sequence set consisting of both the computationally detected orthologs and the homologs was prepared. For these sequence sets, phylogenetic trees were constructed by the neighbor-joining (NJ) method (18). In detail, multiple amino acid sequence alignments and phylogenetic trees were constructed by ClustalW (19) with the options of bootstrap = 1000, seed = 1, kimura, tossgaps, bootlabels = node.

Phylogenetic trees were examined by experts in the field of molecular evolution, who attended the evolutionary annotation meetings described in the introduction. The trees were drawn by NJplot (http://pbil.univ-lyon1.fr/software/njplot.html) and the default rooting was used. Discarding or re-rooting the tree was judged by the experts if necessary. All the ortholog pairs of human and other species detected by the computational analysis were examined (Figure 1). The primary principles of manual curation in Evola to be checked were as follows. [1] Phylogenetic topology between gene tree and species tree is consistent. As a gene tree, the minimum sub-clade including the pair (a part of the tree) was examined. As a species tree of reference, a phylogenetic tree indicating the trifurcation among primates, rodents and Laurasiatherian (dog, cow, etc.) species (20) was used, because the phylogenetic relationship has been controversial among them (21). In fact, we found that ((human–mouse)–dog) clades for some genes and ((human–dog)–mouse) clades for other genes. [2] Outgroup includes either two or more species that are phylogenetically distant from all the species in the sub-clade, or human and other species. In the latter case, human duplicate genes might exist. [3] Available bootstrap values of the corresponding three branches (one between the sub-clade and outgroup, and its two descendants) are all ≥900. The gene pairs consistent with all the principles were defined as 'manually curated orthologs', otherwise their annotation status remained to be 'computationally detected orthologs'.

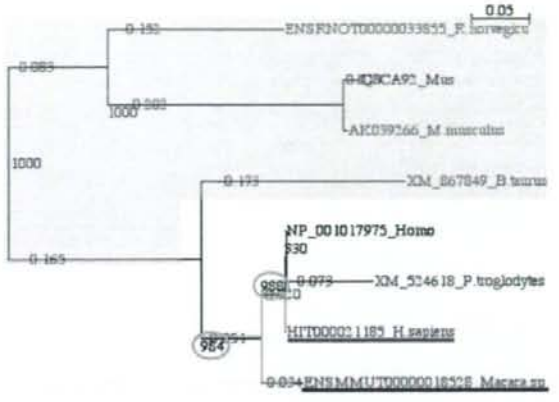


Figure 1. An example of manually curated gene pair from H.sapiens (red underline) and Macaca.sp (blue underline). In this case, conditions of phylogenetic topologies, outgroup species (light gray background) and bootstrap values (two circles) are sufficient (refer to the text). Thus, the pair was defined as a manually curated ortholog.

Table 1. Number of orthologs provided in Evola (release 4.1, June 2007)

Species	Genes	Human genes
Homo sapiens (Human)	18 968	–
Pan troglodytes (Chimpanzee)	16 368	15 615
Macaca sp. (Macaque)[a]	12 037	12 352
Mus musculus (Mouse)	15 570	14 574
Rattus norvegicus (Rat)	15 632	14 302
Canis familiaris (Dog)	14 730	13 916
Bos taurus (Cow)	9375	10 181
Monodelphis domestica (Opossum)	13 201	13 588
Gallus gallus (Chicken)	9266	10 738
Danio rerio (Zebrafish)	12 334	10 468
Tetraodon nigroviridis (Tetraodon)	11 505	9820
Takifugu rubripes (Fugu)	9738	9459

Numbers of genes of both human and other species are listed. Owing to lineage-specific duplication or loss, the numbers are usually different (for example, 15 570 mouse genes are orthologous to 14 574 human genes). 18 968 human genes have at least one ortholog among other 11 species.

[a]*Macaca mulatta, Macaca fascicularis, Macaca fuscata,* etc. are included.

DATABASE CONTENTS

Evola contains two ortholog datasets: (1) more comprehensive set of orthologs (computational analysis); and (2) more reliable orthologs (computational analysis supported by manual curation). In the current Evola (release 4.1), orthology information for 18 968 human genes is available among 11 vertebrates: chimpanzee, macaque, mouse, rat, dog, cow, opossum, chicken, zebrafish, Tetraodon and Fugu (Table 1). Manually curated orthologs occupied 25.4% of all computationally detected ortholog pairs (24 122/94 935) (release 4.1, 2007).

Evola is a sub-database of H-InvDB (9,10,11), and orthology information in Evola is, as 'Evolutionary annotation', a part of the comprehensive human gene annotations in H-InvDB. Thus, orthology information can be utilized with close reference to other annotation in H-InvDB. For example, 2090 human genes with orthology information belonged to H-Inv protein similarity categories of 'hypothetical proteins' (similarity category IV–VI). Molecular functions of these hypothetical proteins can be analyzed using model species. Moreover, cross references between Evola and other annotations in H-InvDB (protein–protein interaction (PPI), expression, polymorphism, disease, etc.) can produce valuable information contributing to the comprehensive understanding of the human genes.

We aimed to develop user-friendly interfaces that provide easy access to a variety of orthology information in Evola. Users can search orthologs in the top page of Evola as well as in the search systems of H-InvDB [simple search, advanced search and navigation system (Navi)]. Users can download data for each human gene on the main page as well as all the data of Evola in the download page. On the main page of Evola (Figure 2),

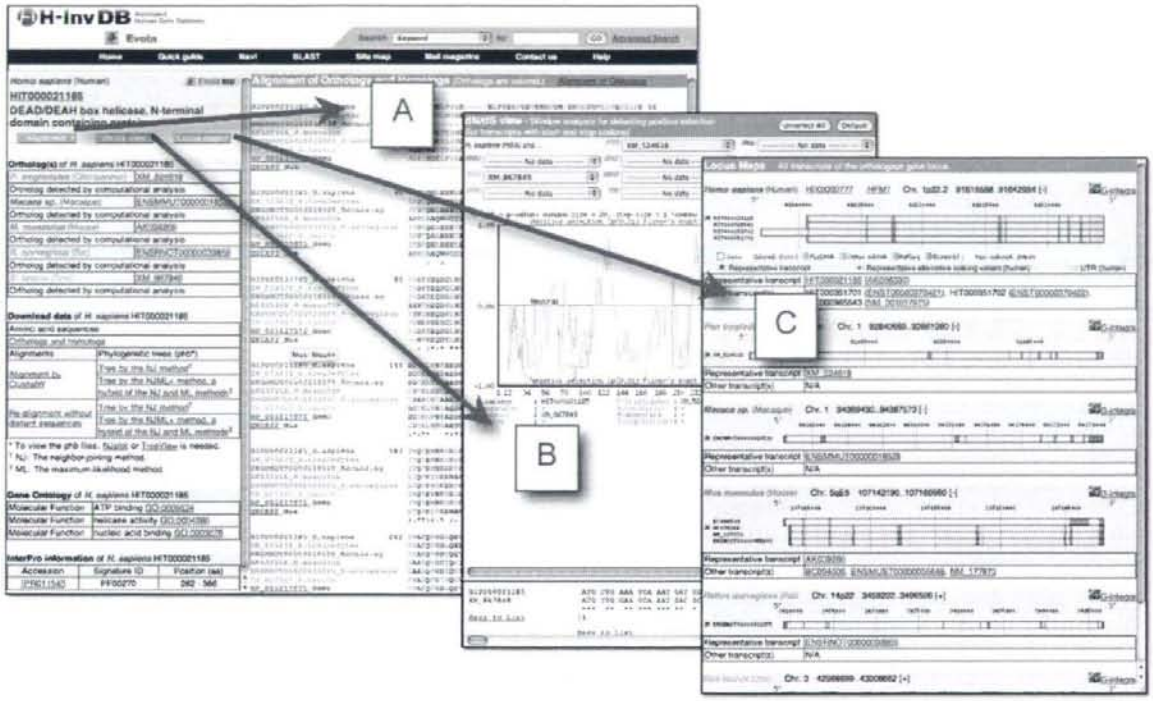


Figure 2. Evola main page. This page is divided into left and right frames. In the left frame, tables of orthologs, download data, Gene ontology, and InterPro are listed. Three green buttons are links to show 'Alignment' (A), 'd_N/d_S view' (B) and 'Locus maps' (C) in the right frame.

the following information for a human gene is available in the left frame: gene name, ortholog list with annotation status, download of sequences, alignments and phylogenetic trees, Gene ontology (22) and InterPro (23). In addition to the set of original ClustalW alignments, another set of alignments, including properly aligned sequences only (24), was also constructed and provided. In the latter sets, sequences with distinctively low identity to other sequences in an alignment were excluded. Based on both alignment sets, phylogenetic trees were constructed by the neighbor-joining method (18) and the NJML + method (25).

In the right frame of the main page, Evola features the three views described below. Users can switch among the views.

Alignment: Multiple alignments of orthologs and homologs (Figure 2A)

Amino acid sequence alignments of orthologs and homologs are displayed. Users can switch from 'Alignment of Orthologs' (default) to 'Alignment of Orthologs and Homologs', or vice versa. Each amino acid residue is color coded as defined in ClustalX (19). Accession numbers and species names of orthologs (human and other species) are colored in their species colors defined in Evola (human in red, mouse in gray, etc.). Accession numbers of homologs are linked to the original data sources of UniProt or RefSeq. While species are labeled by their scientific names (Homo, Mus, etc.), users can activate a popup window giving a species common name by placing the mouse cursor over homolog accession numbers (for example, 'Q5R508_Pongo'). InterPro data in the left frame include positional information on a human gene, and they can be utilized to detect conserved domains in the proteins.

d_N/d_S view: Window analysis detecting regions under positive or negative selection (Figure 2B)

Users can select one or more species for which to show the plots in the graph. In the lower frame under the graph, the pairwise nucleotide sequence alignment of CDSs is shown. The sequence positions (a.a. or codon) appearing in the graph and alignment are those of human genes.

The nonsynonymous to synonymous substitution rate ratio (d_N/d_S) is a commonly used measure of natural selection. In order to visualize positively and negatively selected regions, sliding window analysis was conducted (a 20 codon window with 1 codon stepping; result for the first window appears as a plot at 11th codon of the human gene). The statistical significance (P-value) of the difference between the number of nonsynonymous substitution (n) per synonymous substitutions (s): n/s, and the number of nonsynonymous sites (N) per synonymous sites (S): N/S was calculated by Fisher's exact test.

d_S, d_N, s and n values were estimated by the modified Nei–Gojobori method (26,27). If $d_N/d_S > 1$, the score ($= 1 - P$-value) was plotted above the zero line (neutral), and if $d_N/d_S < 1$, the score [$= -(1 - P$-value)] was plotted below the zero line. The regions plotted above the red line indicate that the sites might be under positive selection ($d_N/d_S > 1$ and $P < 0.01$). Conversely, the regions plotted below the blue line indicate that the sites might be under negative (purifying) selection ($d_N/d_S < 1$ and $P < 0.01$).

Locus maps: Comparative maps of orthologous gene loci (Figure 2C)

Orthologs were detected for representative transcripts (one transcript per gene locus) in Evola. However, there could be transcript variants in gene loci that have different exon–intron structures leading to produce different protein isoforms. Thus, information on other transcripts besides the representative transcript among orthologous gene loci are shown in Locus maps. In the figures, exon/intron structure, coding sequence (CDS) and untranslated regions (UTR) for each transcript are visualized. H-Inv cluster ID (HIX, an identifier of gene locus), Gene symbol, genomic location and a link to 'G-integra', an integrated genome browser of H-InvDB, are available. The flag icon denotes the representative transcript. The blue diamond icon denotes the Representative Alternative Splicing Variant (RASV) that is another representative per transcript group consisting of the same alternative splicing pattern (28). Representative transcripts are also RASVs, and blue diamonds do not appear if there is only one splicing isoform. In the tables, the H-Inv transcript ID (HIX) and original accession numbers (DDBJ/EMBL/GenBank, Ensembl and RefSeq) of the representative transcript and other transcripts are listed.

