

# HERVd: the Human Endogenous RetroViruses Database: update

Jan Pačes<sup>1</sup>, Adam Pavlíček<sup>1,2</sup>, Radek Zika<sup>1</sup>, Vladimir V. Kapitonov<sup>2</sup>, Jerzy Jurka<sup>2</sup> and Václav Pačes<sup>1,\*</sup>

<sup>1</sup>Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Flemingovo 2, CZ-16637 Prague, Czech Republic and <sup>2</sup>Genetic Information Research Institute, 2081 Landings Drive, Mountain View, CA 94043, USA

Received September 15, 2003; Accepted September 30, 2003

## ABSTRACT

**An elaboration of HERVd (<http://herv.img.cas.cz>) is being carried out in two directions. One of them is the integration and better classification of families that diverge considerably from typical retroviral genomes. This leads to a more precise identification of members with individual families. The second improvement is better accessibility of the database and connection with human genome annotation.**

## DATABASE DESCRIPTION

The Human Endogenous RetroViruses Database (HERVd) is designed to identify, store, classify and make accessible retrovirus-like elements that are present in the human genome (1). The source for the database was the output of the human genome project in NCBI (<http://www.ncbi.nlm.nih.gov>) (2) and GoldenPath (<http://genome.ucsc.edu>) (3). Copies of known endogenous retroviruses collected in Repbase Update (4) were detected by RepeatMasker (A. F. Smit and P. Green, unpublished) and processed by the defragmentation algorithm developed by us earlier (1). The database can be searched using several criteria such as HERV families, chromosomal locations or DNA similarities. The sequences, short descriptions and graphic outputs of all entries are available.

## RECENT DEVELOPMENTS

Our effort of the past year has been concentrated in four areas: (i) including nucleotide sequences that diverged from colinearity with the typical retroviral genome [LTR-gag-pol(pro)-env-LTR] and thus considerably increasing the number of HERV families and quantity of data; (ii) better classification of HERV families and thereby increasing the quality of data; (iii) adding of both DNA and protein similarity search and (iv) creating links to other databases thus improving the accessibility of the HERVd and integration with the human genome annotation.

## Data expansion and classification

Classification of HERVs is based on consensus sequences in Repbase Update (deposited by V. V. Kapitonov, A. F. Smit

and J. Jurka) and published literature as appeared in the original HERVd (1). New consensus sequences improved detection of HERVs in the genome. This was especially important for non-autonomous elements that diverged considerably from the typical retroviral genome. The number of HERV families in the database more than doubled compared with the original version (1). The total number of different families in the database is now 150.

## Data accessibility

Another improvement of the database is that HERVs can now be searched by nucleotide sequences for DNA and protein similarity using BLAT (5). In addition, we integrated our database with the human genome annotation. For each element a link to the UCSC Genome Browser (6) is now available.

## ACKNOWLEDGEMENTS

This work was supported by the Center for Integrated Genomics and with grant M023 from the Grant Agency of the Czech Republic.

## REFERENCES

1. Pačes, J., Pavlíček, A. and Pačes, V. (2002) HERVd: database of human endogenous retroviruses. *Nucleic Acids Res.*, **30**, 205–206.
2. Wheeler, D.L., Church, D.M., Federhen, S., Lash, A.E., Madden, T.L., Pontius, J.U., Schuler, G.D., Schriml, L.M., Sequeira, E., Tatusova, T.A. *et al.* (2003) Database resources of the National Center for Biotechnology. *Nucleic Acids Res.*, **31**, 28–33.
3. Karolchik, D., Baertsch, R., Diekhans, M., Furey, T.S., Hinrichs, A., Lu, Y.T., Roskin, K.M., Schwartz, M., Sugnet, C.W., Thomas, D.J. *et al.* (2003) The UCSC Genome Browser Database. *Nucleic Acids Res.*, **31**, 51–54.
4. Jurka, J. (2000) Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet.*, **16**, 418–420.
5. Kent, W.J. (2002) BLAT—the BLAST-like Alignment Tool. *Genome Res.*, **12**, 656–664.
6. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.

\*To whom correspondence should be addressed. Tel: +420 2 20183541; Fax: +420 2 24311019; Email: vpace@img.cas.cz