snoStrip: a Fungi snoRNAome Recipe for Success.

Sebastian Bartschat, Stephanie Kehr, Hakim Tafer, Peter F. Stadler, Jana Hertel

Bioinformatics Group,
Dept. of Computer Science,
University of Leipzig, Germany

contact: sebastian@bioinf.uni-leipzig.de
web server: www.snostrip.bioinf.uni-leipzig.de

Motivation

Small nucleolar RNAs (snoRNAs) are one of the most abundant and evolutionary ancient functional non-coding RNAs. They fulfill an impressive variety of cellular functions ranging from guiding chemical modifications in several ncRNA classes and nucleolytic processing of rRNAs to an involvement in genomic imprinting and alternative splicing.

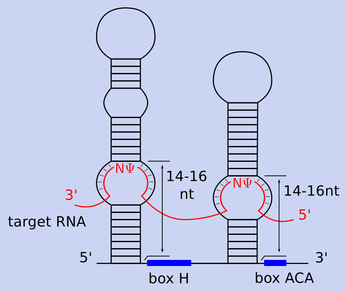

Figure 1: Box H/ACA snoRNA

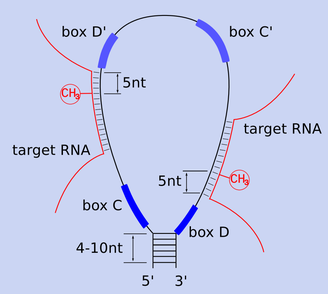

Figure 2: Box C/D snoRNA

They fall mainly into two distinct classes depending on their canonical sequence motifs, box C/D and box H/ACA snoRNAs. However, snoRNAs are **often hard to detect** by means of homology search, due to several reasons:

- lack of overall sequence conservation
- small sequence size
- solely short sequence motifs
- lack of stable structure (box C/D snoRNA)

snoStrip pipeline

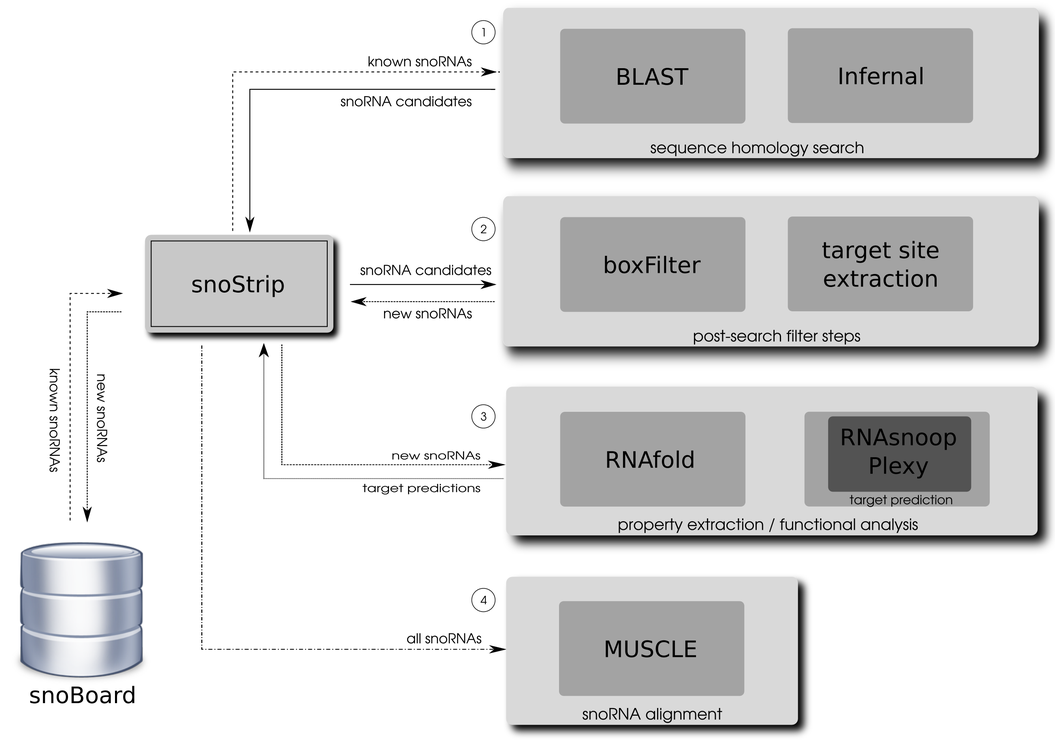

Figure 3: Automatic annotation of snoRNAs with the snoStrip pipeline.

