

Methodology

HBVRegDB: Annotation, comparison, detection and visualization of regulatory elements in hepatitis B virus sequences

Nattanan Panjaworayan, Stephan K Roessner, Andrew E Firth and Chris M Brown*

Address: Department of Biochemistry, University of Otago, Dunedin, New Zealand

Email: Nattanan Panjaworayan - panna478@student.otago.ac.nz; Stephan K Roessner - stephan.roessner@gsf.de; Andrew E Firth - A.Firth@ucc.ie; Chris M Brown* - chris.brown@otago.ac.nz

* Corresponding author

Published: 17 December 2007

Received: 17 October 2007

Virology Journal 2007, 4:136 doi:10.1186/1743-422X-4-136

Accepted: 17 December 2007

This article is available from: <http://www.virologyj.com/content/4/1/136>

© 2007 Panjaworayan et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The many *Hepadnaviridae* sequences available have widely varied functional annotation. The genomes are very compact (~3.2 kb) but contain multiple layers of functional regulatory elements in addition to coding regions. Key regions are subject to purifying selection, as mutations in these regions will produce non-functional viruses.

Results: These genomic sequences have been organized into a structured database to facilitate research at the molecular level. HBVRegDB is a comparative genomic analysis tool with an integrated underlying sequence database. The database contains genomic sequence data from representative viruses. In addition to INSDC and RefSeq annotation, HBVRegDB also contains expert and systematically calculated annotations (e.g. promoters) and comparative genome analysis results (e.g. blastn, tblastx). It also contains analyses based on curated HBV alignments. Information about conserved regions – including primary conservation (e.g. CDS-Plotcon) and RNA secondary structure predictions (e.g. Alidot) – is integrated into the database. A large amount of data is graphically presented using the GBrowse (Generic Genome Browser) adapted for analysis of viral genomes. Flexible query access is provided based on any annotated genomic feature. Novel regulatory motifs can be found by analysing the annotated sequences.

Conclusion: HBVRegDB serves as a knowledge database and as a comparative genomic analysis tool for molecular biologists investigating HBV. It is publicly available and complementary to other viral and HBV focused datasets and tools <http://hbvregdb.otago.ac.nz>. The availability of multiple and highly annotated sequences of viral genomes in one database combined with comparative analysis tools facilitates detection of novel genomic elements.

Background

Hepatitis B virus (HBV) chronically infects about 350 million people worldwide and is a major contributor to liver pathology including hepatitis and carcinoma. A large number of strains, isolates and mutants of the *Hepadna-*

viridae family have been sequenced. For example, a search of Entrez for HBV complete genomes currently (9/2007) retrieves 1114 records, and the Hepatitis Virus Database (HVD) contains over 1000 full-length sequences. The small, just 3.2 kb, genome has been extensively studied –

with a PubMed search for 'HBV genome' resulting in over 2500 publications. This research has shown that the genome is highly packed with information in sequence and structure. This directs processes such as transcription, reverse transcription, replication, nuclear import and export and coding [1-7]. Regulatory elements control this at the DNA, RNA and protein levels, with particular bases known to participate in DNA and RNA elements and also encode more than one protein in alternative frames. During infection the mutation rate is high – estimated to be around 10⁻⁵ to 10⁻⁴ per base per year [8]. This results in a quasi-species infecting a single individual and may result in some DNA sequences from an individual not being representative of the 'fittest' species. Mutants may become prevalent in the population – for example, precore mutations, escape mutations, or antiviral resistance mutations.

Recently several international public databases containing significant hepadnaviral content have become available: the general Viral Reference Sequence genome project [9,10], Hepatitis Virus Database [11], SEQHEPB [12], and the HepSeq database [13]. Each has its own focus and utility. The viral RefSeq genome project is broad but includes 10 *Hepadnaviridae* members. It is searchable through Entrez Genomes and linked to other resources including the protein database, NCBI gMap and gene [10]. The HepSeq database is an epidemiological database focussing on epidemiological, clinical nucleotide sequence and mutational aspects of HBV infection [13]. The Hepatitis Virus Database includes HBV and provides information on genome location and phylogenetic relationships automatically processed from DDBJ [11]. SEQHEPB allows subscribers to analyze genotypes of HBV genomes, including key mutations associated with antiviral resistance [12]. However, there is no tool available to combine expert annotation with similarity search methods for molecular biological research into HBV [14-17].

We describe here a genome-based public domain database for the *Hepadnaviridae*. The database contains data on individual sequences and groups of sequences and facilitates comparative genomic analysis. The complexity of the HBV genome has challenged development of this resource but it will provide a model for other viruses.

Methods

Sequences for analysis

For more detail refer to the documentation in the database. Genome sequences of selected representative viruses of the *Hepadnaviridae* family were retrieved from NCBI. All retrieved Genbank files were split into fasta-formatted and gff-formatted files. As the virus genomes are circular, some of the parsed Genbank files were manually curated in order to be represented correctly.

Processing of data

Multiple sequence alignments were produced with ClustalW [18]. All files were then placed in the MySQL database HBVRegDB.

To identify conserved viral genomic regions, three blast queries (blastn, tblastx and blastx) were performed on RefSeq Virus release 24 with the parameters shown in Table 1. The results were reformatted to create a gff file and the names of the matched sequences were integrated to present them in a meaningful graphical representation. The database will be updated with RefSeq releases.

Results and discussion

Annotations on single viral genomes

The NCBI viral RefSeqs for *Hepadnaviridae* provide the best information about coding regions and protein sequences. However, in general they do not provide information on regulatory signals that are crucial for viral gene expression. Four NCBI taxonomic groupings (Figure 1) were incorporated into HBVRegDB.

Eight HBV genotypes (A-H) have been described that vary in up to 15% of bases. They have very similar genomic organization, but differing global prevalence. Infection has been suggested to result in different clinical outcomes, and this is presumably related to sequence variation [19-25]. Other HBV-like viruses also infect hominids (Hominid HBV, HHBV, ~20% nt divergence from HBV) and rodents (*Orthohepadnavirus*, OHV, ~45% nt divergence from HBV). Closely related viruses infect birds (*Avihepadnavirus*, AHV) with overall similar organization but significant sequence divergence (~60%). These have been used as models to investigate human HBV [26].