FUTURE DIRECTIONS

As our update policy, orthology information in Evola is updated when H-InvDB annotation is updated. One major update and three minor updates per year are scheduled. At the next major update on December 2007, a new duplicate gene family view is planned to be integrated within Evola. Human duplicate gene family data was originally constructed based on both amino acid sequence similarity (29) and orthology information. In the current Evola (release 4.1), parts of human duplicate gene annotation have been already implemented. The human duplicate genes are included in the alignments and phylogenetic trees of orthologs and homologs. Finally, we expect Evola to serve as a new database for evolutionary annotation of human genes. We sincerely welcome any requests and feedback from users.

ACKNOWLEDGEMENTS

We thank the members of Integrated Database Group, Japan Biological Information Research Center for their helpful suggestions. We are also grateful to Craig Gough for critical reading of the manuscript. This work was supported by the Ministry of Economy, Trade and Industry of Japan (METI), and the Japan Biological Informatics Consortium (JBIC). Funding to pay the Open Access publication charges for this article was provided by JBIC.

Conflict of interest statement. None declared.

REFERENCES

1. Tatusov,R.L., Koonin,E.V. and Lipman,D.J. (1997) A genomic perspective on protein families. *Science*, **278**, 631.
2. Wall,D.P., Fraser,H.B. and Hirsh,A.E. (2003) Detecting putative orthologs. *Bioinformatics*, **19**, 1710–1711.
3. Fortna,A., Kim,Y., MacLaren,E., Marshall,K., Hahn,G., Meltesen,L., Brenton,M., Hink,R., Burgers,S., Hernandez-Boussard,T. *et al.* (2004) Lineage-specific gene duplication and loss in human and great ape evolution. *PLoS. Biol.*, **2**, E207.
4. O'Brien,K.P., Remm,M. and Sonnhammer,E.L. (2005) Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res.*, **33**, D476–D480.
5. Birney,E., Andrews,D., Caccamo,M., Chen,Y., Clarke,L., Coates,G., Cox,T., Cunningham,F., Curwen,V., Cutts,T. *et al.* (2006) Ensembl 2006. *Nucleic Acids Res.*, **34**, D556–D561.
6. Wheeler,D.L., Barrett,T., Benson,D.A., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., DiCuccio,M., Edgar,R., Federhen,S. *et al.* (2007) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **35**, D5–D12.
7. Dufayard,J.F., Duret,L., Penel,S., Gouy,M., Rechenmann,F. and Perriere,G. (2005) Tree pattern matching in phylogenetic trees: automatic search for orthologs or paralogs in homologous gene sequence databases. *Bioinformatics*, **21**, 2596–2603.
8. Li,H., Coghlan,A., Ruan,J., Coin,L.J., Heriche,J.K., Osmotherly,L., Li,R., Liu,T., Zhang,Z., Bolund,L. *et al.* (2006) TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res.*, **34**, D572–D580.
9. Imanishi,T., Itoh,T., Suzuki,Y., O'Donovan,C., Fukuchi,S., Koyanagi,K.O., Barrero,R.A., Tamura,T., Yamaguchi-Kabata,Y., Tanino,M. *et al.* (2004) Integrative annotation of 21,037 human genes validated by full-length cDNA clones. *PLoS. Biol.*, **2**, e162.
10. Yamasaki,C., Koyanagi,K.O., Fujii,Y., Itoh,T., Barrero,R., Tamura,T., Yamaguchi-Kabata,Y., Tanino,M., Takeda,J., Fukuchi,S. *et al.* (2005) Investigation of protein functions through data-mining on integrated human transcriptome database, H-Invitational database (H-InvDB). *Gene*, **364**, 99–107.
11. Yamasaki,C., Murakami,K., Fujii,Y., Sato,Y., Harada,E., Takeda,J., Taniya,T., Sakate,R., Kikugawa,S., Shimada,M. *et al.* (2008) The H-Invitational Database (H-InvDB), A comprehensive annotation resource for human genes and transcripts. *Nucleic Acids Res.*, **36**, in press.
12. Fujii,Y., Itoh,T., Sakate,R., Koyanagi,K.O., Matsuya,A., Habara,T., Yamaguchi,K., Kaneko,Y., Gojobori,T., Imanishi,T. *et al.* (2005) A web tool for comparative genomics: G-compass. *Gene*, **364**, 45–52.
13. Schwartz,S., Kent,W.J., Smit,A., Zhang,Z., Baertsch,R., Hardison,R.C., Haussler,D. and Miller,W. (2003) Human-mouse alignments with BLASTZ. *Genome Res.*, **13**, 103–107.
14. Kent,W.J. (2002) BLAT—the BLAST-like alignment tool. *Genome. Res.*, **12**, 656–664.
15. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
16. Rice,P., Longden,I. and Bleasby,A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.
17. Pearson,W.R. (2000) Flexible sequence similarity searching with the FASTA3 program package. *Methods Mol. Biol.*, **132**, 185–219.

18. Saitou,N. and Nei,M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**, 406–425.
19. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
20. Hedges,S.B. and Kumar,S. (2002) Genomics. Vertebrate genomes compared. *Science*, **297**, 1283–1285.
21. Huttley,G.A., Wakefield,M.J. and Easteal,S. (2007) Rates of genome evolution and branching order from whole genome analysis. *Mol. Biol. Evol.*, **24**, 1722–1730.
22. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
23. Mulder,N.J., Apweiler,R., Attwood,T.K., Bairoch,A., Bateman,A., Binns,D., Bork,P., Buillard,V., Cerutti,L., Copley,R. *et al.* (2007) New developments in the InterPro database. *Nucleic Acids Res.*, **35**, D224–D228.
24. Endo,T., Ogishima,S. and Tanaka,H. (2002) ETools: Tools to Handle Biological Sequences and Alignments for Evolutionary Studies. *Genome Inform.*, **13**, 543–544.
25. Ota,S. and Li,W.H. (2001) NJML+: an extension of the NJML method to handle protein sequence data and computer software implementation. *Mol. Biol. Evol.*, **18**, 1983–1992.
26. Nei,M. and Gojobori,T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.*, **3**, 418–426.
27. Zhang,J., Rosenberg,H.F. and Nei,M. (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. USA*, **95**, 3708–3713.
28. Takeda,J., Suzuki,Y., Nakao,M., Barrero,R.A., Koyanagi,K.O., Jin,L., Motono,C., Hata,H., Isogai,T., Nagai,K. *et al.* (2006) Large-scale identification and characterization of alternative splicing variants of human gene transcripts using 56,419 completely sequenced and manually annotated full-length cDNAs. *Nucleic Acids Res.*, **34**, 3917–3928.
29. Gu,Z., Cavalcanti,A., Chen,F.C., Bouman,P. and Li,W.H. (2002) Extent of gene duplication in the genomes of Drosophila, nematode, and yeast. *Mol. Biol. Evol.*, **19**, 256–262.

D22–D24 *Nucleic Acids Research, 2008, Vol. 36, Database issue* *Published online 25 October 2007*
doi:10.1093/nar/gkm889

DDBJ with new system and face

H. Sugawara, O. Ogasawara, K. Okubo, T. Gojobori and Y. Tateno*

Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Research Organization of Information and Systems, Yata, Mishima 411-8540, Japan

Received September 13, 2007; Revised October 1, 2007; Accepted October 2, 2007

ABSTRACT

DDBJ (http://www.ddbj.nig.ac.jp) collected and released 1 880 115 entries or 1 134 086 245 bases in the period from July 2006 to June 2007. The released data contains the high-throughput cDNAs of cricket and high-quality draft genome of medaka among others. Our computer system has been upgraded since March 2007. Another new aspect is an efficient data retrieval tool that has recently been equipped and served at DDBJ. It is called All-round Retrieval for Sequence and Annotation, which enables the user to search for keywords also in the Feature/Qualifier of the International Nucleotide Sequence Database Collaboration (http://www.insdc.org/). We will also replace our home page with a more efficient one by the end of 2007.

INTRODUCTION

Through our service we have witnessed dramatic advancements in biology and the related areas in the past 20 or more years. For example, using genome sequence data for eubacteria, archaebacteria and eukaryotes, some authors constructed a tree of life, which is the phylogenetic tree of the three super-kingdoms (1,2). Others reported a way to predict the number of genes at least in the bacterial world (3). The dramatic advancements prove our simple idea that the more data we collect and serve the more people make use of it for various purposes.

On the other hand, the recent development of sequencing machines such as 454 (by 454 Life Sciences), Solexa (by Illumina, Inc.) and SOLiD (by Applied Biosystems) makes us worrisome as well. According to some estimate, 5–10 tera bases will be sequenced by Solexa at one sequencing facility in a month in the near future. With the further development of the sequencing technology the whole genome of a person may repeatedly be submitted in the near future, as few examples warn (4). To cope with the expected situation of sequencing genes and genomes, we have recently upgraded our computer system and installed an efficient keyword search tool. We think that the new computer and tool serve our data submitters and users better and make our job more effective and efficient.

In this article we will report on the data submissions to DDBJ in the past year, replacement of our computer system with an upgraded one, a new data retrieval tool and a new home page.

DATA SUBMISSIONS TO DDBJ IN THE LAST YEAR

In the period from July 2006 to June 2007, DDBJ collected and released the original data of 1 880 115 entries or 1 134 086 245 bases that were classified into the 19 International Nucleotide Sequence Database Collaboration (INSDC) divisions (5). More than 90% of the submissions came from Japanese researchers, and the rest were mainly from Chinese and Korean researchers.

The released data includes the high-throughput cDNAs (HTC) of cricket, *Gryllus bimaculatus* submitted from Tokushima University (6). The data amount is 32 010 entries that can be obtained through anonymous FTP with the file name, Gryllus_bimaculatus_HTC_070726_1.seq.gz. Also included is 700 Mb of the high-quality draft genome data of medaka, *Oryzias latipes*, which was submitted from University of Tokyo and National Institute of Genetics (7). The data was carefully assembled and upgraded from the WGS data that was reported in our previous paper (8). The given accession numbers are BAAF03000000 (Hd-rR, version 0.9), BAAF04000000 (Hd-rR, version 1.0) BAAE01000000 (HNI) and ACAAA0000001-ACAAA0356693 (5′ SAGE tags). Although draft genome sequences of two fugu (blowfish) species are available, the high-quality draft genome of medaka will be quite useful particularly for the study of vertebrate evolution. The submitters of the genome data discussed, for example, that the medaka genome preserved its ancestral karyotype for more than 300 million years (7).

It is also noted that the current number of bacterial species/strains in the complete bacteria genome data repository, the Genome Information Broker, (GIB, http://gib.genes.nig.ac.jp/) (9), at DDBJ is 569 and keeps on growing rapidly. The species added in the past year include *Methanococcus maripaludis* (by Joint Genome Institute), *Saccharopolyspora erythraea* (by University of Cambridge), *Francisella tularensis* subsp. *tularensis*

*To whom correspondence should be addressed. Tel: +81-55-981-6857; Fax: +81-55-981-6858; Email: ytateno@genes.nig.ac.jp

(by UT Southwestern Medical Center), *Desulfotomaculum reducens* (by Joint Genome Institute), *Burkholderia vietnamiensis* (by Joint Genome Institute), *Herminiimonas arsenicoxydans* (by Genoscope), *Geobacillus thermodenitrificans* (by Nankai University), *Corynebacterium glutamicum* (by RITE) and many others. We also serve a complete virus genome data repository, GIB for Viruses (GIB-V, http://gib-v.genes.nig.ac.jp/) that now contains 31 486 virus genomes and genomic segments.

