Searching for non-coding RNAs and predicting their structure

Young Investigators Day 2011

Christian Höner zu Siederdissen [1]   Ivo L. Hofacker [1]

[1] Institute for Theoretical Chemistry
University of Vienna
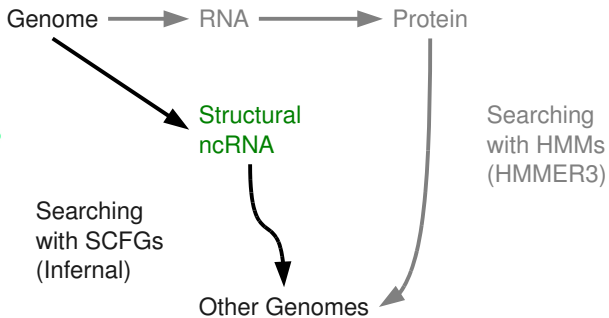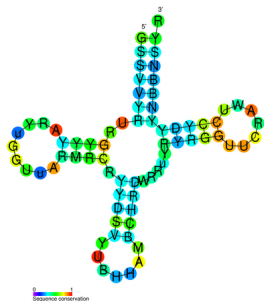
February 3, 2011

universität
wien

GEN-AU
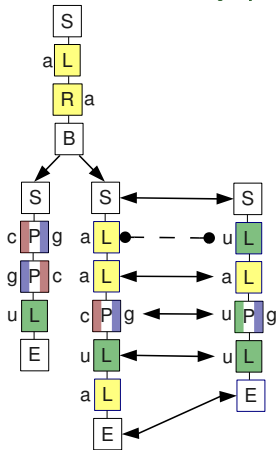GENOME RESEARCH IN AUSTRIA

non-coding RNA search

an improved Infernal grammar

- three meaningful semantics: *alignment*, trace, and structural
- prove alignment grammar unambiguous
- trace (and structural) semantics more likely to capture remote family members
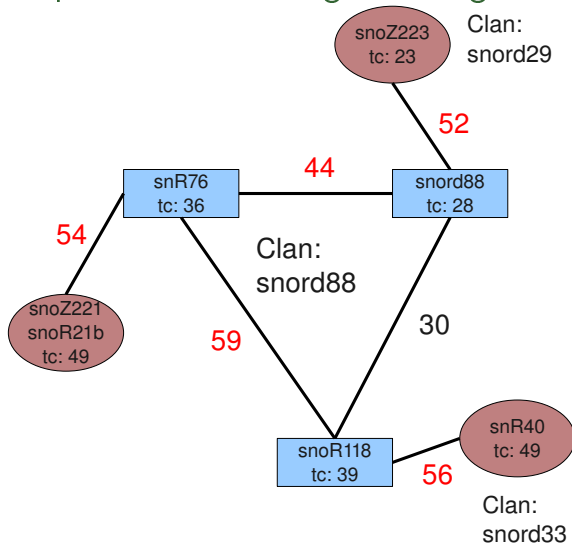- higher sensitivity!

```
alignment 1      alignment 2      trace
ACAGGGG---CAG    ACA---GGGGCAG    ACA[GGGG]CAG
ACA----TTTCAC    ACATTT----CAG    ACA[TTT] CAG
```
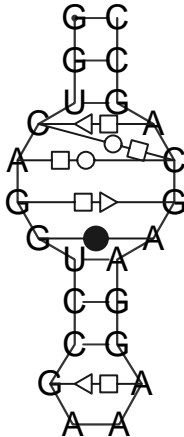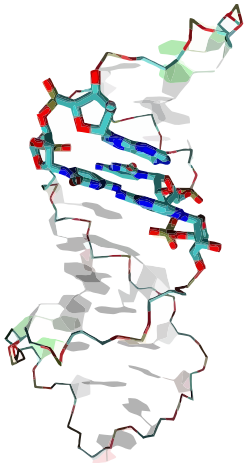
discrimininatory power of RNA family models

- at most one model should produce a hit for any given genome position
- compare *family models* with each other
- find weakly separated models
- higher specificity

# Example: A Clan of High-Scoring Families

- blue: clan members
- red: highest scoring non-member
- red scores: above *trusted cutoff*

extended RNA secondary structures

- canonical Watson-Crick base pairs are the most common pairs
- virtually every RNA structure contains *non-canonical* nucleotide pairs (Eric Westhof)
- non-canonical 2–1 pairings occur as well

parsing 2-diagrams
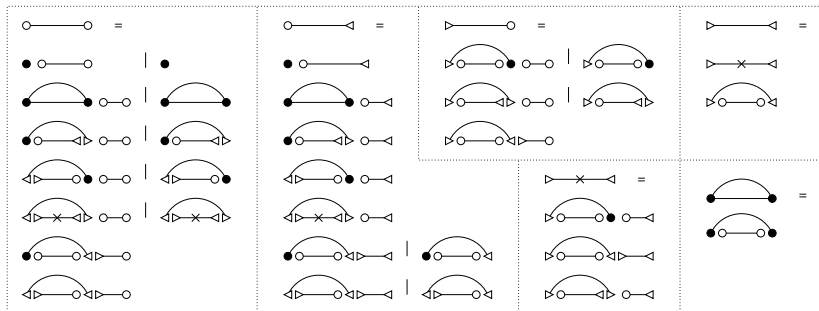
. . . and each pair can be **non-canonical**!

current problems & outlook

we are working on:

- full grammar with prior/posterior models based on *2-diagrams*
- large-scale training (RNAstrand, PDB)
- large-scale convex optimization ($\approx \mathbb{R}^{100000}$)
- derived algorithms

and later this year:

- improved index structures for genome-scale (ncRNA) scans
- *switchable* grammars for scanning algorithms (trace grammar)

📄 Höner zu Siederdissen C, Bernhart S, Stadler PF, Hofacker IL
   A Folding Algorithm for Extended RNA Secondary Structures
   submitted

📄 Höner zu Siederdissen C and Hofacker IL
   Discriminatory Power of RNA Family Models
   Bioinformatics (2010) 26 (18): i453-i459.

📄 Giegerich R and Höner zu Siederdissen C
   Semantics and Ambiguity of Stochastic RNA Family Models
   IEEE/ACM Transactions on Comp. Biology and Bioinformatics
   vol. 8, no. 2, pp. 499-516, Mar./Apr. 2011