As part of our experimental research, a complete HBV genome *adw*, genotype A, derived from a Taiwanese HBV-infected patient was sequenced (a gift from M-H Lin, National Taiwan University). This HBV clone was known to produce viable HBV particles when transfected into

Table 1: The blast parameters used to perform the BLAST queries.

Parameter	Meaning	blastn	tblastx
-e	Expectation value E	100 (10.0)	10(10)
-q	Penalty for nucleotide mismatch	-1 (-3)	-
-r	Reward for nucleotide match	1 (1)	-
-E	Cost to extend a gap	-1 (-2)	-2(1)
-G	Cost to open a gap	-2 (-5)	-8(11)
-W	K-tuple size	7 (11)	2(3)

RefSeq Virus was chosen as the target database as the redundancy in Genbank makes the top matches less informative. The penalty for a mismatch, the cost to open and to extend a gap has been specifically set to detect distant similarity matches. BLOSUM62 was used for tblastx.

Viruses

Virus sequences analysed		
Viruses analysed here with links to ICTV and NCBI taxonomy		
Name	Family/Genus	Taxid
Hepadnaviridae	Family 00.030.0	10404
Orthohepadnavirus, (OHV) mammalian hepatitis B-type viruses	Genus 00.030.0.01	10405
Hepatitis B virus (HBV)	Species 00.030.0.01.001	10407
Avihepadnavirus (AHV) avian hepatitis B-type viruses	Genus 00.030.0.02	10437

Figure 1
Screenshot of a table indicating genomic sequences analyzed in HBVRegDB.

cells [27]. This sequence was highly annotated and submitted to INSDC via EMBL (EMBL ACC: AM282986; Figure 2). The annotation was done by extracting biological information from the literature or other sequence records, based on the functional conservation. This sequence has the most complex annotation per nucleotide (3–13 annotations per base) of any sequence in the public sequence database. It represents the limits of this type of annotation and of parsers implemented to interpret it. It was intended to annotate most experimentally proven features on the sequence. This will lead to annotation of features that may be of lesser importance (e.g. the S protein myristoylation site [28]) or alternative splicing [29,30].

Sequences were numbered to begin at the EcoRI site position (if present). This was chosen because, for most sequences, the numbering is common until base ~1910, where the numbering diverges. An alternative logical numbering scheme is also used from position 1 in the pregenomic RNA [31]. Protein sequences are separately described using the standard numbering [32].

Annotated DNA elements

Two Direct Repeat DNA primer-binding sites – DR1 and DR2 – are involved in replication and may be preferential sites for viral integration into the host genome [4]. Promoters – preC/C promoter and TATA box [33], S1 promoter, S2 promoter, X promoter. Transcriptional regulatory elements – NRE α , NRE β and NRE γ . (reviewed in [4]. Enhancers – Enh I and Enh II, which modulate mRNA synthesis. The functions of Enh1 and EnhII were demonstrated for the HBV *ayw* subtype [4]. Protein binding sites. DNA binding sites within the central core domain of Enh I. Binding sites of C/EBP, p53, IRF α , HNF3, HNF4, RFX1, AP1, NF1, CREAB, ATF2, RXR:PPAR and COUP1 (reviewed in [34]. An element within Enh II, box α , which is essential for function of the enhancer *in vivo*. The non-canonical polyadenylation (TATAAA) signal

used by all transcripts [33,35] followed by the poly (A) cleavage site (nucleotide 1930). An indicative variation which represents a nucleotide transition from 'A' to 'G' at nucleotide position 1896 changing a preC tryptophan to a termination codon [36]. There are many functional and non-functional variants of HBV and it is not the focus of this database to show them; this is done by existing databases – e.g. HepSEQ and SEQHEPB [12,13].

Annotated RNA elements

Five mRNAs – preC, pgRNA, S1, S2, and X, all ending at the common poly (A) cleavage site including alternative splice variants of these transcripts [27,37]. RNA regulatory elements – post-transcriptional regulatory element (PRE; reported to be an important RNA export element [38–40]), splicing regulatory element (SRE) 1–3 [41], conserved stem-loop structures within the HBV PRE, PRE HSL α , PRE HSL β [38], and the critical RNA epsilon element structure required for replication and packaging [42].

Annotated protein coding sequences

Eight CDS were annotated on the sequence – preC, C, P, X, large S, middle S, small S and C0. C0 is a small CDS not annotated on most HBV genomes. It is involved in regulation of translation of the P and C CDSs and is conserved in all HBV genotypes [31]. Protein domains: P – Terminal protein, Spacer, Reverse transcriptase, RNase H. [32].

This highly annotated nucleotide sequence can be downloaded from HBVRegDB in formats designed for use in software that will read Genbank format. A number of the most sophisticated parsers were tested by directly retrieving the entry from an INSDC database (NCBI Genome Browser, Artemis, Apollo (free), VectorNTI (free for academics)). These had differing levels of ability to represent complex annotation, with features (e.g. the P CDS) crossing the origin of a circular genome and complex descriptors (e.g. mRNA, alternative splices) parsed more or less