NEW COMPUTER SYSTEM

In July 2007, we celebrated the 20th anniversary of the public release of the DNA data. Our first release in July 1987 contained only 66 entries or 108 970 bases that were typed in from published papers. These numbers may be impressive in the comparison with the corresponding ones as of June 2007, 13 371 690 entries or 8 988 178 758 bases. This tremendous increase in the numbers perhaps reflects the remarkable advancement of research in biology and the related areas in Japan in the past 20 years. The ever-increasing amount of the data also makes us worry about our hardware and software facilities.

In March 2007, we completely replaced our computer system with an upgraded one. Major upgraded aspects are as follows. (i) The increase in the number of entries in making the flat files from 300 000 to 1 000 000 entries/day, (ii) the decrease in processing time in making a huge flat file; in case of four rice chromosomes, from 110 to 13 min, (iii) the decrease in processing time from 120 to 13 min for updating the live-list that lists the accession numbers and dates of the public release of the released entries; it is weekly updated to exchange the information about the currently released data with the EMBL Bank and GenBank, (iv) the increase in the number of ESTs in data processing from 40 000 to 800 000 entries/h and (v) the increase in the number of queries accepted at once by 1.5 times. Therefore, we will be able to cope with the increase in the number of data submissions for the next several years.

NEW KEYWORD SEARCH TOOL

Recently, we have installed a high-speed keyword search tool, All-round Retrieval for Sequence and Annotation (ARSA, http://arsa.ddbj.nig.ac.jp/top-e.html). The search logic behind ARSA is called SIGMA, which was invented by Arikawa and his colleagues (10,11). For a given query SIGMA makes it possible to retrieves all the right entries by checking the contents of a database just once, no matter how the query is complicated. The one time checking makes keyword search fast. SIGMA does not need an index file, which means that search can be made against the currently available data. SIGMA is implemented on the Shunsaku search engine developed by Fujitsu. The search engine operates in parallel for divided data, which makes the search even faster. ARSA also has a large scalability with an increasing amount of data. In theory, one search can be completed within 10 s irrespective of the data size and the query formula. If the data increases more than 10 times larger than the current amount, however, we may have to increase the number units in the Shunsaku accordingly to keep the present search speed.

ARSA covers 23 databases including DDBJ, UniProt, PFAM, PDB and LENZYME. A special feature of ARSA is that it can also incorporate the terms defined by the Feature/Qualifier of INSDC. While this feature is very helpful for us to annotate the submitted data, it enables our user to perform data retrieval by using terms in the Feature/Qualifier. For example, you can search for CDSs (protein coding sequences) located on human Y chromosome, as shown in Figure 1. In the figure, the query formula is given on the top, and a part of the hit entries is given below with the accession numbers. By clicking one of the numbers you can see its contents. HUM in the last column stands for the human division. You can download the search result in Flat File, FASTA or XML, and also choose the items in the search results to be displayed on the computer screen and directly download them in tab-limited format. We also provide you with WebAPI (http://xml.nig.ac.jp/>http://xml.nig.ac.jp/) (12) so that you can customize ARSA by writing a program in Perl or JAVA. We will soon include KEGG (http://www.genome.ad.jp/kegg/) in ARSA and make the 24 databases simultaneously retrievable for common keywords.

NEW FACE OF DDBJ

We updated our home page (HP) in 2005 (13). We are again in the process of updating it rather drastically this time. Since the present HP holds many contents that have been added in an irregular sequence without much consideration for consistency, it is not really convenient now for our data submitters and users. The main point of the updating thus is to reach the almost every content with three clicks or less, which is now a common practice in making use of a HP. In the new HP when you click one of our main services, data submission, data retrieval, ftp/SOAP, statistics and inquiry, you can get the whole view of all contents for each service at once, and easily go to the one of them that you wish. The new HP will replace the present one by the end of 2007. We hope the new HP on the new computer system and tool will be more attractive to our data submitters and users worldwide.

ACKNOWLEDGEMENTS

We thank all DDBJ staff for the collection, annotation and distribution of the original data. We are also grateful to H. Yamada and T. Hosokawa for providing us with valuable information on ARSA. DDBJ is funded by Ministry of Education, Culture, Sports, Science and Technology (MEXT) with Management experience grants for national university cooperation. Funding to pay the Open Access publication charges for this article was provided by Japan Society for Promotion of Science Grant no. 16255006.

Conflict of interest statement. None declared.

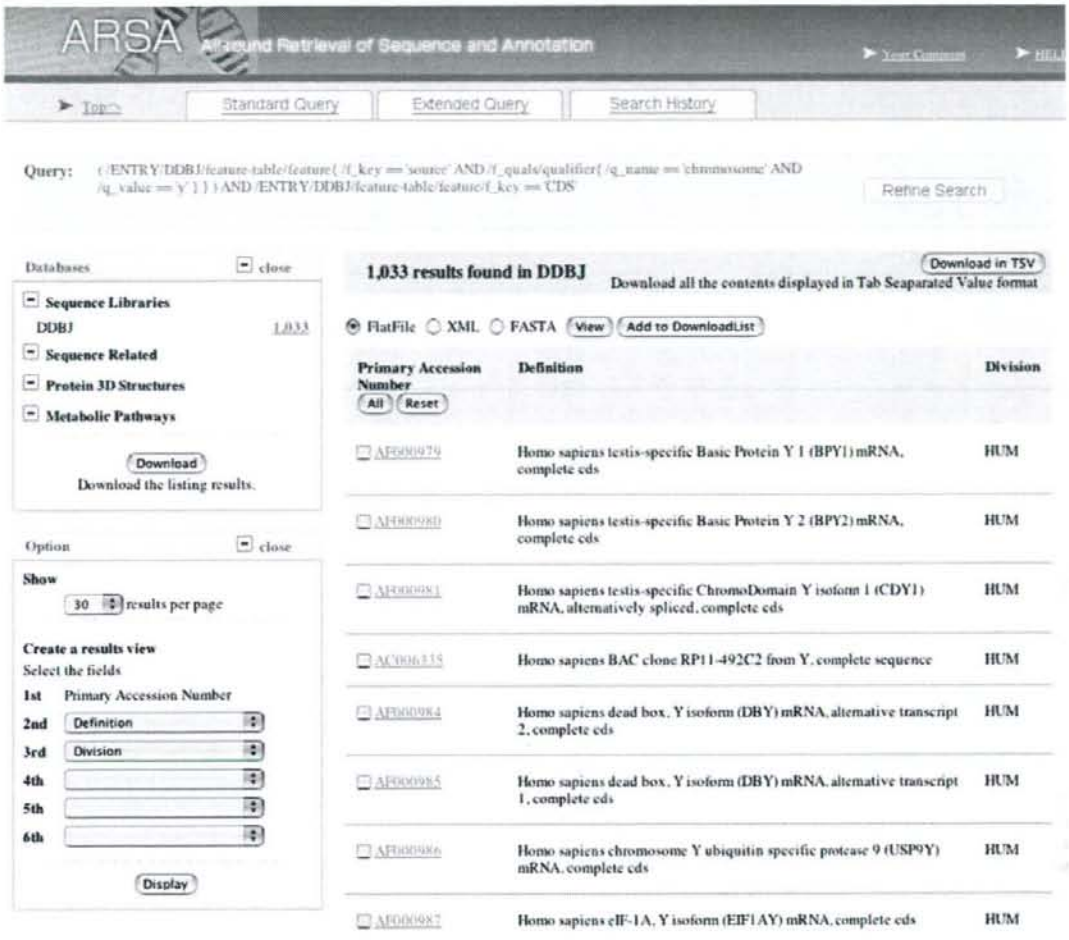


Figure 1. Example of keyword search by ARSA. Keywords used are 'source' (Feature), 'chromosome' (Qualifier belonging to 'source') and CDS (Feature). 'Chromosome' has a value attribute to which 'Y' is given for specifying chromosome Y.

REFERENCES

1. Wolf,Y.I., Rozogin,I.B., Grishin,N.V. and Koonin,E.V. (2002) Genome trees and the tree of life. *Trends Genet.*, **18**, 472–479.
2. Fukami-Kobayashi,K., Minezaki,Y., Tateno,Y. and Nishikawa,K. (2007) A tree of life based on protein domain organizations. *Mol. Biol. Evol.*, **24**, 1181–1189.
3. Kosuge,T., Abe,T., Okido,T., Tanaka,N., Hirahata,M., Maruyama,Y., Mashima,J., Tomiki,A., Kurokawa,M. *et al.* (2006) Exploration and grading of possible genes in 183 bacterial strains by a common fine protocol lead to new genes: Gene Trek in Prokaryote Space (GTPS). *DNA Res.*, **13**, 245–254.
4. Levy,S., Sutton,G., Ng,P.C., Feuk,L., Halpern,A.L., Walenz,B.P., Axelrod,N., Huang,J., Kirkness,E.F. *et al.* (2007) The diploid genome sequence of an individual human. *PLoS Biol.*, **5**, e254.
5. Okubo,K., Sugawara,H., Gojobori,T. and Tateno,Y. (2006) DDBJ in preparation for overview of research activities behind data submissions. *Nucleic Acids Res.*, **34**, D6–D9.
6. Shinmyo,Y., Mito,T., Uda,T., Nakamura,T., Miyawaki,K., Ohuchi,H. and Noji,S. (2006) Brachyenteron is necessary for morphogenesis of the posterior gut but not for anteroposterior axial elongation from the posterior growth zone in the intermediate-germband cricket *Gryllus bimaculatus*. *Development*, **133**, 4539–4547.
7. Kasahara,M., Naruse,K., Sasaki,S., Nakatani,Y., Qu,W., Ahsan,B., Yamada,T., Nagayasu,Y., Doi,K. *et al.* (2007) The medaka draft genome and insights into vertebrate genome evolution. *Nature*, **447**, 714–719.
8. Sugawara,H., Abe,T., Gojobori,T. and Tateno,Y. (2007) DDBJ working on evaluation and classification of bacterial genes in INSDC. *Nucleic Acids Res.*, **35**, D13–D15.
9. Fumoto,M., Miyazaki,S. and Sugawara,H. (2002) Genome Information Broker (GIB): data retrieval and comparative analysis system for completed microbial genomes and more. *Nucleic Acids Res.*, **30**, 66–68.
10. Arikawa,S. (1981) One-way sequential search systems and their powers. *Bull. Math. Stat.*, **19**, 69–85.
11. Arikawa,S., Shinohara,T. and Takeya,S. (1989) SIGMA: A text database management system. *Berliners Informatik Tag*, 72–81.
12. Sugawara,H. and Miyazaki,S. (2003) Biological SOAP servers and web services provided by the public sequence data bank. *Nucleic Acids Res.*, **31**, 3836–3839.
13. Tateno,Y., Saitou,N., Okubo,K., Sugawara,H. and Gojobori,T. (2005) DDBJ in collaboration with mass-sequencing teams on annotation. *Nucleic Acids Res.*, **33**, D25–D28.

