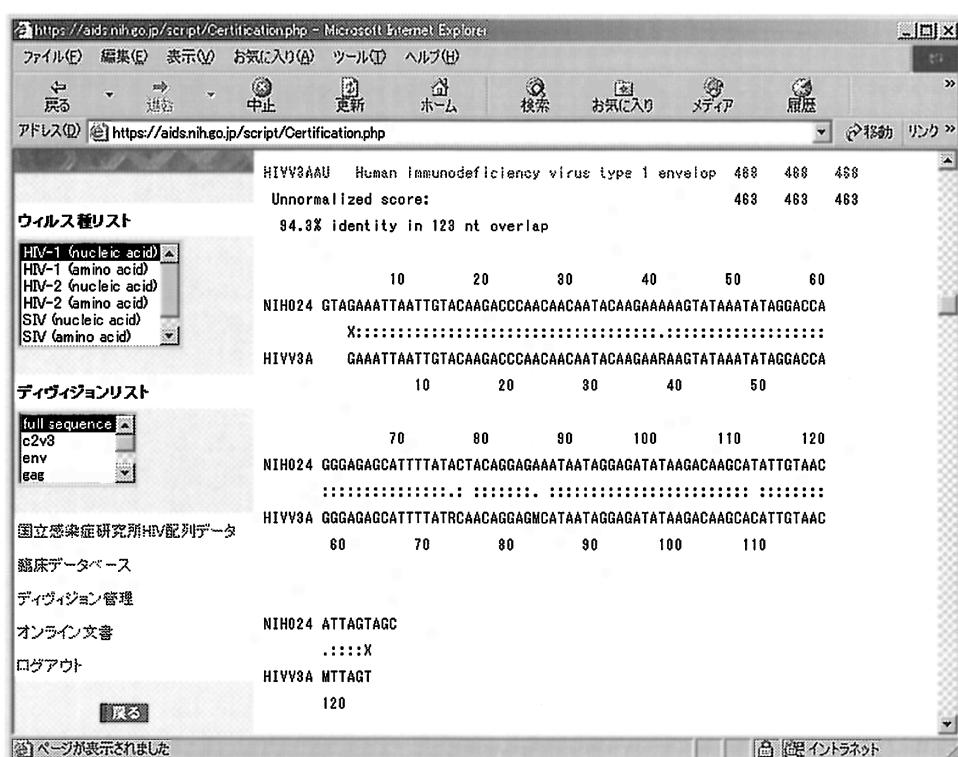


図 3 遺伝子相同性検索結果画面。目的配列に対して相同性の高い配列が検索され、その結果がアライメント表示されます。

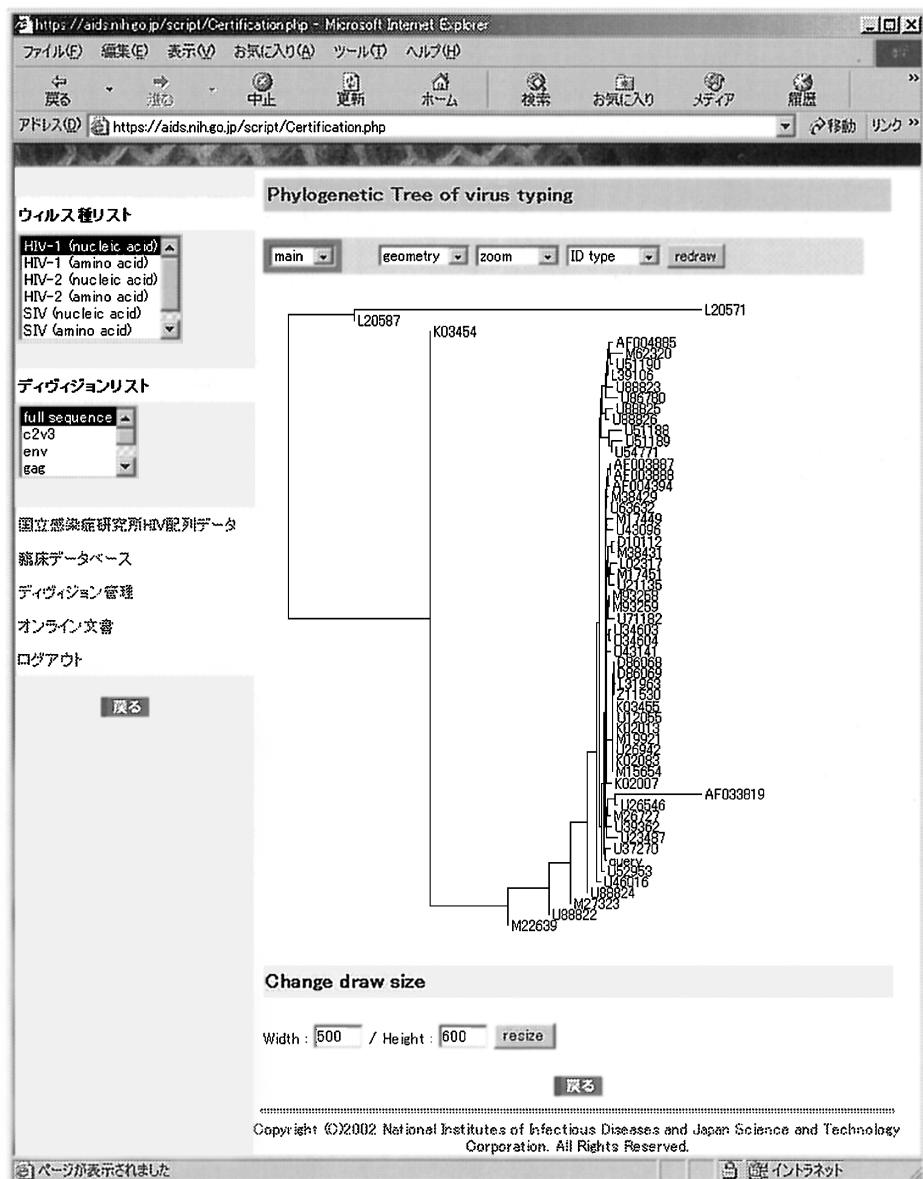


図 4 HIV 遺伝子サブタイピング結果画面。遺伝子系統樹解析により、目的配列のサブタイプが決定されます。

標準的に使用されています。一方で、生物学的活性を表しているという点では本法の方が優れていると考えられています。

c) C2V3 遺伝子解析データ（361 例）：ウイルス表面蛋白の中で、ウイルスの中和に最も重要な部分の DNA 配列データ。約 300 対の塩基配列（一部データは V3 部のみの 120 対）からなります。同部は宿主免疫により攻撃を受けやすく、それから逃れるべく容易に変異するため多様性に富みます。その多様性の限界を見極める事はワクチン開発に大きく貢献すると考えられています。同部はまたウイルスの進化を解析するにも絶好の箇所とされており、事実、

同部の DNA 配列の違いにより遺伝子学的サブタイピングを含めた分子進化学的解析がなされています。

d) V3 部蛋白構造解析データ（269 例）：まず、前項 C2V3 部遺伝子データより V3 部アミノ酸配列を演繹します。演繹されたアミノ酸配列と NMR 構造解析による既知 V3 部蛋白構造データを元に、蛋白構造解析ソフトウェア (MOE) によりその 3 次元構造を分子力学的に計算し PDB 形式に変換し搭載しました。先に述べたように、同部は中和抗体の標的あるいは分子進化学的解析の絶好の標的です。その分子力学的構造解析は、中和抗体のデザインや構造学的多様性解析に役立ちます。