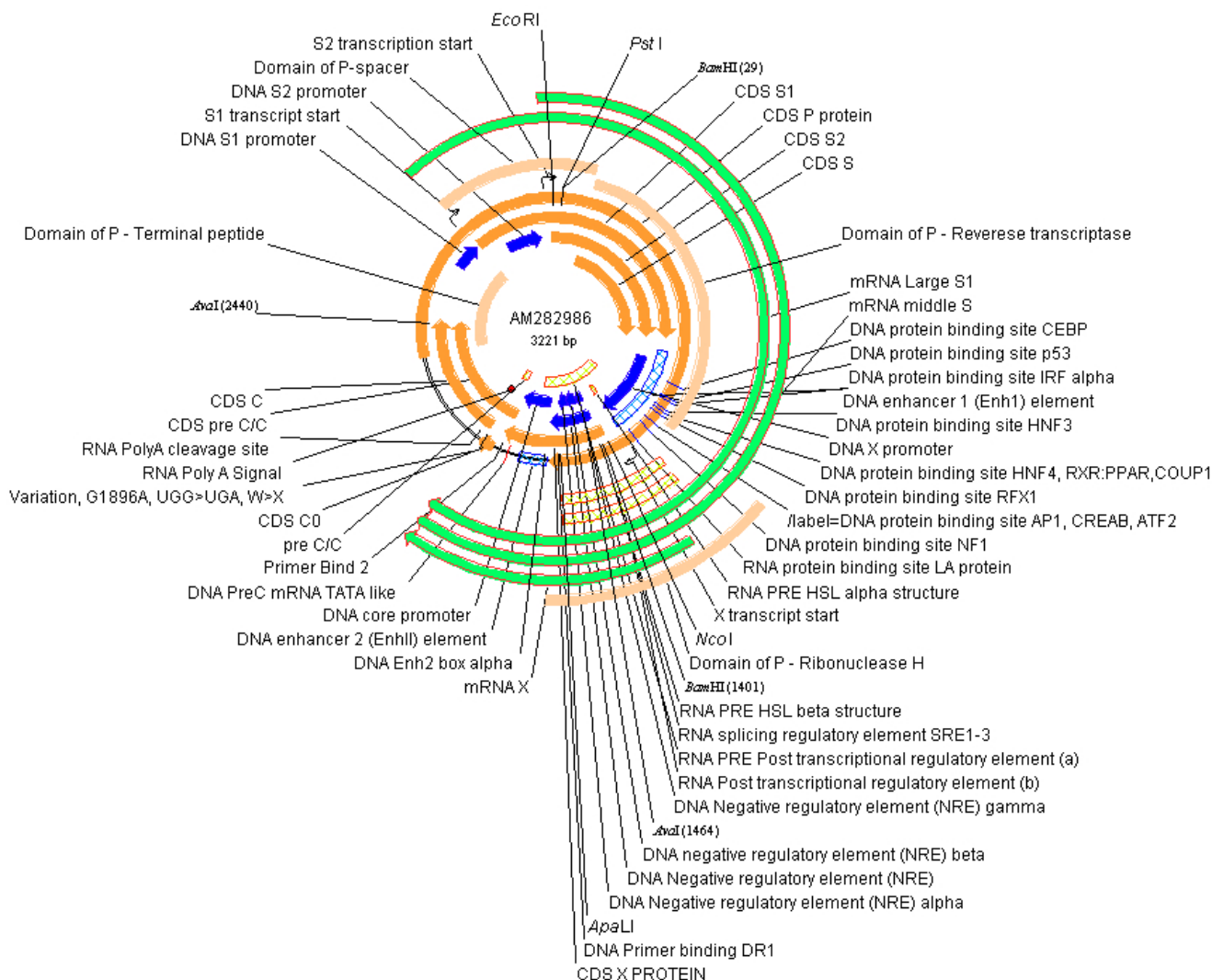


Figure 2
The highly annotated reference sequence (AM282986) in Genbank format visualized by VectorNTI.

well. In HBVRegDB we provide two slightly modified annotations of this HBV genome. One for more accurate circular parsing into VectorNTI, and another for linear browsers (e.g. GBrowse, Argo). A graphical representation of this annotated sequence in VectorNTI is shown in Figure 2. Although it can represent circular genomes, this format becomes difficult to interpret with many annotations.

HBVRegDB provides a tool to map these annotations onto another HBV sequence by performing a pairwise alignment.

HBV, rodent and avian hepadnaviral RefSeq genomes

Key additional regulatory elements were added to HBV genotype C (RefSeq NC_003977), Woodchuck RefSeq (WHV; NC_004107) and Duck HBV RefSeq

(NC_001344). These modified sequences are indicated by 'm' e.g. NC_003977m. The additional features for WHV include woodchuck post-transcriptional regulatory element (WPRE), which is reported to enhance gene expression delivered by retroviral vectors for gene therapy [43], and WRE α , WRE β and WRE γ , whose sequences are conserved within the mammalian hepadnaviruses and are essential for WRE function [44].

**Annotations on multiple sequence alignments
HBV_HBVRegDB_32**

This is an annotation of the 23 NCBI genotyping sequences, with other members of genotypes B-F added from [32]. The most highly annotated sequences: NC_003977m RefSeq (genotype C) and AM282986 (genotype A) are included in the alignment. This alignment is

available in VectorNTI format (apr) with annotation (e.g. Figure 3) and also in formats that cannot be automatically annotated (msf and aln).

HHBV_HBVRegDB_21

Hominid HBV. This is an alignment of human, gibbon, gorilla, chimp and orang-utan HBV genomes from [17], NC_0003977m group.