Journal of Viral Hepatitis, 2008 doi:10.1111/j.1365-2893.2008.00997.x

Two simultaneous hepatitis B virus epidemics among injecting drug users and men who have sex with men in Buenos Aires, Argentina: characterization of the first D/A recombinant from the American continent

J. Trinks,[1,2,*] M. L. Cuestas,[1,*] Y. Tanaka,[3] V. L. Mathet,[1,2] M. L. Minassian,[1,2] C. W. Rivero,[1,2] J. A. Benetucci,[4] E. D. Gímenez,[5] M. Segura,[6] M. C. Bobillo,[7] D. Corach,[2,7] P. D. Ghiringhelli,[2,8] D. O. Sánchez,[2,9] M. M. Ávila,[2,6] L. A. M. Peralta,[2,6] F. Kurbanov,[3] M. C. Weissenbacher,[2,6] P. Simmonds,[10] M. Mizokami[3] and J. R. Oubiña[1,2] [1]*Centro para el Estudio de Hepatitis Virales, Depto. de Microbiología, Fac. de Medicina, Universidad de Buenos Aires (UBA), Buenos Aires, Argentina;* [2]*Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina;* [3]*Department of Clinical Molecular Informative Medicine, Nagoya City University Graduate School of Medical Sciences, Kawasumi, Nagoya, Japan;* [4]*Fundación de Ayuda al Inmunodeficiente (FUNDAI), Buenos Aires, Argentina;* [5]*Hospital Zonal General de Agudos 'Dr. Isidoro G. Iriarte', Quilmes, Argentina;* [6]*Centro Nacional de Referencia para el SIDA, Depto. de Microbiología, Fac. de Medicina, UBA, Buenos Aires, Argentina;* [7]*Servicio de Huellas Digitales Genéticas y Cátedra de Genética Molecular, Fac. de Farmacia y Bioquímica, UBA, Buenos Aires, Argentina;* [8]*Laboratorio de Ingeniería Genética y Biología Celular Molecular, Universidad Nacional de Quilmes, Bernal, Argentina;* [9]*Universidad Nacional de San Martín, San Martín, Argentina; and* [10]*Virus Evolution Group, Laboratory for Clinical and Molecular Virology, University of Edinburgh, Summerhall, Edinburgh, UK*

Received February 2008; accepted for publication April 2008

SUMMARY. Previous studies have revealed that hepatitis B virus (HBV)/D and HBV/F predominate among blood donors from Buenos Aires, Argentina. In the present study, blood samples from two high-risk groups were analysed: 160 corresponding to street- and hospital-recruited injecting drug users [81.2% showing the 'anti-hepatitis B core antigen (anti-HBc) only' serological pattern] and 20 to hepatitis B surface antigen $(HBsAg)^+$/$anti\text{-}HBc^+$ men who have sex with men. HBV genotypes were assigned by polymerase chain reaction amplification followed by restriction fragment length polymorphism and confirmed by nucleotide sequencing of two different coding regions. HBV DNA was detected in 27 injecting drug users (16.9%, occult infection prevalence: 7.7%), and 14 men who have sex with men (70%). HBV/A prevailed among injecting drug users (81.8%) while HBV/F was predominant among men who have sex with men (57.1%). The high predominance of HBV/A among injecting drug users is in sharp contrast to its low prevalence among blood donors ($P = 0.0006$) and men who have sex with men ($P = 0.0137$). Interestingly, all HBV/A S gene sequences obtained from street-recruited injecting drug users encoded the rare serotype *ayw*1 and failed to cluster within any of the known A subgenotypes. Moreover, one of the HBV strains from a hospital-recruited injecting drug user was fully sequenced and found to be the first completely characterized D/A recombinant genome from the American continent. Data suggest that two simultaneous and independent HBV epidemics took place in Buenos Aires: one spreading among injecting drug users and another one sexually transmitted among the homosexual and heterosexual population.

Keywords: Argentina, hepatitis B virus, HBV recombinant, molecular epidemiology.

Abbreviations: HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; HCV, hepatitis C virus; HIV, human immunodeficiency virus; IDU, injecting drug user; MSM, men who have sex with men; NJ, neighbour-joining; pre-C/C, Precore/core; RFLP, restriction fragment length polymorphism.

Correspondence: Dr José Raúl Oubiña, Centro para el Estudio de Hepatitis Virales, Depto. de Microbiología, Fac. de Medicina, Universidad de Buenos Aires, Paraguay 2155, Piso 11, 1121 Buenos Aires, Argentina. E-mail: joubina@fmed.uba.ar
*These authors contributed equally to this study.

INTRODUCTION

Eight major genotypes of HBV (A to H) with distinct geographical distribution have been defined [1]. Some of these genotypes are subdivided into subgenotypes. For example, genotype A is classified into three major subgroups: A1 (in Asia and Africa), A2 (in Europe and the USA) and A3 (in Central West Africa) [2]. Likewise, genotype F is represented by four subgenotypes: F1 which includes subclades 1a (Central America) and 1b (Alaska and Argentina), F2 (widely present in Central and South America), F3 (Venezuela and Colombia) and F4 (Bolivia and Argentina) [2–4]. On the other hand, prior to the definition of the genotypes, HBV strains were distinguished by serological analysis into nine different HBsAg subtypes named *adw2*, *adw4*, *ayw1*, *ayw2*, *ayw3*, *ayw4*, *adrq+*, *adrq-* and *ayr* [1].

Previous reports have shown that genotype F is the most prevalent among Argentine chronic patients [5,6]. Furthermore, França *et al.* [22] observed in the same country a prevalence of 64% for genotype F, 17.3% each for genotypes A and D, and 1.3% for genotype C among blood donors. Noteworthy, the distribution of HBV genotype F is uneven throughout Argentina since it predominates in the northern part of the country (88.9%), in contrast with Buenos Aires city area where 40.9% of the 22 samples analysed were classified as HBV/D, 31.8% as HBV/F, 22.7% as HBV/A and 4.6% as HBV/C.

HBV molecular epidemiology data regarding risk groups, such as injecting drug users (IDUs) and men who have sex with men (MSM), are still lacking in Argentina and scarce in the rest of the world.

Bearing in mind both the emerging evidence of HBV genotype relevance in the course of the infection [7], as well as the lack of information associating a given HBV genotype with its route of transmission, the aims of the present study were to retrospectively assess the HBV genotype prevalence among the above mentioned at risk populations from Argentina and to analyse their genetic diversity and phylogenetic relatedness.

MATERIALS AND METHODS

Informed consent and Ethics Committee

All subjects had provided their informed written consent to carry out this study. This protocol received approval of the Independent Ethics Committee on Research (CIEI-FM-UBA).

Samples

During the period 2000–2001, serum samples and epidemiological data were collected from 80 street-recruited subjects from Buenos Aires with past or ongoing history of injecting drug addiction. Recruitment was done by the so-called snowball sampling procedure [8].

In order to rule out the possibility that results corresponding to street-recruited IDUs were merely the direct consequence of their behaviour as a close community due to the sampling technique chosen in this study, we decided to include a second group consisting of an identical number of unrelated IDUs, who individually visited public hospitals and treatment centres in Buenos Aires city and its suburbs. Their serum samples were obtained between 1995 and 2006 (most of them-except 16-during 2001–2002). IDUs patients were only enrolled if either HBsAg or total immunoglobulin to hepatitis B core antigen (total anti-HBc) were detected.

In addition, 20 MSM from Buenos Aires, previously enrolled in a cohort study [9] and whose serum samples had been collected between February and December 2003, were also included in this study. MSM were only included if both markers (HBsAg and total anti-HBc) were positive and no current or past history of injecting drug use was reported.

Serological assays

Serum samples had been stored at −70 °C until use. HBsAg and total anti-HBc were tested with commercially available kits (AxSYM, Abbot, IL, USA). Serological tests for human immunodeficiency virus (HIV) (Bio-Rad, Fujirebio, Tokyo, Japan) and hepatitis C virus (HCV) (AxSYM, Abbot, IL, USA) were also performed following the manufacturer's instructions.

HBV DNA PCR amplification, nucleotide sequencing and genotype assignment

By using a DNA extraction kit (Macherey, Nagel, Germany), HBV DNA was extracted from 200 μL of serum from all patients.

Initially, the HBV S gene was partially amplified by a previously described nested PCR protocol [10], yielding an amplicon of 585 bp (nucleotide position 203 to 787). HBV DNA was also looked for by three further methods: (i) another partial S gene PCR [11], yielding an amplicon of 541 bp (nucleotide position 256 to 796), and exceptionally: (ii) a nested PCR [10] followed by a boosted PCR (by using the same primers employed during the nested step) – as an attempt to make very faint bands useful for DNA sequencing-, as well as (iii) a nested PCR performed by combination of primers used by Zeng *et al.* (first round) and by Lindh *et al.* (second round) and its corresponding boosted PCR (third round).

Amplicons of expected sizes were detected in 1% agarose gels and subsequently subjected to restriction fragment length polymorphism (RFLP) [10,11]. In some samples, very faint bands were observed; consequently, the RFLP pattern was under the level of detection for proper HBV genotype assignment and the sample was considered to be untypeable.

In order to examine the possibility of an HBV recombination event, the Precore/core (pre-C/C) region was partially

amplified as well (nucleotide position 1756–2451), as previously described [12].

Subsequently, full-length HBV sequencing [13] was attempted in all DNA samples showing either inconsistencies/ambiguities of phylogenetic analysis of partial S and pre-C/C sequences or discrepancy between S gene RFLP and phylogenetic genotype assignment. However, full length sequences were only obtained from five samples (S-IDU4, S-IDU8, S-IDU11, S-IDU19 and H-IDU7) due to a combination of technical factors (low sera availability and very faint bands). Due to the retrospective nature of the study and the limited serum volume available, HBV DNA viral load was not measured.

Appropriate precautions and procedures were strictly followed to avoid cross-contamination [14,15]. PCR products were bidirectionally sequenced by using the Big-Dye Termination chemistry (Applied Biosystem, USA).

The GenBank/EMBL/DDBJ accession numbers of the sequences reported in this paper are: EU185765–EU185767, EU185781–EU185789 (street-recruited IDUs), EU185768–EU185780, EF467999 (hospital-recruited IDUs) and EU185741-EU185764 (MSM).

Sequence analysis

DNA alignments of the nucleotide sequences obtained and GenBank sequences corresponding to HBV genotypes A–H were generated with the CLUSTAL X program. Phylogenetic trees were constructed using neighbour-joining (NJ) analysis included within the PHYLIP package version 3.5 c [16]. A sequence identity matrix was recorded for each the pre-C/C and S gene partial sequences, by using the BioEdit Sequence Alignment Editor, version 7.0.1.

Detection of HBV DNA recombination and identification of recombination sites

All complete genome sequences from Argentine strains were examined with other HBV genotypes for the presence of potential recombination event(s) by running the bootscan analysis available in the SIMPLOT (version 3.5.1) [17] and by using the Grouping Scan included in the Simmonic 2005 Sequence Editor Package (version 1.5) [18]. Moreover, and in order to confirm the results obtained, a third method to detect recombinants developed by one of the authors (P. D. Ghiringhelli, unpublished data) was used as well. In this method, alignments were carried out with the CLUSTAL X program (default parameters: [19,20]) between sequence pairs, always involving the putative recombinant candidate and one representative sequence of each one of the putative parental genotypes. The relative similarities were calculated using the CLUSTAL X consensus symbol (* and blank space) as the input sequence, in an overlapping windows-based strategy. Arbitrary values of +1 for identical (*) and −1 for nonidentical (blank spaces) residues, to obtain the similarity profiles were assigned. The sum of assigned values for each residue in each window (101 nucleotides) was divided by the window width and allotted to the central position to generate the plots. Pairs of profiles were superimposed and analysed with the aim of detecting cross points between them. In order to find a good relation between graph complexity and cross-point detection sensitivity, windows length of 101 residues were scanned.

The breakpoints were estimated by mapping the informative sites using SIMPLOT (version 3.5.1) [17] and further confirmed by means of chi-square analysis [21].