Results

Evolution of snoRNAs in 63 fungal species:

start set:
- 5 different organisms (marked with *)
- 231 experimentally verified snoRNAs

outcome:
- **123 snoRNA families**
  including: 67 box C/D families
  56 box H/ACA families

- **3564 snoRNA sequences**
  including: 2565 box C/D snoRNAs
  999 box H/ACA snoRNAs

- **129 analyzed targets**
  including: 71 box C/D targets
  58 box H/ACA targets

- still 15 orphan snoRNA families

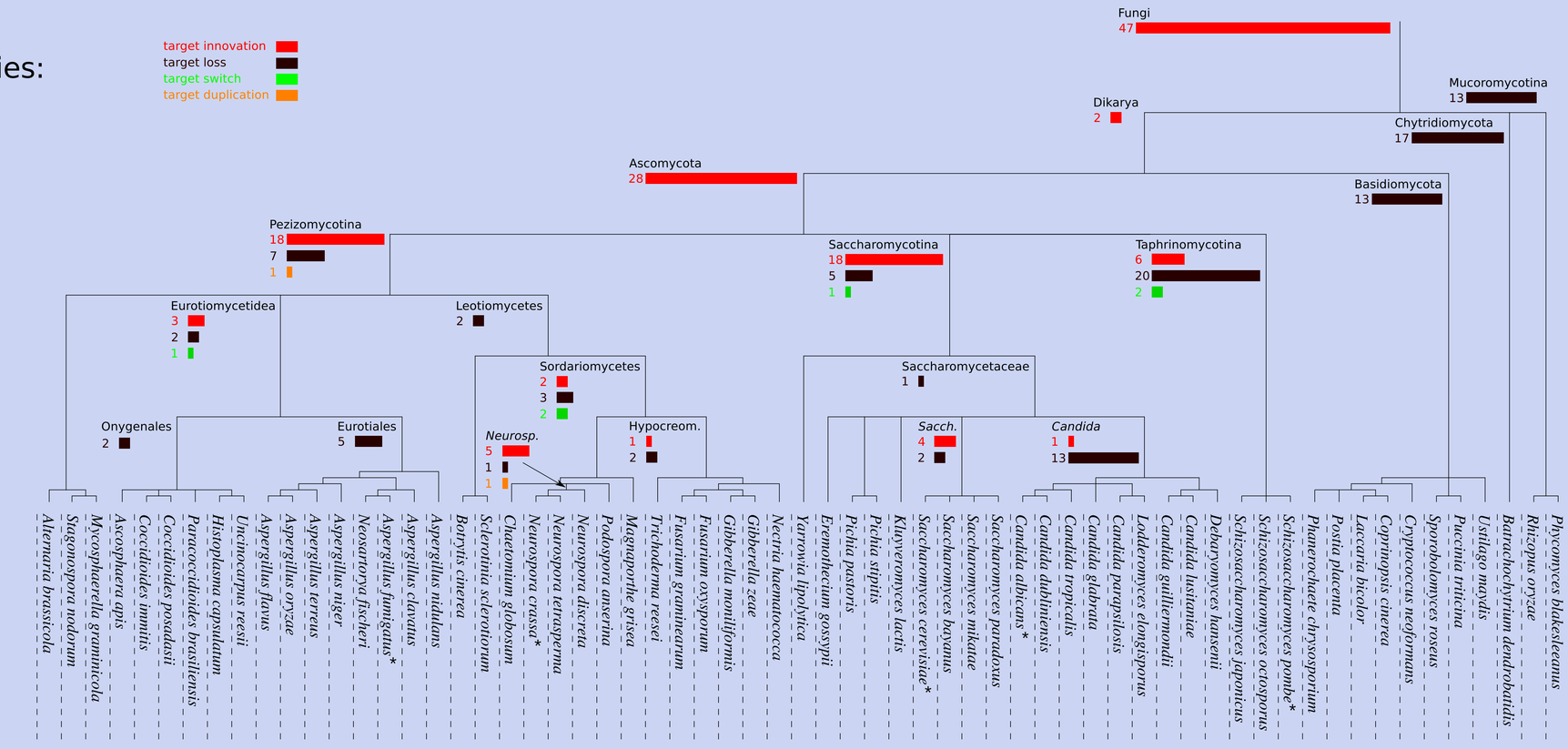

Figure 4: Target innovation, deletion, duplication, and switches during the evolution of snoRNA in fungal species

snoRNA clans

Several snoRNA families underwent major rearrangements during their evolution, including target switches, loss of target sites, and the (re)innovation of target sites.

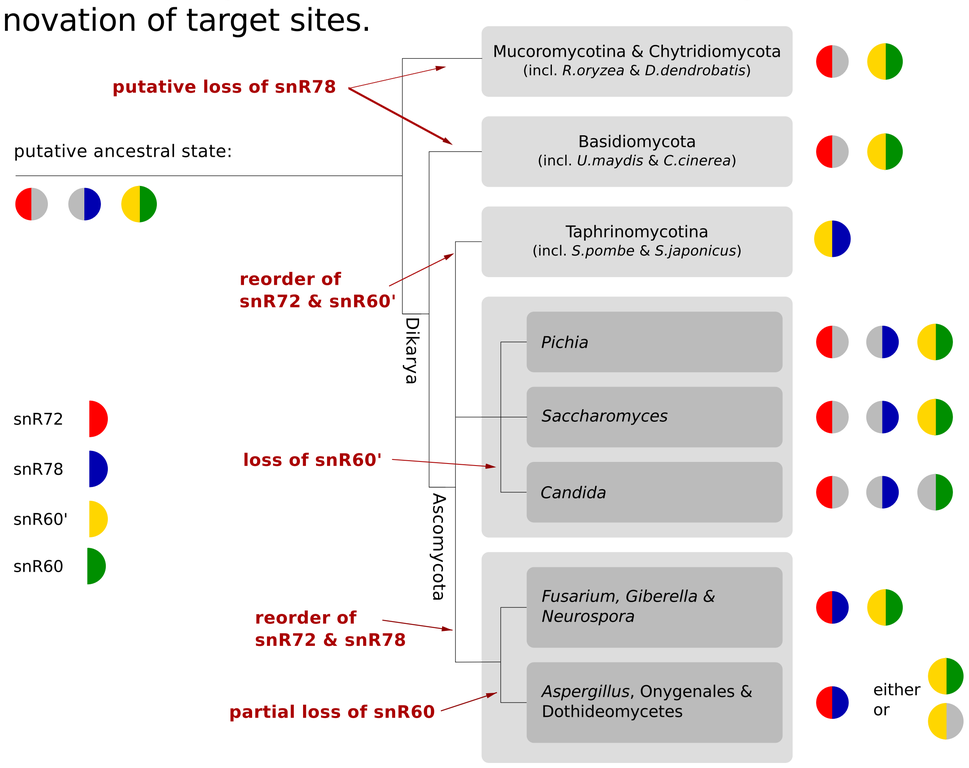

Figure 5: snoRNA clan containing snoRNA families snR60, snR72, and snR78

snoStrip web server

Two different operating modes:

1) genome-wide snoRNA annotation

input: (novel) fungal genome
output: - homologes of known snoRNAs
- putative targets
- alignments w.r.t. sequence conservation
- alignments w.r.t. target interaction conservation
benefit: - whole snoRNAome annotation w.r.t. known snoRNAs
- functional analysis of identified sequences

2) single sequence conservation

input: snoRNA candidates + box motif annotation
output: - homologs in fungi
- reconciliation of known snoRNA families
benefit: - verify candidates w.r.t. to their conservation
- preventing mis- or double annotations

Acknowledgement

This work was funded in part by the European Union under the auspices of the FP-7 project QUANTOMICS.

UNIVERSITÄT LEIPZIG

Quantomics
From Sequence to Consequence :
Tools for the Exploitation of Livestock Genomes