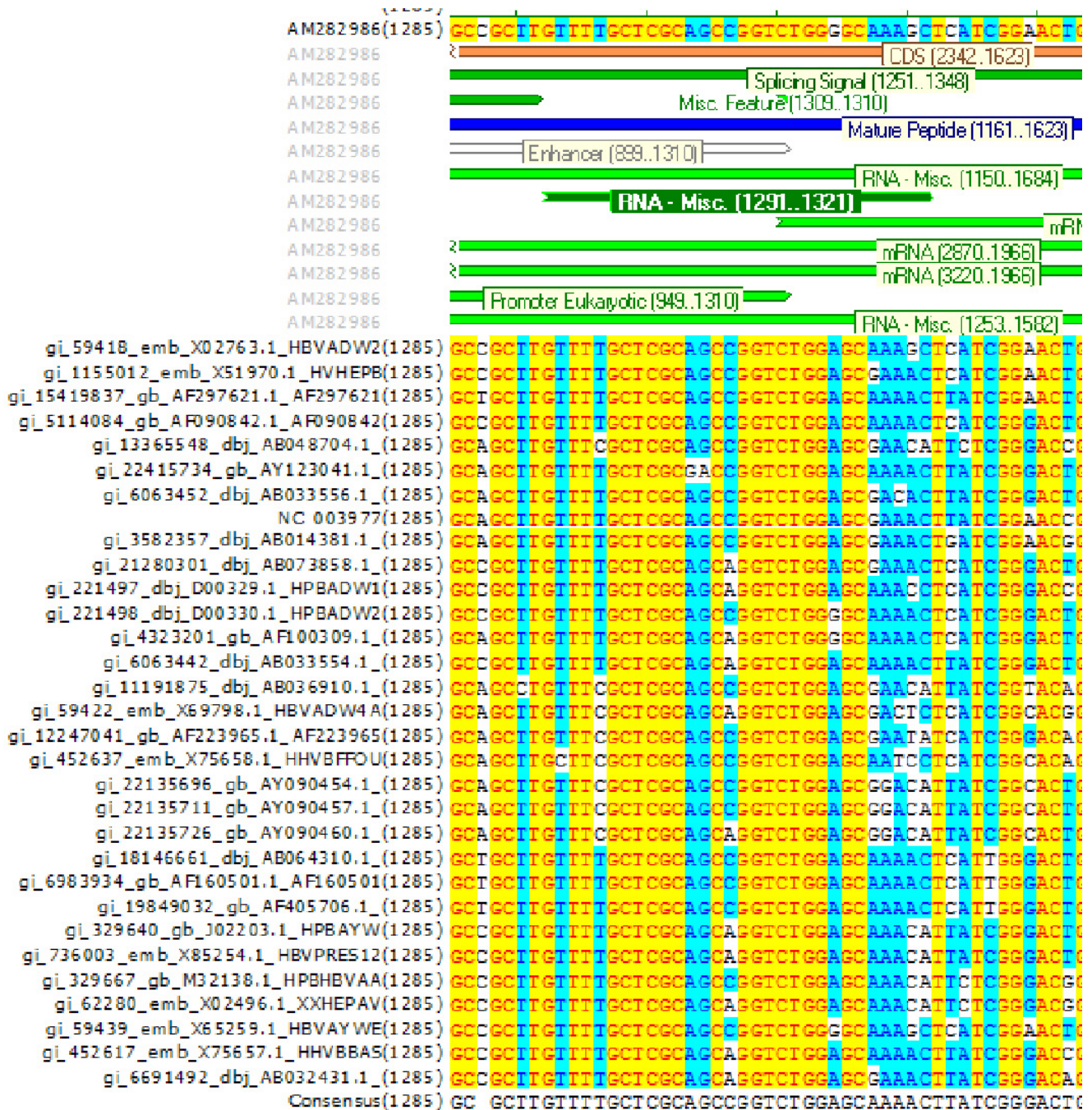


Figure 3
Multiple sequence alignment of HBV_HBVRegDB_32. The figure shows the region of the conserved RNA secondary structure known as HBV SL α (nucleotide 1292–1321, [38]). The annotation of genotype A (AM282986) is shown.

OHV_HBVRegDB_12

Orthohepadnavirus (OHV). This is an alignment of primate and rodent HBV genomes, NC_003977m group.

AHV_HBVRegDB_5

Avihepadnavirus. This is alignment of avian HBV genomes, NC_001344m group

Web-based graphical representation using GBrowse

A set of 65 representative HBV sequences from these groups of alignments is available using Gbrowse. For HBVRegDB the GBrowse software package was chosen because of its flexible configuration and efficient handling of large amounts of data, although a limitation here is the lack of ability to represent circular genomes. Annotations of conserved elements consist of large amounts of data, e.g. more than 30,000 records for one viral genome. GBrowse uses a Bio::DB::GFF schema in a MySQL data-

base and a fetch request is answered by the database query engine in a satisfactory time of ~20 seconds.

Underlying MySQL database

A version of MySQL was installed and configured for GBrowse. A Bio::DB::GFF schema was created and integrated into HBVRegDB to store the virus genome sequence data, annotations, statistical track data, and textual information. The five core tables of the HBVRegDB MySQL database contain additional taxonomic information, virus group relationships, web page search related data, and a comprehensive link reference list. An overview of the entire application information flow is shown schematically in Figure 4.

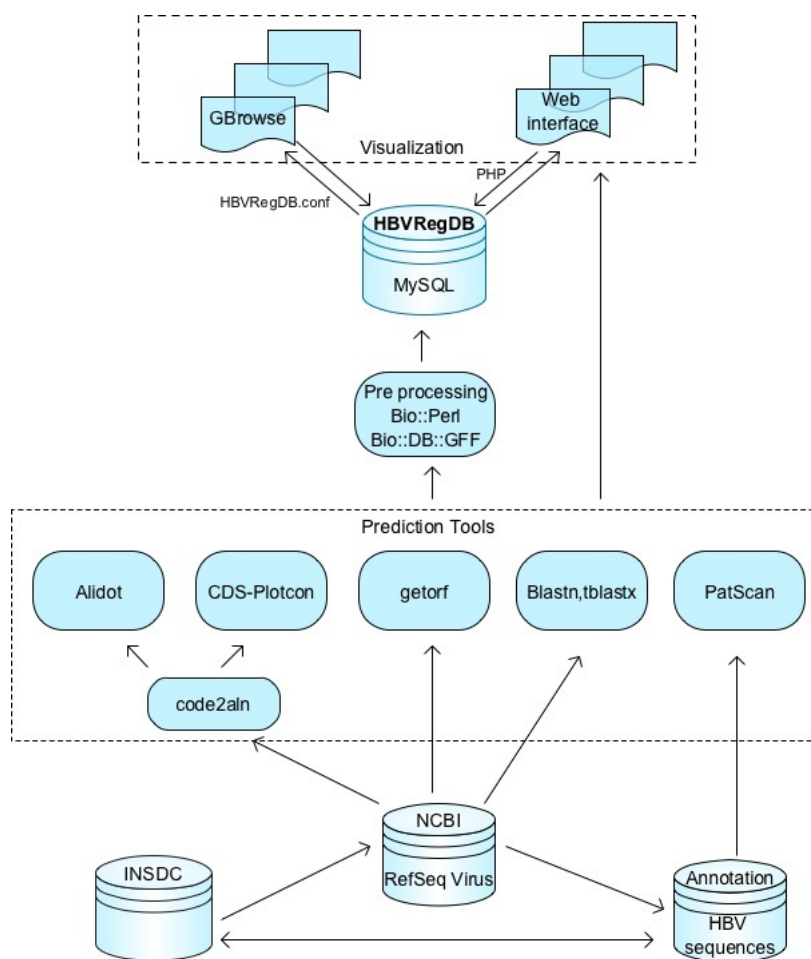


Figure 4

Schematic overview of the information flow in HBVRegDB. Boxes denote data sources. Cylinders represent database components.