Statistical analysis

Statistical differences were evaluated by two methods: (a) the chi-square calculation with Yates' correction, and (b) the two-tailed *t*-test. A P value < 0.05 was used as an indicator of statistical significance.

RESULTS

Serological status and HBV genotypes of street-recruited IDUs (S-IDUs)

Street-recruited IDUs were divided into three groups according to different HBV serological patterns. The first group included six IDUs (83.3% male; mean age ± SD = 27.6 years ± 6.9) who were HBsAg positive but total anti-HBc negative. The second one included 10 IDUs (90% male; 28.7 years ± 5.0) and showed positivity for both HBsAg as well as total anti-HBc. Finally, a third group encompassed 64 IDUs (82.8% male; 32.5 years ± 6.3) who exhibited the sole presence of total anti-HBc. The serological status for HCV and HIV from all HBV DNA PCR positive samples is depicted in Table 1.

HBV S gene was amplified in 20 out of 80 samples (25%): three patients from the first group (50%), eight patients from the second group (80%) and nine from the third one (14.1%). In 5 of these 20 cases, negligible yields of DNA products were obtained from PCR, even after performing a third round of amplification. Therefore, only 15 samples out of these 20 were appropriate for an RFLP analysis [11]. Two out of 15 samples (13.3%), which belonged to the 'anti-HBc only' group, were classified as genotype C. The remaining samples (86.7%) were characterized as genotype A, whose assignment was further confirmed by performing a second RFLP method, originally described for characterization of genotypes A–F [11] (Table 1).

Furthermore, the HBV S coding region was successfully sequenced in 9 out of these 15 samples and all of them were confirmed as belonging to genotype A. These nine Argentine HBV strains clustered separately from HBV/A1, HBV/A2 and HBV/A3 strains (Fig. 1a). The phylogenetic analysis revealed that all the samples were closely related to each other, with (mean ± SD) 98.7 ± 0.8 nucleotide identity, as

Table 1 Major characteristics of the 20 HBV PCR (+) street-recruited IDUs (S-IDUs) from Buenos Aires, Argentina

				Serological status					S gene genotype		
Sample	Gender	Age	Date of blood sample collection	HBsAg	Anti-HBc Ab	Anti-HCV Ab	Anti-HIV Ab	Sexual orientation	RFLP	DNA sequencing (serotype)	Pre-C/C region genotype*
S-IDU1	M	26	March 2001	+	–	–	–	Heterosexual	A	N.D.	A2
S-IDU2	F	22	July 2000	+	–	–	–	Bisexual	A	N.D.	N.D.
S-IDU3	M	26	September 2000	+	–	–	–	Heterosexual	N.D.	N.D.	N.D.
S-IDU4†	M	28	August 2000	+	+	+	+	Bisexual	A	Au‡ (*ayw*1)	A2
S-IDU5	M	28	August 2000	+	+	+	–	Heterosexual	A	Au‡ (*ayw*1)	N.D.
S-IDU6	M	29	September 2000	+	+	+	+	Heterosexual	A	Au‡ (*ayw*1)	A2
S-IDU7	F	28	August 2000	+	+	+	–	Heterosexual	N.D.	N.D.	N.D.
S-IDU8†	M	26	February 2001	+	+	+	+	Heterosexual	A	Au‡ (*ayw*1)	A2
S-IDU9	M	40	June 2000	+	+	+	–	Heterosexual	A	Au‡ (*ayw*1)	N.D.
S-IDU10	M	25	June 2000	+	+	+	–	Heterosexual	N.D.	N.D.	N.D.
S-IDU11†	M	33	October 2000	+	+	+	+	Heterosexual	A	Au‡ (*ayw*1)	A2
S-IDU12	M	29	August 2000	–	+	+	+	Bisexual	N.D.	N.D.	N.D.
S-IDU13	M	24	August 2000	–	+	+	+	Heterosexual	C	N.D.	N.D.
S-IDU14	M	37	November 2000	–	+	+	+	Heterosexual	N.D.	N.D.	N.D.
S-IDU15	M	29	July 2000	–	+	–	–	Heterosexual	A	N.D.	N.D.
S-IDU16	M	29	July 2000	–	+	+	+	Heterosexual	A	Au‡ (*ayw*1)	A1
S-IDU17	M	30	July 2000	–	+	+	+	Heterosexual	A	N.D.	N.D.
S-IDU18	F	24	March 2001	–	+	+	+	Heterosexual	C	N.D.	N.D.
S-IDU19†	M	45	June 2000	–	+	+	–	Heterosexual	A	Au‡ (*ayw*1)	A2
S-IDU20	M	31	September 2000	–	+	+	+	Heterosexual	A	Au‡ (*ayw*1)	N.D.

N.D., not done due to very faint bands, even after performing a nested PCR followed by a boosted step. IDUs, injecting drug users; Pre-C/C, Precore/core. *Genotypes were assigned by DNA sequencing. †Full-length HBV sequence was obtained.‡HBV genotype A, subgenotype unidentified.

compared with 95.9 ± 1, 97.4 ± 0.8 and 95.6 ± 0.9 when S-IDUs strains were matched with HBV/A1, HBV/A2 and HBV/A3 sequences (n = 5 for each subgenotype), respectively. Estimated intra-group percentage nucleotide identity was (mean ± SD) 98 ± 0.5 for HBV/A1, 98 ± 0.6 for HBV/A2 and 99 ± 0.5 for HBV/A3. On the other hand, estimated inter-group percentage nucleotide identity consisted of (mean ± SD): 97 ± 0.9, HBV/A1 *vs* HBV/A2; 97 ± 0.4, HBV/A1 *vs* HBV/A3; and 96 ± 0.9, HBV/A2 *vs* HBV/A3. Moreover, all S-IDUs strains encoded the subtype *ayw*1 (Tables 1 and 2), as defined by the presence of Arg^{122}, Pro^{127}, Phe^{134}, Ala^{159} and Lys^{160} [1].

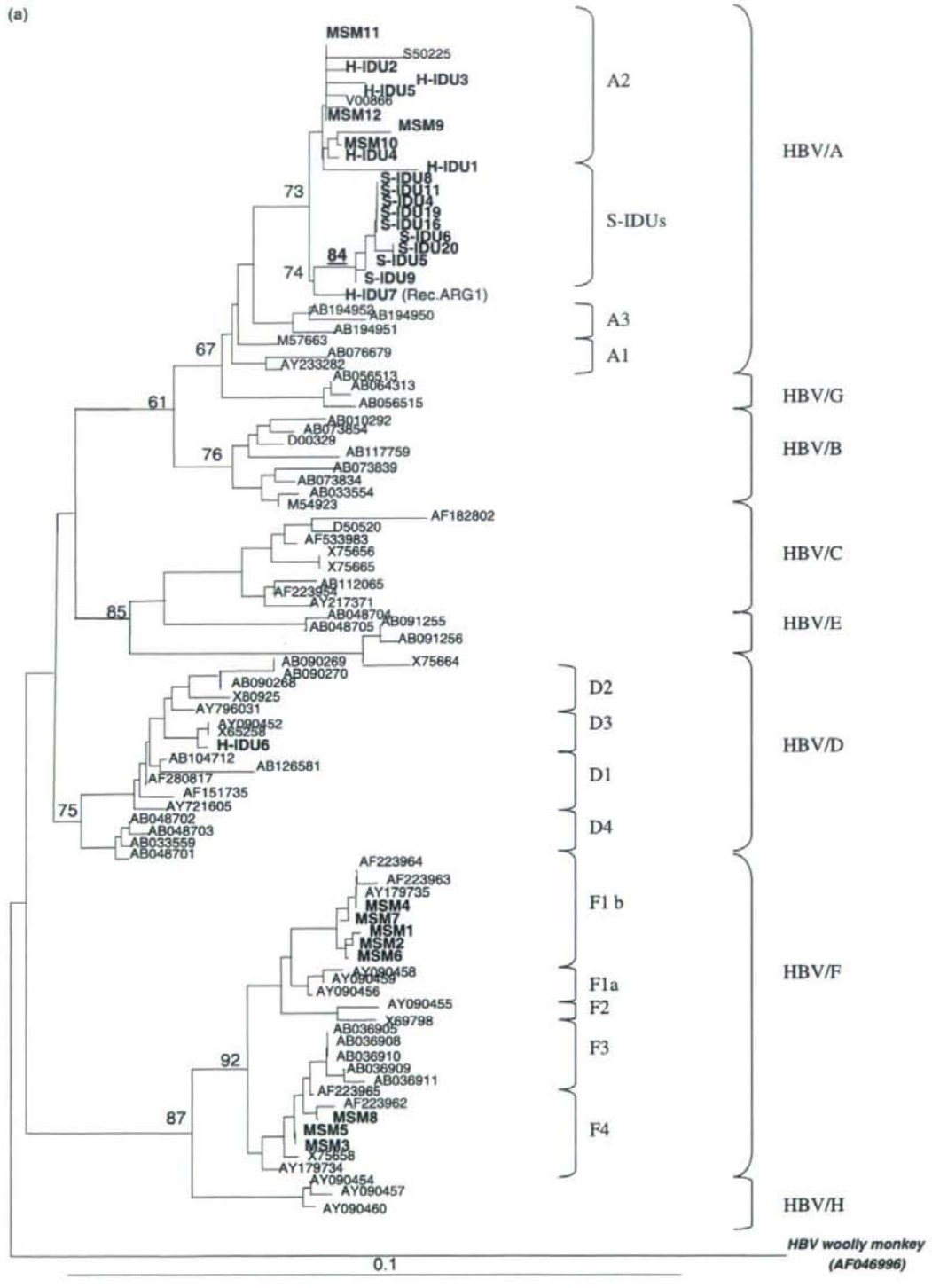
HBV pre-C/C region was also PCR amplified, sequenced and subjected to phylogenetic analysis in 7 out of 15 samples (Table 1). All of them were assigned to HBV genotype A: one of the samples clustered together with the HBV/A1 strains, while the remaining six were included within the HBV/A2 subgenotype (trees available upon request).

Four out of the 15 samples with detectable DNA (samples S-IDU4, S-IDU8, S-IDU11 and S-IDU19) were further analysed, and their HBV full-length DNA sequence conclusively classified them within the HBV/A2 subgenotype (Fig. 1b).

Serological status and HBV genotypes of hospital-recruited IDUs (H-IDUs)

These 80 patients were divided into three groups according to the above-mentioned different HBV serological patterns. The first group included 13 IDUs (92.3% male; mean age ± SD = 34.9 years ± 12.3) who showed positivity for both HBsAg as well as total anti-HBc. The second one included 1 IDU (male; 41 years old) positive for HBsAg but negative for total anti-HBc. Finally, a third group encompassed 66 IDUs (83.3% male; 36.2 years ± 6.3) who belonged to the 'anti-HBc only' group. The serological status for HCV and HIV from all HBV DNA PCR positive samples is depicted in Table 3.

Fig. 1 A phylogenetic neighbour-joining tree constructed by using (a) partial HBV S gene and (b) full-length genome sequences. Strains isolated from Argentinean IDUs and MSM are indicated in bold. Sample MSM14 was not included because it corresponds to a mixed A/F infection. Bootstrap statistical analysis was performed by using 1000 data sets; these values are indicated in the tree roots. S-IDUs: street-recruited injecting drug users; H-IDUs: hospital-recruited injecting drug users; MSM: men who have sex with men.



