Abstract

Retrocopies of protein-coding genes, reverse transcribed and inserted into the genome copies of mature RNA, have been categorized as pseudogenes with no biological importance. However, recent studies showed that they play an important role in the genomes evolution and shaping interspecies differences. Here, we present RetrogeneDB, a database of retrocopies in 62 animal genomes. RetrogeneDB contains information about retrocopies, their genomic localization, parental genes, ORF conservation, and expression. To our best knowledge, this is the most complete retrocopies database providing information for dozens of species previously never analyzed in the context of protein-coding genes retroposition. The database is available at http://retrogeneadb.amu.edu.pl.

Key words: retroposition, gene duplication, retrogene, database.
mammals account for 10–25% of retrocopies. In nonmammalian animals, the fraction is much higher, considerably over 50% and in some species close to 100. However, the conservation of the ORF over the length of alignment does not automatically imply that a retrocopy is efficiently translated, even if it is expressed. In selected species, we also identified expressed retrocopies based on the RNA-seq data. Because of the high similarity to parental genes, in the process of reads mapping, we made sure they uniquely and perfectly map to retrocopies (supplementary file S1, Supplementary Material online). This led to the underestimation of retrocopies expression level but prevented false-positive predictions of expressed retrocopies. Approximately 10–20% of mammalian retrocopies are expressed in at least one library at minimal level of 1 RPM (reads per million mapped). In lizard, this number is higher with almost 40% of expressed retrocopies. Majority of expressed retrocopies in marsupials, egg-laying mammals, and nonmammalian species have conserved ORFs. However, in placental mammals, the fraction of expressed retrocopies with conserved ORF is lower, from only 30% in human up to 65% in horse.

All the data are stored in MySQL database (www.mysql.com, last accessed September 2013), and the web interface was developed using Django framework (www.djangoproject.com, last accessed January 2014). The database is available at http://retrogenedb.amu.edu.pl (last accessed April 26, 2014) and can be searched either from the retrocopy or the parental gene perspective. The retrocopy search can be done based on

**Fig. 1.** Example of RetrogeneDB record with selected data.
the genomic localization, key words, parental gene name, and retrocopy ID, and results can be filtered based on the retrocopy type, ORF conservation, or expression. In addition, a JBrowse genome browser was implemented allowing retrocopy inspection in the genomic context (fig. 1). The search from parental gene perspective enables to identify all retrocopies of a given gene or all orthologs, which were retroposed in any other species. Users can also perform sequence-based search using BLAST tool.

**Supplementary Material**

Supplementary file S1 is available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

**Acknowledgments**

This work was supported by the National Science Centre of Poland grant number 2013/09/N/NZ2/01221 to M.K. and grant number 2011/01/N/NZ2/01701 to J.C. and European Union grant PIRSES-GA-2009-247633 to IM.

**References**

Abyzov A, Iskow R, Gokcumen O, Radke DW, Balasubramanian S, Pei B, Habegger L. 2013. Analysis of variable retroduplications in human populations suggests coupling of retrotransposition to cell division. *Genome Res.* 23(12):2042–2052.

Balasubramanian S, Zheng D, Liu Y-J, Fang G, Frankish A, Carriero N, Robilotto R, Cayting P, Gerstein M. 2009. Comparative analysis of processed ribosomal protein pseudogenes in four mammalian genomes. *Genome Biol.* 10(R2).

Betrán E, Wang W, Jin L, Long M. 2002. Evolution of the phosphoglycerate mutase processed gene in human and chimpanzee revealing the origin of a new primate gene. *Mol Biol Evol.* 19:654–663.

Brosius J. 1991. Retroposons—seeds of evolution. *Science* 251:753.

Ciomborowska J, Rosikiewicz W, Skłarczyk D, Makalowski W, Makalowska I. 2013. "Orphan" retrogenes in the human genome. *Mol Biol Evol.* 30:384–396.

Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, et al. 2014. Ensembl 2014. *Nucleic Acids Res.* 42:D749–D755.

Kaessmann H, Vinckenbosch N, Long M. 2009. RNA-based gene duplication: mechanistic and evolutionary insights. *Nat Rev Genet.* 10:19–31.

Kheifiz A, Duret L, Mouchiroud D. 2005. HOPPSiGEN: a database of human and mouse processed pseudogenes. *Nucleic Acids Res.* 33:1–15.

Meyer LR, Zweig AS, Hinrichs AS, Karolchik D, Kuhn RM, Wong M, Sloan CA, Rosenbloom KR, Roe G, Rhead B, et al. 2013. The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res.* 41:D64–D69.

Navarro FC, Galante PA. 2013. RCPedia: a database of retrocopied genes. *Bioinformatics* 29:1235–1237.

Prendergast GC. 2001. Actin’ up: RhoB in cancer and apoptosis. *Nat Rev Cancer.* 1:162–168.

Szczęśniak MW, Ciomborowska J, Nowak W, Rogozin IB, Makalowska I. 2011. Primate and rodent specific intron gains and the origin of retrogenes with splice variants. *Mol Biol Evol.* 28:333–337.

Vinckenbosch N, Dupanloup I, Kaessmann H. 2006. Evolutionary fate of retroposed gene copies in the human genome. *Proc Natl Acad Sci U S A.* 103:3220–3225.

Yu Z, Morais D, Ivanga M, Harrison PM. 2007. Analysis of the role of retrotransposition in gene evolution in vertebrates. *BMC Bioinformatics* 8:308.