Statistical and similarity search annotations on single sequences

Potential protein coding regions

Where annotated, CDSs are shown. For consistency, and for HBV genomes for which not all CDS sequences are annotated, potential ORFs >100 aa were calculated with getorf (EMBOSS). Coding regions, which extended over the virtual end of the viral genome sequence were automatically assigned and represented as two parts (e.g. Figure 5). This process also shows ORFs that could potentially be initiated at different ATG codons. For example (Figure 5) the predicted S ORFs, ORF 2 and 3, for which there is experimental evidence, or the predicted nested ORF 1, which could arise by internal initiation within P.

Similarity searches against other viral genomes

Blastn and tblastx were used to detect distant sequence similarities using selected parameters, shown in Table 1. The blastn parameters chosen will detect short exact matches. Tblastx was used to search a six-frame translated sequence against the protein database with hits of greater than two similar amino-acids analyzed. This could identify novel coding regions in query sequences, along with the CDS-plotcon analyses or alternative approaches [14]. Matches are shown in Figure 6. Blastn mainly finds other *Hepadnaviridae*, whereas tblastx (with these parameters) is able to detect more dissimilar matches, e.g. matches between reverse transcriptases from HBV and retroviruses (boxed).

Specific regulatory elements

As an example, PatScan [45] as implemented in Transtern [46], was used to identify polyadenylation sites by search-

ing the corresponding pattern AUWAAA. The output files were parsed and gff-formatted files were created and uploaded into the database.

User-added custom tracks

The user can add tracks in gff format. The search procedure above can be followed using online tools to annotate any motif that can be described by a regular expression, RNA descriptor or matrix. A description of this procedure is provided on the website.

Conserved primary and secondary structural elements

A way to detect functional elements in genomes is to look for conserved columns in multiple sequence alignments. However, many columns of CDSs show conservation due to constraints on the encoded protein. CDS-Plotcon is specifically designed to look for conserved functional elements within CDSs, independent of the protein coding constraints. To find conserved RNA structures, the program Alidot [47] was used. For each multiple sequence alignment, the Alidot and CDS-Plotcon results were gff-formatted and uploaded into the database. An example is shown in Figure 7. CDS-plotcon predicts unusually high conservation (higher than that required by the coding capacity in, for example, the boxed region). Similarly the predicted RNA secondary structure (Alidot) has higher than expected conservation in this region. This predicted region is the epsilon element, which is highly conserved in structure and function and is required for viral replication.

Web interface

All virus genomes in the database can be queried by browsing a table in which information about grouping,

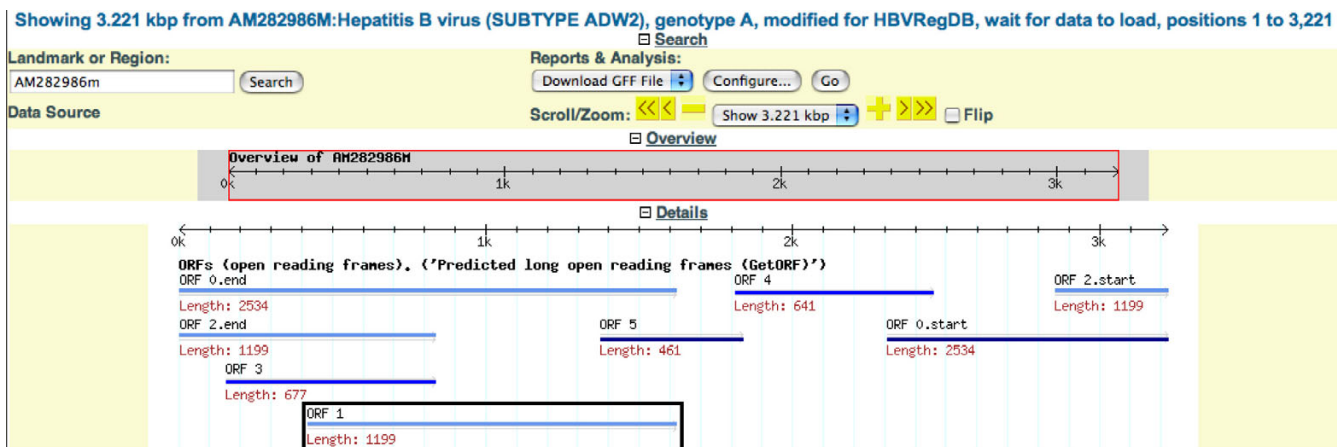


Figure 5
The top part of the screenshot showing annotations of the AM282986m Hepatitis B virus genome. Calculated ORFs are represented as bars. This analysis indicates ORFs that could potentially be initiated at different ATG codons. For example, the predicted nested ORF 1 (marked by box).