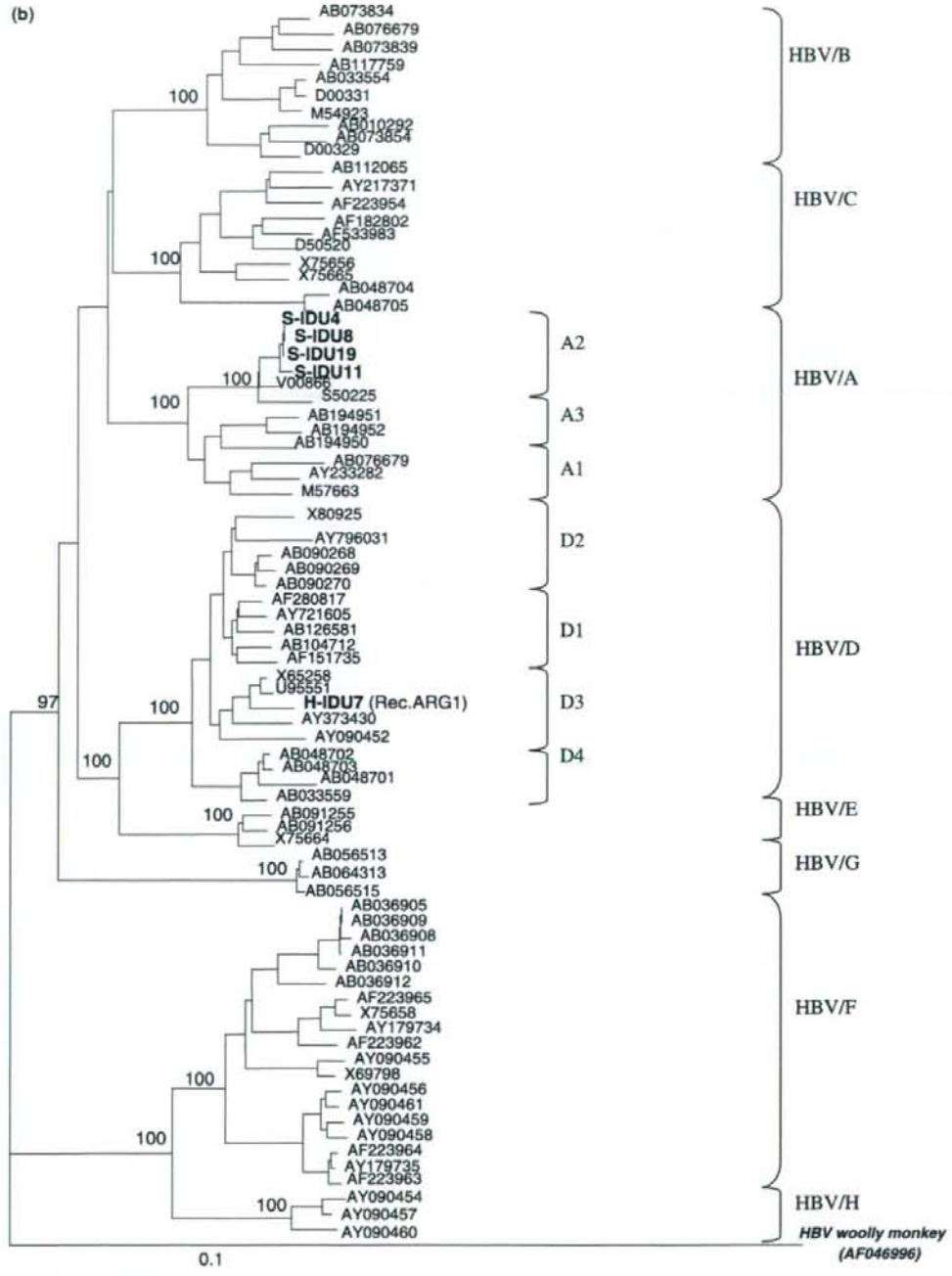


Fig. 1 Continued.

Table 2 Comparison of the amino acid residues of the S gene region – encompassing positions 100 to 210 – of HBV/Au (HBV genotype A, subgenotype unidentified) isolates with amino acid sequences retrieved from the GenBank and assigned to HBV/A1, HBV/A2 and HBV/A3.

	S region (aa 100–210) Codon number†									
HBV isolates	103 I	122 (*d/y*)‡ R	127 P	134 F	159 A	160 (*w/r*)‡ K	166 G	168 A	207 S	209 L
Au										
S-IDU4	.	.	.	.	.	.	.	.	.	V
S-IDU5	.	.	.	.	.	.	A	.	.	.
S-IDU6	.	.	.	.	.	.	.	.	N	.
S-IDU8	.	.	.	.	.	.	.	.	.	V
S-IDU9	.	.	.	.	.	.	.	.	.	.
S-IDU11	.	.	.	.	.	.	.	.	.	V
S-IDU16	.	.	.	.	.	.	.	.	N	.
S-IDU19	.	.	.	.	.	.	.	.	.	V
S-IDU20	.	.	.	.	.	.	G/A	.	.	.
A1										
AB076679	M	K	.	.	V	.	A	V	N	.
M57663	M	K	.	.	.	.	A	V	N	.
AB076678	M	.	.	.	.	.	A	V	N	.
AB116089	M	K	.	.	.	.	A	V	N	.
AB116083	M	K	.	.	.	.	A	V	N	.
AB297625	M	K	.	.	.	.	A	V	N	V
AY233282	.	K	.	.	.	.	A	V	N	.
A2										
V00866	M	K	.	.	.	.	A	V	.	V
AY233280	M	K	.	.	.	.	A	V	.	V
AJ309369	M	K	.	.	.	.	A	V	.	V
AF536524	M	K	.	.	.	.	A	V	.	V
AY128092	M	K	.	.	.	.	A	V	.	V
AB014370	M	K	.	.	.	.	A	V	.	V
S50225	M	K	.	.	.	.	A	V	.	V
A3										
AB194951	M	.	.	.	.	.	A	V	N	.
AB194952	M	.	.	.	.	.	A	V	N	.
AB194950	M	.	.	.	.	.	A	V	N	.
AM184125	M	.	.	.	.	.	A	V	N	.
AM184126	M	.	.	.	.	.	A	V	N	.

Dots indicate amino acid identity. †Amino acid residue 1 refers to the first amino acid coded by the S gene in genotype A sequences. ‡These two columns show key amino acidic positions for subtype assignment.

Both HBV S and pre-C/C coding regions were amplified and successfully sequenced in 7 out of 80 samples (8.7%): 6 samples from the first group (46.1%), none from the second group (0%) and 1 from the third (1.5%) (Table 3).

In six out of these seven samples, phylogenetic trees constructed by using both the S and pre-C/C nucleotide sequences revealed no discrepancy in genotype assignment between the two analysed coding regions. Five samples clustered within the subgenotype A2 strains and one within subgenotype D3 isolates (Table 3, Fig. 1a).

The S gene of the remaining sample (H-IDU7) was partly amplified and assigned to genotype D by RFLP. However, when the PCR products were sequenced and subjected to phylogenetic analysis, the sample was placed far from genotype D branch and close – but not interspersed – to sequences assigned to subgenotype A2 (Fig. 1a). As a consequence of this discrepancy, the HBV S gene sequence was further analysed by the Simplot and Simmonic programs and results confirmed by a third method developed by one of the authors (P.D.G.; graphics available upon

Table 3 Major characteristics of the seven HBV PCR (+) hospital-recruited IDUs (H-IDUs) from Buenos Aires, Argentina

				Serological status					S gene genotype		
Sample	Gender	Age	Date of blood sample collection	HBsAg	Anti-HBc Ab	Anti-HCV Ab	Anti-HIV Ab	Sexual orientation	RFLP	DNA sequencing (serotype)	Pre-C/C region genotype*
H-IDU1	M	33	June 2001	+	+	+	+	Heterosexual	A	A2 (*adw2*)	A2
H-IDU2	M	33	October 2001	+	+	–	+	Bisexual	A	A2 (*adw2*)	A2
H-IDU3	M	30	April 2002	+	+	–	+	Heterosexual	A	A2 (*adw2*)	A2
H-IDU4	M	28	September 1999	+	+	+	+	Heterosexual	A	A2 (*adw2*)	A2
H-IDU5	M	23	February 2001	+	+	+	+	Bisexual	A	A2 (*adw2*)	A2
H-IDU6	M	35	December 2006	+	+	+	+	Heterosexual	D	D3 (*ayw3*)	D3
H-IDU7†	M	26	May 1995	–	+	+	+	Heterosexual	D	A2 (*adw2*)	D3

IDUs, injecting drug users; Pre-C/C, Precore/core. *Genotypes were assigned by DNA sequencing. †Full-length HBV sequence was obtained.

request). An HBV D/A recombination event was clearly shown at the S gene. Consequently, this genomic recombination fully justified the RFLP assignment as a genotype D sample.

Furthermore, the full-length HBV genome sequence of this sample was amplified and subsequently sequenced. The phylogenetic analysis revealed that the complete genome sequence belonged to subgenotype D3 (Fig. 1b). However, evidence of recombination between HBV/A and HBV/D was confirmed by implementing a bootscanning analysis, as observed in Fig. 2. Plots confirming the recombination event were observed after performing genomic analysis with the Simmonic program and the method developed by the authors (P.D.G.; both graphics, available upon request). This strain-named Rec.ARG1- exhibited a recombinant HBV/A2 segment which corresponded to the nucleotide positions 147 to 636 of the S gene inserted in a backbone corresponding to HBV/D3.

HBV genotypes in MSM

Serum samples from 20 MSM (mean age ± SD = 30.9 years ± 6.2), who showed positivity for HBsAg as well as total anti-HBc, were analysed. HBV S gene was amplified in 13/20 (65%) samples and the pre-C/C region in 11/20 (55%) samples. Epidemiological data and the serological status for HCV and HIV from the HBV PCR positive MSM (70%) are shown in Table 4.

Phylogenetic analysis of the S and pre-C/C coding regions (Table 4, Fig. 1a) revealed that eight samples were assigned to genotype F (57.1%): five belonged to clade F1b and three to subgroup F4. Of the six remaining samples, five clustered together with the subgenotype A2 strains (35.7%) and one sample exhibited a mixed A/F infection (7.1%), as detected by RFLP (available upon request).

DISCUSSION

This is the first study carried out in Latin America which analyses and characterizes HBV DNA isolated from two groups with different risk factors: IDUs sharing needles and syringes and MSM exhibiting high-risk sexual behaviour.

Three major topics deserve to be analysed: (i) the observation of two simultaneous HBV epidemics involving different HBV genotypes among IDUs in comparison with MSM, HBV chronic carriers and blood donors (the latter two studied by Telenta *et al.* [6], and by França *et al.* [22], respectively, and herein used as a preliminary approach to the HBV molecular epidemiology within the general population, whose data are still unknown); (ii) the complete characterization of a novel D/A recombinant strain; and (iii) the observation of isolates clustering in a previously nonreported divergent genotype A clade at the S gene.

HBV molecular epidemiology data regarding the genotype prevalence among IDUs are still scarce and the few reports available are widely divergent [23–31]. Taking into account that several of these studies concerning HBV typing were performed at different time periods, it is unclear whether the reported association of genotype D and IDUs [25–31] reflects a differential route of transmission, a genotype shifting in a given area when two periods are compared or – alternatively – a mere reflection of the genotype circulating in the general population at a given time.