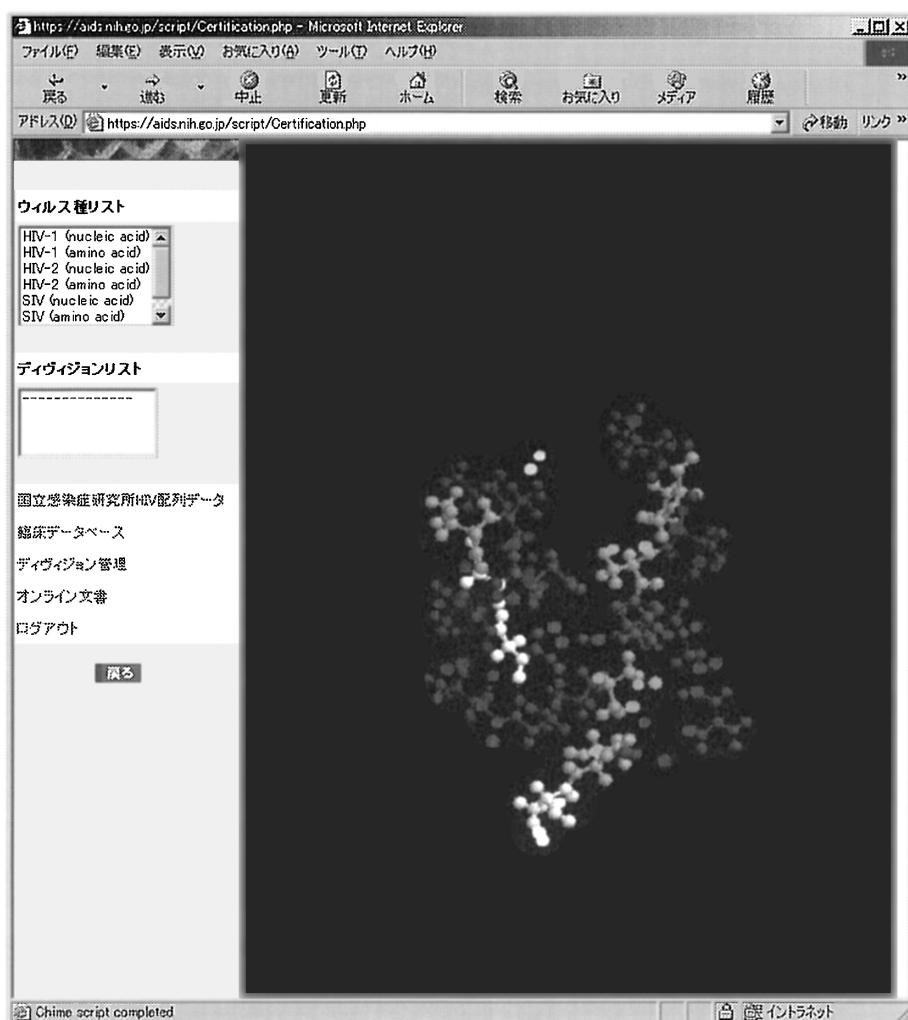


図 5 HIV-V3 部蛋白構造表示画面。計算科学の手法により予測された HIV-V3 loop 部分のウイルス蛋白構造を閲覧できます。ただし、ユーザ側で plug-in 設定が必要です。または単体の蛋白構造閲覧ソフト（PDB 形式ファイル対応）が必要です。

e) 対応臨床データ*：搭載している臨床データは次の通りです。生年、性別、推定出生地〔国別〕、推定居居住地〔国別、日本についてはブロック別表示〕、感染経路、病期、抗体検査結果〔ウェスタンプロット法〕、CD4 細胞数、血漿ウイルス量、薬剤履歴、抗ウイルス剤耐性。(オリジナルサンプル提供施設数 72、提供 HIV 感染者数 572、提供者の出身国：日本、タイ、ミャンマー、ベトナム、ルワンダ、ケニア、ブラジル)

* アカウントユーザのみに制限公開

使用言語

英語（デフォルト）、日本語、タイ語（一部）

利用法

利用法を WEB 上で閲覧参照できるよう設定しています。

なお、本 DB のデータは、旧厚生省「HIV 感染者発症予防・治療に関する研究班（1989–1997）」、旧厚生省「HIV 感染症の疫学研究班・母子感染に関する研究グループ（1997–2000）」、科学技術振興事業団国際共同研究事業「クレイド E 型エイズワクチン研究プロジェクト（1998–2003）」¹⁰⁾により収集・解析されたものです。

サンプルの採取にあたっては、2000 年度までは採取時の旧厚生省および旧科学技術庁の指針等に、2001 年度以降は、「ヒトゲノム・遺伝子解析研究に関する倫理指針」¹¹⁾に従い説明を行い同意が得られています。