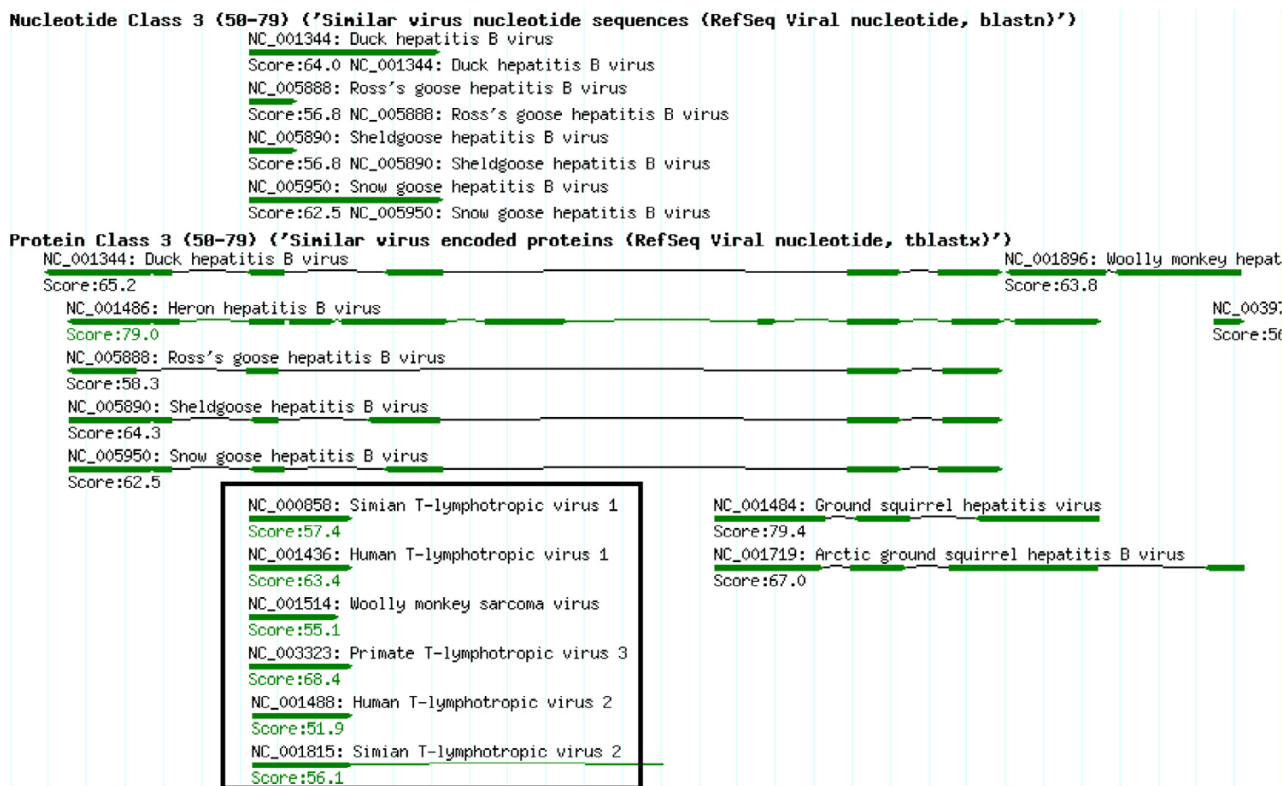


Figure 6
HBVRegDB-formatted results of the blastn and tblastx query (AM282986) against all viral sequences from RefSeq. Blast results from HBVRegDB are grouped in different classes based on match scores. This figure displays the results from Class 3 (scores between 50–79). Notably, parameters used in HBVRegDB are adjusted to allow matching of short sequences (-W 2 -G 8 -E 2). For example, the tblastx of HBVRegDB returns the hit of the short motif (YMDD) of the HBV P protein to the YMDD of the P protein from Human T-lymphotropic virus, Simian T-lymphotropic virus I and Woolly monkey sarcoma virus (boxed).

genotype, virus name and links to NCBI and to the graphical visualization of the sequence including annotations are provided. There is a comprehensive list of links to related web sites, which is intended to complement research using HBVRegDB. Tutorials in the form of web pages guide users through common analyses, such as:

- Comparison of your sequence to a well-annotated HBV genome.
- Testing for conservation of a sequence across genomes.
- Testing for conservation of an RNA secondary structure across genomes.
- Repeating similarity searches against HBVRegDB Sequences, RefSeq viral genomes and proteins.

Conclusion and future studies

Focused public domain viral databases have been developed, particularly for HIV, HCV and influenza, but for most viruses this is not available. Part of the approach described here can be generalized to any viral genomes. A preliminary analysis of all ~4000 viral segments in RefSeq has been done, building on the HBVRegDB database, and a comparative viral database (CompVirusDB) is being developed

Authors' contributions

NP carried out the annotations on single viral genomes, multiple sequence analysis, design of basis for HBVRegDB (e.g. content and structure) and drafted the manuscript. SKR developed web interface and a comparative genomic analysis tool with an integrated underlying HBV viral database. AEF developed a CDS-plotcon programme for detecting functional elements within coding regions. CMB substantially contributed to conception and design of the HBVRegDB, analysis of similarity searches against