The high predominance of HBV/A among epidemiologically related (street-recruited) and unrelated (hospital-recruited) IDUs from Buenos Aires is in sharp contrast to its lower prevalence among three different populations residing in the same area: HBV carriers studied in the middle nineties [6] ($P = 0.0066$), blood donors examined during this decade [22] ($P = 0.0006$) and the MSM group with high-risk sexual behaviours included in the present

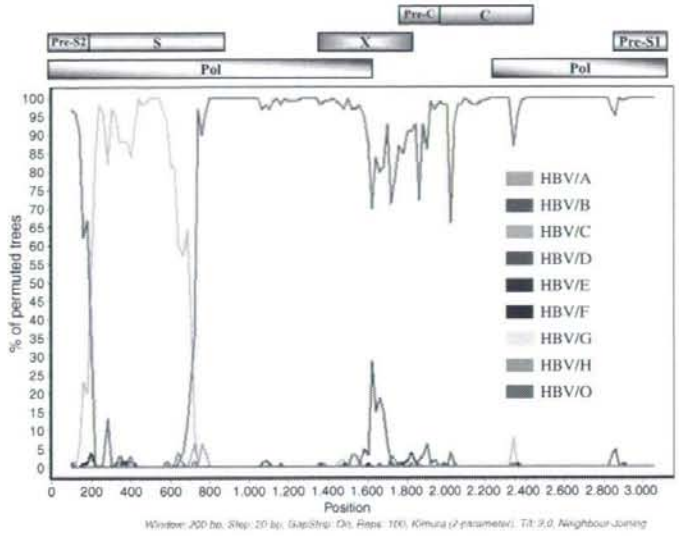


Fig. 2 Bootscanning analysis of the full-length genome of query sequence Rec.ARG1 to representative sequences of putative parental genotypes (A and D), other genotypes (B–C and E–H) and outgroup (O). Graph was generated using SIMPLOT (version 3.5.1) with a window size of 200 bp, step size of 20 bp, gap-strip off, Kimura and transition/transversion ratio: 2. Rec.ARG1 exhibits a recombinant HBV/A2 segment which corresponds to the nucleotide positions 147 to 636 of the S gene inserted in a backbone assigned to HBV/D3. GenBank accession numbers for the representative sequences of the eight HBV genotypes included in the phylogenetic analyses, are as follows: genotype A (AB076679, M57663, AB076678, AB116089, AB116083, AF297625, AY233282, V00866, AY233280, AJ309369, AF536524, AY128092, AB014370, S50225, AB194951, AB194952, AB194950, AM184125 and AM184126); genotype B (D00329, AB010292, AB073854, X97850, AB073839, AB076679, AB033554, AB033555, D00331 and AB117759); genotype C (AB112065, AY217371, AB112348, AF182802, D50520, AF533983, X75656, X75665, AB048704 and AB048705); genotype D (AB104712, AF151735, AF280817, AY721605, AB126581, X80925, AY796031, AB090268, AJ344117, AB090269, AB090270, AY090452, X65258, U95551, AB048701, AB048702, AB048703 and AB033559); genotype E (X75664, AB091255 and AB091256); genotype F (AY090456, AY090458, AY090459, AY090461, AF223964, AF223963, AY179735, AY090455, X69798, AB036905, AB036908, AB036909, AB036910, AB036911, X75663, AF223965, AY179734, X75658 and AB166850); genotype G (AB056513, AB056515 and AB064313); and genotype H (AY090460, AY090454 and AY090457). The full-length genome of the woolly monkey hepadnavirus, the most divergent primate hepadnavirus, was included as an outgroup sequence (AF046996).

study ($P = 0.0137$). Genotype G coinfection among S-IDUs samples was ruled out by using G-specific primers for PCR amplification (data not shown; [32]). As previously stated, the serum samples from all groups were collected throughout the same period of time. Thus, such dissimilar prevalence should not be ascribed to a genotypic shift among samples obtained at significantly distant dates. Moreover, since genotype D and F – but not A – appeared to predominate among blood donors in Buenos Aires city, it is suggested that HBV/A predominance among IDUs could not be considered a mere reflection of HBV strains circulating in the general population. At least two possibilities should be taken into account: (i) two independent HBV epidemics occurring simultaneously in Buenos Aires and related to the route of transmission; or (ii) a putative association between a given HBV genotype and an additional factor (HIV-1 and/or HCV co-infections, occult HBV infection, etc.).

Interestingly, earlier studies on HIV-1/HBV coinfected subjects showed that HBV/A predominates among Japanese [33] and Spanish MSM [28,29] while HBV/D prevails among Spanish IDUs [28,29]. However, in our study no statistically significant difference in HBV genotype prevalence was found when either high-risk subjects coinfected with HIV-1 ($n = 16$ for IDUs and $n = 4$ for MSM) were compared to those who were seronegative for the latter ($n = 6$ for IDUs and $n = 10$ for MSM). Moreover, no difference between genotype prevalence was observed after analysing HCV coinfected and HCV noninfected HBV patients nor after the presence or absence of occult infections was recorded. Furthermore, no statistically significant difference in HBV genotype distribution was observed between

Table 4 Major characteristics of the 14 HBV PCR (+) MSM from Buenos Aires, Argentina

Sample	Age	Date of blood sample collection	Number of sexual partners*	Sexual intercourse with men from a foreign country?	Serological status				S/Core PCR	Genotype† (serotype)
					HBsAg	Anti-HBc Ab	Anti-HCV Ab	Anti-HIV Ab		
MSM1	30	February 2003	11–50	Germany	+	+	–	–	+/–	F1b (*adw4*)
MSM2	35	February 2003	6–10	Spain	+	+	–	–	+/+	F1b (*adw4*)
MSM3	41	March 2003	6–10	No	+	+	–	–	+/–	F4 (*adw4*)
MSM4	26	April 2003	6–10	Brazil	+	+	–	+	+/+	F1b (*adw4*)
MSM5	21	May 2003	1	Paraguay, Brazil and USA	+	+	–	+	+/–	F4 (*adw4*)
MSM6	36	July 2003	2–5	No	+	+	–	–	+/+	F1b (*adw4*)
MSM7	23	October 2003	0	No	+	+	–	–	+/+	F1b (*adw4*)
MSM8	35	November 2003	11–50	No	+	+	–	–	+/+	F4 (*adw4*)
MSM9	33	November 2003	1	No	+	+	–	+	+/+	A2 (*adw2*)
MSM10	22	April 2003	2–5	USA	+	+	–	+	+/+	A2 (*adw2*)
MSM11	19	August 2003	1	No	+	+	–	–	+/+	A2 (*adw2*)
MSM12	35	August 2003	1	Australia	+	+	–	–	+/+	A2 (*adw2*)
MSM13	32	August 2003	2–5	No	+	+	–	–	–/+	A2
MSM14	34	September 2003	2–5	Colombia and USA	+	+	–	–	+/+	Mixed A / F

MSM, men who have sex with men. *In the last 6 months; †Genotypes were assigned by DNA sequencing. No discrepancies in genotyping results were observed between RFLP and both S gene as well as pre-C/C region sequencing.

MSM subjects and blood donors. As a whole, these data suggest that two simultaneous HBV epidemics took place in Buenos Aires city: one blood-borne infection associated with the HBV/A genotype and spreading among IDUs, and another one sexually transmitted among the homosexual and heterosexual population.

Interestingly, recent molecular studies have revealed two simultaneous epidemics caused by different HIV-1 subtypes in Buenos Aires [34]. Noteworthy, the serum samples analysed by Avila *et al.* [34] were collected in the same period of time as the samples characterized in the present study. Therefore, in contrast to what was previously stated for HIV-1, HBV showed similar patterns of infection in the homosexual and heterosexual populations from Buenos Aires between 2000 and 2003.

There is fairly general agreement in considering HCV-infected patients as those with the highest prevalence of occult HBV infection. This type of persistent infection mostly appears due to a strong suppression of HBV replication and gene expression probably induced by HCV 'core' protein [35]. In our study, 81.25% of the recruited IDUs were included in the 'anti-HBc only' group, being 86.15% of them coinfected with HCV. HBV DNA was detected in 10 samples from the 'anti-HBc only' group, which implies that the prevalence of occult HBV infection among Argentine IDUs is 7.7%. Moreover, among them, a novel intergenotypic D/A recombinant strain (showing different breakpoints from all D/A recombinants already documented) was isolated from a HCV coinfected patient. This is the first full-length HBV D/A recombinant genome characterized from the American continent.

At the S gene region, all sequenced street-recruited IDUs samples were genotyped as HBV/A, although subgenotype assignment was not possible (HBV genotype A, subgenotype unidentified). Probably due to the snowball recruitment technique employed, data suggest that this group of patients behaves as a closed community since no HBV/A isolate obtained from unrelated hospital-recruited IDUs ($n = 5$) exhibited the same degree of nucleotide divergence at the S gene.

To obtain a complete characterization of these nine Au isolates, the full-length genome from four of them and the pre-C/C region from three strains could be sequenced, ascribing six isolates to subgenotype A2. However, the remaining strain clustering within the Au sequences at the S gene, grouped within the A1 sequences at the pre-C/C region as confirmed by phylogenetic analysis (available upon request), the presence of nucleotide substitutions T1809 and T1812 considered to be characteristic of HBV/A1 [36], and a (mean ± SD) 98.1 ± 0.01 nucleotide identity when compared to A1 sequences ($n = 7$), in contrast to 97.3 ± 0.008 and 97.6 ± 0.006 with respect to A2 ($n = 7$) and A3 ($n = 5$) sequences, respectively.

As a whole, these data might suggest that a recombination event involving a previously unrecognized HBV/A sequence and an A2 or A1 genome could have arisen among S-IDUs at the S coding region. Moreover, these unusual strains showed the serological subtype *ayw*1, which is considered to be rare among isolates assigned to genotype A [1]. With the exception of the A3 isolates and a few A1 strains from Africa and the Philippines, which are *ayw*1, the HBV/A2 isolates exhibit all serotype *adw*2. Therefore, these unique Argentine samples

represent the first strains classified as genotype A encoding serotype *ayw*1 in the American continent.

HBV/A when compared to other genotypes may be associated with higher risk to develop chronic carrier state after acute infection in adults (discussed in [37]). Particularly, this was explained by differences in the replicative ability of different genotypes [38]. Persistence of the viral infection may also be associated with the dose of virus exposed to the host immune system [39]. Obviously, intravenous injection and sexual contact result in different HBV infection doses. These factors may have contributed to the coexistence of two simultaneous and independent epidemics in Buenos Aires, suggesting that some HBV genotypes might be predominantly associated to different risk factors/behaviours, as observed among IDUs and MSM populations in Argentina. Further clinical and experimental studies are required to determine an eventual association of certain HBV genotypes with a particular transmission route and the development of the chronic carrier stage in adults. This should shed some light in the development of both efficient vaccination and treatment strategies for populations where different genotypes are distributed.

ACKNOWLEDGEMENTS

The authors thank to FUNDAI staff, Rubén Marone, Psy.D. (Nexo Asociacon Civil) and Prof. Diana Rossi (Intercambios Asociación Civil) for their assistance during the collection of the serum samples and epidemiological data, and to Jorge Rey, M.D. (Depto. de Hemoterapia e Inmunohematología, Htal. de Clínicas "José de San Martín", UBA) for performing the serological assays. Also, we would like to thank Maria Victoria Illas for manuscript preparation.

This study was funded in part by PICT 10871/02 and 440/06 (ANPCyT), UBACyT M057 (University of Buenos Aires), PIP 6065 and 6118 (CONICET), Pan American Health Organization of the World Health Organization, 2000–2001.