本 DB の公開（一部制限）については、国立感染症研究所・医学研究倫理審査委員会により平成 14 年 6 月に承認されました。

図 6 臨床検査データ検索画面：各種キーワードによる検索が可能です。

おわりに

エイズワクチン開発に必要な情報をデータベース化して公開（一部制限）し、関連研究者に活用してもらうことを目的として HIV 感染症統合データベースを構築しました。データの独占はエイズワクチン開発の遅延に繋がり好ましくありません。反対に、データの共有は幅広いアプローチでの研究推進に繋がり、ワクチン開発を加速します。本 DB の意義はそこにあります。一方、データの共有は個人情報の漏洩という危険につながります。しかしながら、本 WEB サーバ上には個人名は勿論、生年月日の月日、市町村レベルでの居住地等の情報は一切存在しません。すなちサーバ上のデータは連結不可能匿名化されています。さらに HIV 遺伝子情報以外の臨床データ部分へのアクセスを HIV 関連研究者のみに制限する事により、学会・論文発表データと同様に個人情報が厳しく保護されます。な

お、本 DB にはヒト遺伝子情報は含まれていません。

本 DB の最大の特色は、HIV 遺伝子情報を含むウイルス学的データに加えて臨床データを統合している点です。さらに HIV 遺伝子情報に基づいたウイルス蛋白 3 次元構造データを併載している点も他 DB には見られない特色です。HIV 遺伝子情報が主体の既存の DB¹²⁾では前述の幅広いアプローチでの研究推進に限界があります。本 DB の有効活用によりエイズワクチン・薬剤開発の進展が期待できます。

本 DB は、科学技術振興機構の研究情報データベース化事業 (<http://dbs.jst.go.jp/>) の 1 つとして国立感染症研究所と科学技術振興機構と共同で開発されました。

謝辞：アドバイザリー委員としてご協力いただいた次の諸先生に深謝致します。岡慎一先生（国立国際医療センター

エイズ治療・研究開発センター), 喜多恒和先生(防衛医科大学校産婦人科), 佐藤成大先生(岩手医科大学細菌学), 佐藤裕徳先生(国立感染症研究所遺伝子解析室), 椎野禎一郎先生(国立感染症研究所エイズ研究センター), 白幡聰先生(産業医科大学小児科), 杉浦亘先生(国立感染症研究所エイズ研究センター), 高松純樹先生(国立名古屋病院付属病院輸血部), 武部豊先生(国立感染症研究所エイズ研究センター), 永井美之先生(富山県衛生研究所), 福武勝幸先生(東京医科大学臨床病理部), 山田兼雄先生(エイズ予防財団)。また, 山口(加畠)由美先生(産業技術総合研究所生物情報解析研究センター)にも遠く米国(当時)よりご協力をいただきました事, 感謝致します。