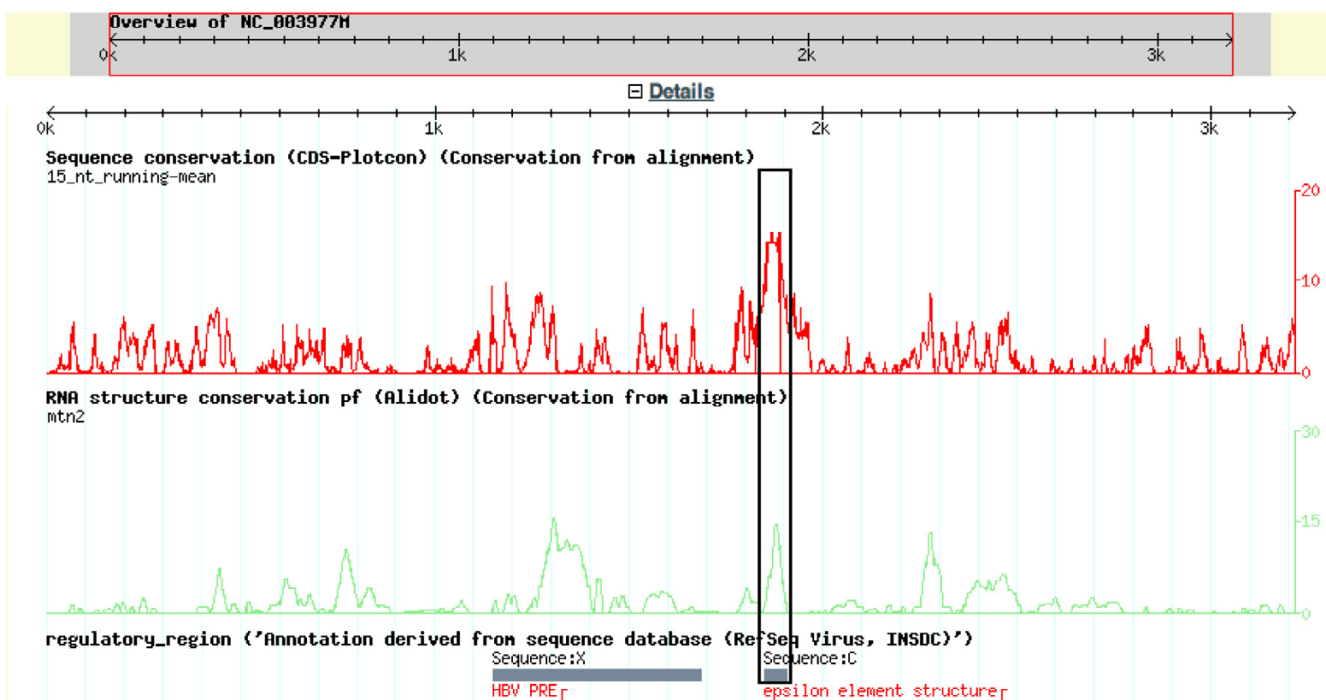


Figure 7

Conservation analyses (CDS-Plotcon and Alidot) of the virus group with the reference sequence NC_003977. For example, the known epsilon element is predicted to have conserved primary sequence (CDS-plotcon) and conserved secondary structure (Alidot), as experimentally demonstrated.

other viral genomes and preparation of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

NP by a scholarship from the Royal Thai Government, SKR was funded by the University of Otago Virology Theme, AEF by a Post-Doctoral fellowship from ForST (NZ), CMB by a grant from the NZ HRC.

References

- Beck J, Nassal M: **Hepatitis B virus replication.** *World J Gastroenterol* 2007, **13(1)**:48-64.
- Chan HL, Sung JJ: **Hepatocellular carcinoma and hepatitis B virus.** *Semin Liver Dis* 2006, **26(2)**:153-161.
- Jeong JK, Yoon GS, Ryu WS: **Evidence that the 5'-end cap structure is essential for encapsidation of hepatitis B virus pregenomic RNA.** *J Virol* 2000, **74(12)**:5502-5508.
- Moolla N, Kew M, Arbutnot P: **Regulatory elements of hepatitis B virus transcription.** *J Viral Hepat* 2002, **9(5)**:323-331.
- Osiowy C, Giles E, Tanaka Y, Mizokami M, Minuk GY: **Molecular evolution of hepatitis B virus over 25 years.** *J Virol* 2006, **80(21)**:10307-10314.
- Tong S: **Mechanism of HBV genome variability and replication of HBV mutants.** *J Clin Virol* 2005, **34 Suppl 1**:S134-8.
- Glebe D, Urban S: **Viral and cellular determinants involved in hepadnaviral entry.** *World J Gastroenterol* 2007, **13(1)**:22-38.
- Chen WN, Oon CJ: **Hepatitis B virus mutants: an overview.** *J Gastroenterol Hepatol* 2002, **17 Suppl**:S497-9.
- Bao Y, Federhen S, Leipe D, Pham V, Resenchuk S, Rozanov M, Tatusov R, Tatusova T: **National center for biotechnology information viral genomes project.** *J Virol* 2004, **78(14)**:7291-7298.
- Pruitt KD, Tatusova T, Maglott DR: **NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.** *Nucleic Acids Res* 2007, **35(Database issue)**:D61-5.
- Hirahata M, Abe T, Tanaka N, Kuwana Y, Shigemoto Y, Miyazaki S, Suzuki Y, Sugawara H: **Genome Information Broker for Viruses (GIB-V): database for comparative analysis of virus genomes.** *Nucleic Acids Res* 2007, **35(Database issue)**:D339-42.
- Yuen LK, Ayres A, Littlejohn M, Colledge D, Edgely A, Maskill WJ, Locarnini SA, Bartholomeusz A: **SEQHEPB: A sequence analysis program and relational database system for chronic hepatitis B.** *Antiviral Res* 2006.
- Gnaneshan S, Ijaz S, Moran J, Ramsay M, Green J: **HepSEQ: International Public Health Repository for Hepatitis B.** *Nucleic Acids Res* 2007, **35(Database issue)**:D367-70.
- Firth AE, Brown CM: **Detecting overlapping coding sequences with pairwise alignments.** *Bioinformatics* 2005, **21(3)**:282-292.
- Firth AE, Brown CM: **Detecting overlapping coding sequences in virus genomes.** *BMC Bioinformatics* 2006, **7**:75.
- McGinnis S, Madden TL: **BLAST: at the core of a powerful and diverse set of sequence analysis tools.** *Nucleic Acids Res* 2004, **32(Web Server issue)**:W20-5.
- Stocsits RR, Hofacker IL, Fried C, Stadler PF: **Multiple sequence alignments of partially coding nucleic acid sequences.** *BMC Bioinformatics* 2005, **6**:160.
- 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(22)**:4673-4680.
- Chu CJ, Lok AS: **Clinical significance of hepatitis B virus genotypes.** *Hepatology* 2002, **35(5)**:1274-1276.
- Du H, Li T, Zhang HY, He ZP, Dong QM, Duan XZ, Zhuang H: **Correlation of hepatitis B virus (HBV) genotypes and mutations in basal core promoter/precore with clinical features of chronic HBV infection.** *Liver Int* 2007, **27(2)**:240-246.
- Huang ZM, Huang QW, Qin YQ, Huang CH, Qin HJ, Zhou YN, Xu X, Lu CL: **Clinical characteristics and distribution of hepatitis B virus genotypes in Guangxi Zhuang population.** *World J Gastroenterol* 2005, **11(41)**:6525-6529.