REFERENCES

1 Kramvis A, Kew M, François G. Hepatitis B virus genotypes. *Vaccine* 2005; 23: 2409–2423.
2 Schaefer S. Hepatitis B virus taxonomy and hepatitis B virus genotypes. *World J Gastroenterol* 2007; 13: 14–21.
3 Huy TTT, Ushijima H, Sata T, Abe K. Genomic characterization of HBV genotype F in Bolivia: genotype F subgenotypes correlates with geographical distribution and T^{1858} variant. *Arch Virol* 2006; 151: 589–597.
4 Livingston SE, Simonetti JP, McMahon BJ *et al.* Hepatitis B virus genotypes in Alaska native people with hepatocellular carcinoma: preponderance of genotype F. *J Infect Dis* 2007; 195: 5–11.
5 Mbayed VA, Barbini L, López JL, Campos RH. Phylogenetic analysis of the hepatitis B virus (HBV) genotype F including Argentine isolates. *Arch Virol* 2001; 146: 1803–1810.
6 Telenta PFS, Palacios Poggio G, López JL, Gonzalez J, Lemberg A, Campos RH. Increased prevalence of genotype F hepatitis B virus isolates in Buenos Aires, Argentina. *J Clin Microbiol* 1997; 35: 1873–1875.
7 Kramvis A, Kew MC. Relationship of genotypes of hepatitis B virus to mutations, disease progression and response to antiviral therapy. *J Viral Hepat* 2005; 12: 456–464.
8 Weissenbacher M, Rossi D, Radulich G *et al.* High seroprevalence of bloodborne viruses among street-recruited injection drug users from Buenos Aires, Argentina. *Clin Infect Dis* 2003; 37: S348–S352.
9 Segura M, Sosa Estani S, Marone R *et al.* Buenos Aires cohort of men who have sex with men: recruitment, retention and seroincidence of HIV and other sexually-transmitted infections. *Proceedings of the Third IAS Conference on HIV Pathogenesis and Treatment*. Rio de Janeiro, Brazil, 24–27 July 2005.
10 Zeng GB, Wen SJ, Wang ZH, Yan L, Sun J, Hou JL. A novel hepatitis B virus genotyping system by using restriction fragment length polymorphism patterns of S gene amplicons. *World J Gastroenterol* 2004; 10: 3132–3136.
11 Lindh M, Andersson AS, Gusdal A. Genotypes, nt 1858 variants, and geographic origin of hepatitis B virus – large-scale analysis using a new genotyping method. *J Infect Dis* 1997; 175: 1285–1293.
12 Birkenmeyer LG, Mushahwar IK. Detection of hepatitis A, B and D virus by the polymerase chain reaction. *J Virol Methods* 1994; 49: 101–112.
13 Sugauchi F, Mizokami M, Orito E *et al.* A novel variant genotype C of hepatitis B virus identified in isolates from Australian Aborigines: complete genome sequence and phylogenetic relatedness. *J Gen Virol* 2001; 82: 883–892.
14 Borst A, Box ATA, Fluit AC. False-positive results and contamination in nucleic acid amplification assays: suggestions for a Prevent and Destroy strategy. *Eur J Clin Microbiol Infect Dis* 2004; 23: 289–299.
15 Kwok S, Higuchi R. Avoiding false positives with PCR. *Nature* 1989; 339: 237–238.
16 Felsenstein J. PHYLIP Inference Package Version 3.5. Departments of Genetics, University of Washington, Seattle, 1993.
17 Lole KS, Bollinger RC, Paranjape RS *et al.* Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 1999; 73: 152–160.
18 Simmonds P, Midgley S. Recombination in the genesis and evolution of hepatitis B virus genotypes. *J Virol* 2005; 79: 15467–15476.
19 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. *Nucl Acids Res* 1994; 2: 4673–4680.
20 Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl Acids Res* 1997; 25: 4876–4882.
21 Robertson DL, Hahn BH, Sharp PM. Recombination in AIDS viruses. *J Mol Evol* 1995; 40: 249–259.

22 França PH, González JE, Munné MS *et al.* Strong association between genotype F and hepatitis B virus (HBV) e antigen-negative variants among HBV-infected Argentinean blood donors. *J Clin Microbiol* 2004; 42: 5015–5021.
23 Datta S, Banerjee A, Chandra PK, Mahapatra PK, Chakrabarti S, Chakravarti R. Drug trafficking routes and hepatitis B in injecting drug users, Manipur, India. *Emerg Infect Dis* 2006; 12: 1954–1957.
24 Lin CL, Liu CJ, Chen PJ, Lai MY, Chen DS, Kao JH. High prevalence of occult hepatitis B virus infection in Taiwanese intravenous drug users. *J Med Virol* 2007; 79: 1674–1678.
25 Swenson PD, Van Geyt C, Alexander ER *et al.* Hepatitis B virus genotypes and HBsAg subtypes in refugees and injection drug users in the United States determined by LiPA and monoclonal EIA. *J Med Virol* 2001; 64: 305–311.
26 Christensen PB, Krarup HB, Niesters HGM *et al.* Outbreak of hepatitis B among injecting drug users in Denmark. *J Clin Virol* 2001; 22: 133–141.
27 Lindh M, Horal P, Norkrans G. Acute hepatitis B in Western Sweden – genotypes and transmission routes. *Infection* 2000; 28: 161–163.
28 Pérez-Olmeda M, Núñez M, García-Samaniego J, Ríos P, González-Lahoz J, Soriano V. Distribution of hepatitis B virus genotypes in HIV-infected patients with chronic hepatitis B: therapeutic implications. *AIDS Res Hum Retroviruses* 2003; 19: 657–659.
29 Ramos B, Núñez M, Martín-Carbonero L *et al.* Hepatitis B virus genotypes and lamivudine resistance mutations in HIV/Hepatitis B virus-coinfected patients. *J Acquir Immune Defic Syndr* 2007; 44: 557–561.
30 van Steenbergen JE, Niesters HG, Op de Coul EL *et al.* Molecular epidemiology of hepatitis B virus in Amsterdam 1992–1997. *J Med Virol* 2002; 66: 159–165.
31 Alam MM, Zaidi SZ, Shaukat S *et al.* Common Genotypes of Hepatitis B virus prevalent in injecting drug abusers (addicts) of North West Frontier Province of Pakistan. *Virol J* 2007; 4: 63–68.
32 Kato H, Orito E, Sugauchi F *et al.* Determination of hepatitis B virus genotype G by polymerase chain reaction with hemi-nested primers. *J Virol Methods* 2001; 98: 153–159.
33 Koibuchi T, Hitani A, Nakamura T *et al.* Predominance of genotype A HBV in an HBV-HIV-1 dually positive population compared with an HIV-1-negative counterpart in Japan. *J Med Virol* 2001; 64: 435–440.
34 Avila MM, Pando MA, Carrion G *et al.* Two HIV-1 epidemics in Argentina: different genetic subtypes associated with different risk groups. *J Acquir Immune Defic Syndr* 2002; 29: 422–426.
35 Chen SY, Kao CF, Chen CM *et al.* Mechanisms for inhibition of hepatitis B virus gene expression and replication by hepatitis C virus core protein. *J Biol Chem* 2003; 278: 591–607.
36 Tanaka Y, Hasegawa I, Kato T *et al.* A case-control study for differences among hepatitis B virus infections of genotypes A (subtypes Aa and Ae) and D. *Hepatology* 2004; 40: 747–755.
37 Ozasa A, Tanaka Y, Orito E *et al.* Influence of genotypes and precore mutations on fulminant or chronic outcome of acute hepatitis B virus infection. *Hepatology* 2006; 44: 326–334.
38 Sugiyama M, Tanaka Y, Kato T *et al.* Influence of hepatitis B virus genotypes on the intra- and extracellular expression of viral DNA and antigens. *Hepatology* 2006; 44: 915–924.
39 Bocharov G, Ludewig B, Bertoletti A *et al.* Underwhelming the immune response: effect of slow virus growth on CD8 + -T-lymphocyte responses. *J Virol* 2004; 78: 2247–2254.

Available online at www.sciencedirect.com
ScienceDirect
Virus Research 132 (2008) 174–180
Virus Research
www.elsevier.com/locate/virusres

Transmission of hepatitis B virus (HBV) genotypes among Japanese immigrants and natives in Bolivia

Anis Khan [a], Yasuhito Tanaka [a], Hidetsugu Saito [b], Hirotoshi Ebinuma [b], Hitoshi Sekiguchi [c], Hideaki Iwama [d], Go Wakabayashi [e], Toshio Kamiya [f], Fuat Kurbanov [a], Abeer Elkady [a], Masashi Mizokami [a,*]

[a] *Department of Clinical Molecular Informative Medicine, Nagoya City University, Graduate School of Medical Sciences, Kawasumi 1, Mizuho, Nagoya 467-8601, Japan*
[b] *Department of Medicine, Keio University, 35 Shinano-machi, Shinjuku-ku, Tokyo 160-8582, Japan*
[c] *Department of Laboratory Medicine, Keio University Hospital, 35 Shinano-machi, Shinjuku-ku, Tokyo 160-8582, Japan*
[d] *International Medical Association, Keio University, 35 Shinano-machi, Shinjuku-ku, Tokyo 160-8582, Japan*
[e] *Department of Surgery, Iwate Medical College, Iwate Medical University School of Medicine, 19-1 Uchimaru, Morioka 020-8505, Japan*
[f] *Clinica Kamiya, Av Mons. Rivero 265, Santa Cruz, Bolivia*

Received 16 November 2007; received in revised form 6 December 2007; accepted 6 December 2007
Available online 22 January 2008

Abstract

Hepatitis B virus genotypes are associated with transmission pattern, virological and clinical features and outcome of the chronic infection course. HBV genotypes other than Genotype F (HBV/F) are considered a reflection of human migration into South America. A total of 487 individuals in Bolivia, including Japanese immigrants ($n = 287$) and natives ($n = 200$), were screened for HBV serological markers. Overall 22/487 (4.5%) of the subjects were positive for HBsAg, 217/487 (44.5%) for anti-HBc and 162/487 (33.3%) for anti-HBs. Genotypes were determinable in 22 cases by EIA, followed by sequencing and phylogenetic analysis in 17 cases. HBV genotype distribution in Japanese and Bolivians was HBV/F (4 and 8); HBV/C (5 and 3); and HBV/B (1 and 1), respectively. Phylogenetic analyses of nine complete and eight partial (HBsAg/pre-core/core region) genomes, revealed that HBV/F strains cluster with previously reported regional strains, whereas HBV/B and HBV/C strains belonged to Asian subgenotype B2 (Ba) and C2 (Ce), respectively. Japanese immigrants might have introduced HBV/B and HBV/C to natives in Bolivia, conversely, exposed to the indigenous HBV/F. This report provides evidence of an inter-communities transmission of HBV revealed by its genotypes. Further study is required to investigate peculiarities of the genotypes in different ethnic groups in Bolivia.

Keywords: Bolivians; Japanese immigrants; HBV; Genotypes; Transmission

1. Introduction

Reports indicate that about two billion people are exposed to HBV and 350 million of them have chronic infection around the world. Morbidity and mortality in chronic HBV infection is associated with development of liver cirrhosis (LC) and hepatocellular carcinoma (HCC). Populations is South and East Asia, sub-Saharan Africa, and Central and South America show particularly high frequencies of HBV infection that may be maintained through mother to child perinatal transmission (Chen et al., 2004) or horizontal transmission in childhood (Dumpis et al., 2001). Knowledge of hepatitis B infection and genotype distribution is necessary to design and evaluate the preventive measures, such as universal immunization.

HBV strains infecting humans show antigenic and genetic heterogeneity and are currently classified into eight genotypes that differ in nucleotide sequence by >8% and subgenotypes by >4%. These genotypes have geographical distribution: A and D have global distributions; genotypes B and C are found predominantly in East and Southeast Asia; genotype E in West Africa; and genotypes F and H are considered indigenous to Central and South America (Miyakawa and Mizokami, 2003; Norder et

* Corresponding author. Tel.: +81 52 853 8292; fax: +81 52 842 0021.
E-mail address: mizokami@med.nagoya-cu.ac.jp (M. Mizokami).

0168-1702/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.virusres.2007.12.005