

# YIDB: the Yeast Intron DataBase

Pascal J. Lopez and Bertrand Séraphin\*

EMBL, Meyerhofstraße 1, D-69117, Germany

Received August 31, 1999; Revised and Accepted October 4 1999

## ABSTRACT

**The Yeast Intron DataBase (YIDB) contains currently available information about all introns encoded in the nuclear and mitochondrial genomes of the yeast *Saccharomyces cerevisiae*. Introns are divided according to their mechanism of excision: group I and group II introns, pre-mRNA introns, tRNA introns and the HAC1 intron. Information about the host genome, the type of RNA in which they are inserted and their primary structure are provided together with references. For nuclear pre-mRNA introns, transcription frequencies, as determined by microarray experiments, have also been included. This updated database is accessible at: <http://www.embl-heidelberg.de/ExternalInfo/seraphin/yidb.html>**

## INTRODUCTION

Introns are sequences present in various types of gene that need to be removed from the primary transcript in order to form a functional RNA. Introns are present in all classes of RNAs (rRNA, tRNA, mRNA, etc.) and have been found in various genomes (eukaryotic, prokaryotic, organelles, viruses, etc.). Intron sequences have to be precisely recognized and eliminated from pre-RNA to allow for functional protein or RNA synthesis. In a few cases, introns are involved in the regulation of the expression of their host genes, are alternatively spliced, correspond to mobile genetic elements or code themselves for protein or functional RNA. They constitute therefore a remarkable evolutionary tool.

Introns are classified according to their excision mechanisms. In yeast, five different classes can be distinguished. (i) Group I introns are spliced via two guanosine-initiated transesterification reactions. Some of these introns are autocatalytic. Group I introns share limited sequence similarity but have conserved secondary and tertiary structures (1). (ii) Group II introns are spliced via two transesterification reactions similar to the ones occurring for nuclear pre-mRNA introns. Some group II introns are also autocatalytic. They have conserved 5' and 3' extremities as well as secondary and tertiary structures (2). *In vivo* splicing of group I and II introns often depends on proteins (3). (iii) Splicing of nuclear spliceosomal introns also involves two transesterification reactions. This process is catalyzed by a large ribonucleoprotein machinery named the spliceosome. This complex and dynamic enzyme is made by the ordered assembly onto the intron of five snRNPs (small nuclear ribonucleoproteins) as well as non-snRNP proteins; in total more

than 100 proteins may contribute to the spliceosome (4). Conserved sequences are present at the 5' and 3' ends of spliceosomal introns, at their branchpoint and, to a lesser extent, in flanking exon sequences. (iv) Introns are also found in nuclear tRNA genes. They show little sequence conservation but are always inserted at the same location in tRNAs. The removal of these introns is catalyzed by proteins (5). (v) Finally, a special type of intron is found in the nuclear HAC1 gene (6). Splicing of this intron is catalyzed by protein factors, some of which are shared with tRNA introns. The splicing of this intron is highly regulated.

A compilation of intron sequences can help not only to analyze the chromosomal organization of yeast genomes, but also to define consensus sequences and nucleotide contents, features that are crucial for systematic gene identification. Moreover, the database can help by taking into account quantitative data about expression such as transcription levels, to gain a better insight into cellular processes (7).

## DATABASE CONTENT AND ORGANIZATION

The Yeast Intron DataBase (YIDB) contains information for the five different types of introns found in the nuclear and mitochondrial genomes of the budding yeast *Saccharomyces cerevisiae*. The data are derived from primary databases (SGD, YPD, MIPS and EMBL) completed with information retrieved from the literature. This information was manually edited to add or remove entries that contained annotation errors and to take into account new experimental evidence.

A summary table shows the overall representation of introns within the *S.cerevisiae* genomes. Information for each type of intron is presented in independent sections as follows.

- The location, name as well as specific information about intron mobility and polymorphic presence for the 13 group I and II mitochondrial introns.
- A list of 254 spliceosomal introns and putative introns present in 249 genes. Partial sequences of the exon 1, 5' and 3' splice sites, and of the branchpoint are provided as well as the intron size. When available, we also provide the transcription frequency of the corresponding genes (i.e., number of transcripts produced per unit of time) determined by microarray experiments (8). A related database containing complementary information about spliceosomal intron can be found at [http://www.cse.ucsc.edu/research/compbio/yeast\\_introns.html](http://www.cse.ucsc.edu/research/compbio/yeast_introns.html) (9).
- A list of the 61 nuclear tRNA genes that contain an intron alongside with their location, cognate amino-acid identity and intron length.

\*To whom correspondence should be addressed. Tel. +49 6221 387486; Fax: +49 6221 387518; Email: seraphin@embl-heidelberg.de

Links to related entries from the EMBL, MIPS or SGD databases as well as to some relevant references from the PubMed database are provided.

#### **DATABASE AVAILABILITY AND CITATION**

Access to the YIDB for academic usage is possible through the World Wide Web at

<http://www.embl-heidelberg.de/ExternalInfo/seraphin/yidb.html>

Users of the database should cite the present publication as reference. Comments, corrections and new entries are welcome.

#### **REFERENCES**

1. Michel,F. and Westhof,E. (1990) *J. Mol. Biol.*, **216**, 585–610.
2. Michel,F. and Ferat,J.L. (1995) *Annu. Rev. Biochem.*, **64**, 435–461.
3. Lambowitz,A.M. and Belfort,M. (1993) *Annu. Rev. Biochem.*, **62**, 587–622.
4. Burge,C.B., Tuschl,T. and Sharp,P.A. (1999) In Gesteland,R.F., Cech,T.R. and Atkins,J.F. (eds), *The RNA World*. 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
5. Abelson,J., Trotta,C.R. and Li,H. (1998) *J. Biol. Chem.*, **273**, 12685–12688.
6. Gonzalez,T.N., Sidrauski,C., Dorfler,S. and Walter,P. (1999) *EMBO J.*, **18**, 3119–3132.
7. Lopez,P.J. and Séraphin,B. (1999) *RNA*, **5**, 1135–1137.
8. Holstege,F.C., Jennings,E.G., Wyrick,J.J., Lee,T.I., Hengartner,C.J., Green,M.R., Golub,T.R., Lander,E.S. and Young,R.A. (1998) *Cell*, **95**, 717–728.
9. Spingola,M., Grate,L., Haussler,D. and Ares,M.,Jr (1999) *RNA*, **5**, 221–234.