22. Kidd-Ljunggren K, Myhre E, Blackberg J: **Clinical and serological variation between patients infected with different Hepatitis B virus genotypes.** *J Clin Microbiol* 2004, **42(12)**:5837-5841.
23. Locarnini SA: **Clinical relevance of viral dynamics and genotypes in hepatitis B virus.** *J Gastroenterol Hepatol* 2002, **17 Suppl 3**:S322-S328.
24. Ogawa M, Hasegawa K, Naritomi T, Torii N, Hayashi N: **Clinical features and viral sequences of various genotypes of hepatitis B virus compared among patients with acute hepatitis B.** *Hepatology Res* 2002, **23(3)**:167-177.
25. Wai CT, Fontana RJ: **Clinical significance of hepatitis B virus genotypes, variants, and mutants.** *Clin Liver Dis* 2004, **8(2)**:321-52, vi.
26. Schaefer S: **Hepatitis B virus taxonomy and hepatitis B virus genotypes.** *World J Gastroenterol* 2007, **13(1)**:14-21.
27. Chiu CM, Yeh SH, Chen PJ, Kuo TJ, Chang CJ, Chen PJ, Yang WJ, Chen DS: **Hepatitis B virus X protein enhances androgen receptor-responsive gene expression depending on androgen level.** *Proc Natl Acad Sci U S A* 2007, **104(8)**:2571-2578.
28. Prassolov A, Hohenberg H, Kalinina T, Schneider C, Cova L, Krone O, Frolich K, Will H, Sirma H: **New hepatitis B virus of cranes that has an unexpected broad host range.** *J Virol* 2003, **77(3)**:1964-1976.
29. Loeb DD, Mack AA, Tian R: **A secondary structure that contains the 5' and 3' splice sites suppresses splicing of duck hepatitis B virus pregenomic RNA.** *J Virol* 2002, **76(20)**:10195-10202.
30. Suzuki T, Kajino K, Masui N, Saito I, Miyamura T: **Alternative splicing of hepatitis B virus RNAs in HepG2 cells transfected with the viral DNA.** *Virology* 1990, **179(2)**:881-885.
31. Chen A, Kao YF, Brown CM: **Translation of the first upstream ORF in the hepatitis B virus pregenomic RNA modulates translation at the core and polymerase initiation codons.** *Nucleic Acids Res* 2005, **33(4)**:1169-1181.
32. Stuyver LJ, Locarnini SA, Lok A, Richman DD, Carman WF, Dienstag JL, Schinazi RF: **Nomenclature for antiviral-resistant human hepatitis B virus mutations in the polymerase region.** *Hepatology* 2001, **33(3)**:751-757.
33. Paran N, Ori A, Haviv I, Shaul Y: **A composite polyadenylation signal with TATA box function.** *Mol Cell Biol* 2000, **20(3)**:834-841.
34. Alcantara FF, Tang H, McLachlan A: **Functional characterization of the interferon regulatory element in the enhancer I region of the hepatitis B virus genome.** *Nucleic Acids Res* 2002, **30(9)**:2068-2075.
35. Perfumo S, Amicone L, Colloca S, Giorgio M, Pozzi L, Tripodi M: **Recognition efficiency of the hepatitis B virus polyadenylation signals is tissue specific in transgenic mice.** *J Virol* 1992, **66(11)**:6819-6823.
36. Tong SP, Li JS, Vitvitski L, Kay A, Treepo C: **Evidence for a base-paired region of hepatitis B virus pregenome encapsidation signal which influences the patterns of precore mutations abolishing HBe protein expression.** *J Virol* 1993, **67(9)**:5651-5655.
37. Wu HL, Chen PJ, Lin MH, Chen DS: **Temporal aspects of major viral transcript expression in Hep G2 cells transfected with cloned hepatitis B virus DNA: with emphasis on the X transcript.** *Virology* 1991, **185(2)**:644-651.
38. Donello JE, Beeche AA, Smith GJ 3rd, Lucero GR, Hope TJ: **The hepatitis B virus posttranscriptional regulatory element is composed of two subelements.** *J Virol* 1996, **70(7)**:4345-4351.
39. Huang J, Liang TJ: **A novel hepatitis B virus (HBV) genetic element with Rev response element-like properties that is essential for expression of HBV gene products.** *Mol Cell Biol* 1993, **13(12)**:7476-7486.
40. Huang ZM, Yen TS: **Role of the hepatitis B virus posttranscriptional regulatory element in export of intronless transcripts.** *Mol Cell Biol* 1995, **15(7)**:3864-3869.
41. Heise T, Sommer G, Reumann K, Meyer I, Will H, Schaal H: **The hepatitis B virus PRE contains a splicing regulatory element.** *Nucleic Acids Res* 2006, **34(1)**:353-363.
42. Knaus T, Nassal M: **The encapsidation signal on the hepatitis B virus RNA pregenome forms a stem-loop structure that is critical for its function.** *Nucleic Acids Res* 1993, **21(17)**:3967-3975.
43. Zufferey R, Donello JE, Trono D, Hope TJ: **Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors.** *J Virol* 1999, **73(4)**:2886-2892.
44. Donello JE, Loeb JE, Hope TJ: **Woodchuck hepatitis virus contains a tripartite posttranscriptional regulatory element.** *J Virol* 1998, **72(6)**:5085-5092.
45. Dsouza M, Larsen N, Overbeek R: **Searching for patterns in genomic data.** *Trends Genet* 1997, **13(12)**:497-498.
46. Jacobs GH, Stockwell PA, Tate WP, Brown CM: **Transterm--extended search facilities and improved integration with other databases.** *Nucleic Acids Res* 2006, **34(Database issue)**:D37-40.
47. Hofacker IL, Fekete M, Stadler PF: **Secondary structure prediction for aligned RNA sequences.** *J Mol Biol* 2002, **319(5)**:1059-1066.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