文 献

- 1) 仲宗根正, 北村勝彦, 渡邊くほみ, 吉崎ひとみ, 本多三男: 遺伝子解析より見た HIV 感染症の現況と将来。臨床医 20 : 356-360, 1994.
- 2) 本多三男, 北村勝彦, 岡本ゆかり, 渡邊くほみ, 吉崎ひとみ, 福島誉子, 長繩聰, 宮本鋼, 染谷健二, 山田兼雄, 山崎修道: 臨床 HIV 株の解析とそのコンセンサス HIV 株の主要中和領域を用いたリコンビナント BCG ワクチンの感染防御能。臨床血液 36 : 435-441, 1995.
- 3) Yamanaka T, Fujimura Y, Ishimoto S, Yoshioka A, Konishi M, Narita N, Mimaya J, Meguro T, Nakasone T, Okamoto Y, Yoshizaki H, Yamada K, Honda M : Correlation of titer of antibody to principal neutralizing domain of HIV MN strain with disease progression in Japanese hemophiliacs seropositive for HIV type 1. AIDS Res Hum Retrovir 13 : 317-326, 1997.
- 4) Yasuda S, Iwasaki M, Oka S, Naganawa S, Nakasone T, Honda M, Sata T, Kojima A, Matsuda S, Takemori T, Tsunetsugu-Yokota Y : 1998. Detection of HIV-gag p24-specific antibodies in sera and saliva of HIV-1-infected adults and in sera of infants born to HIV-1-infected mothers. Microbiol Immunol 42 : 305-311, 1998.
- 5) Nakasone T, Totani R, Yamazaki S, Honda M : HIV-1 subtype A in Japan. AIDS 12 : 950-952, 1998.
- 6) Yoshino N, Naganawa S, Nakasone T, Imura S, National Cooperative Study Investigators on Vertical Transmission of HIV-1, Kita T, Honda M : Vertical transmission of human immunodeficiency virus type 1 (HIV-1) in Japan, 1989-1997 : presence of two subtypes B and E with subtype E predominance. Acta Pediatrica Japonica 40 : 503-508, 1998.
- 7) Nakasone T, Takamatsu J, Watanabe K, Naganawa S, Someya K, Yoshino N, Kaizu M, Ohsu T, Takizawa M, Izumi Y, Kawahara M, Hara T, Fujimura Y, Yamada K, Nagai Y, Yamazaki S, Honda M : Decline in the HIV-1 isolation rate in Japan : A 12 year observation. Microbiol Immunol 44 : 949-952, 2000.
- 8) Hara T, Yoshino N, Takayama N, Minamitani M, Naganawa S, Ohkubo H, Takizawa M, Izumi Y, Kantake M, Suzuki S, Takano M, Kita T, Totani R, Nagai Y, Honda M, Nakasone T : Presence of multiple HIV type 1 subtypes among mothers and children in Japan. AIDS Res and Hum Retrovir 17 : 569-575, 2001.
- 9) Naganawa S, Sato S, Nossik D, Takahashi K, Hara T, Tochikubo O, Kitamura K, Honda M, Nakasone T : First report of CRF03_AB recombinant HIV type 1 in injecting drug users in Ulkraine. AIDS Res and Hum Retrovir 18 : 1145-1149, 2002.
- 10) <http://www.jst.go.jp/inter/aids/index-j.htm>
- 11) http://www.mext.go.jp/a_menu/shinkou/shisaku/gensoku.htm
- 12) Los Alamos 研究所・HIV Database, <http://hiv-web.lanl.gov/content/index>

Development of HIV Infectious Diseases Integrated Database

Tadashi NAKASONE¹⁾, Takashi HARA^{1),2)}, Kenji SOMEYA^{1,2)}, Kazuho IKEO³⁾,
Takashi GOJOBORI³⁾, Naoki YAMAMOTO¹⁾ and Mitsuo HONDA¹⁾

¹⁾ AIDS Research Center, National Institute of Infectious Diseases, Tokyo, Japan

²⁾ Japan Science and Technology Agency, Tokyo, Japan

³⁾ Center for Information Biology and DNA Data Bank of Japan, Shizuoka, Japan

Objective : In order to make the wide use of the accumulated data for the development of HIV/AIDS vaccines, we developed the HIV Infectious Diseases Integrated Database (HIV-DB ; URL = <https://aids.nih.go.jp>) and made it available to any HIV-related researchers via Internet.

Outline : The HIV-DB provides virological and genetic data, such as results of virus isolation, nucleotide sequences information on HIV-1 and structural information on the virus proteins, integrated with indispensable clinical data, which have been collected and analyzed at the National Institute of Infectious Diseases (NIID) since 1988.

Contents : The HIV-DB includes the HIV DNA database of the DNA Databank of Japan (DDBJ), results of virus isolation (n=3,217), C2V3 DNA data (n=361), V3 loop structural data (PDB format, n=269), and corresponding clinical data (birth year, sex, CD4 lymphocyte count, plasma virus load, drug history, etc.), which have been obtained from 572 HIV-1-infected individuals in Japan. Several functions including FASTA similarity search analysis against the HIV DNA database of DDBJ, phylogenetic tree analysis, genosubtyping of HIV-1, creation of individual user divisions, browsing 3D view of V3 loop structure and search for corresponding clinical data, are also available on the WEB site.

Remarks : The HIV-DB was developed by collaboration of NIID and the Japan Science and Technology Corporation.

Key words : HIV infections diseases, database, HIV gene, HIV protein 3D structure