

Computational comparative genomics for discovery of bacterial *cis*-regulatory RNAs

W.L. Ruzzo, Computer Science/Genome Sciences, Univ. of Washington, Seattle, WA

Fronting for: Zizhen Yao, Zasha Weinberg, Jeff Barrick, Adam Roth, Jane N. Kim, Jeremy Gore, Joy Xin Wang,

Elaine R. Lee, Kirsten F. Block, Narasimhan Sudarsan, Shane Neph, Martin Tompa, and Ronald R. Breaker.

Goal: Discovery of novel *cis*-regulatory noncoding RNA in bacteria

Multifaceted Challenge:

- Motif representation
- Motif inference
- Motif search/refinement (more examples/better models)
- Incorporation of relevant biological knowledge
- Expert analysis

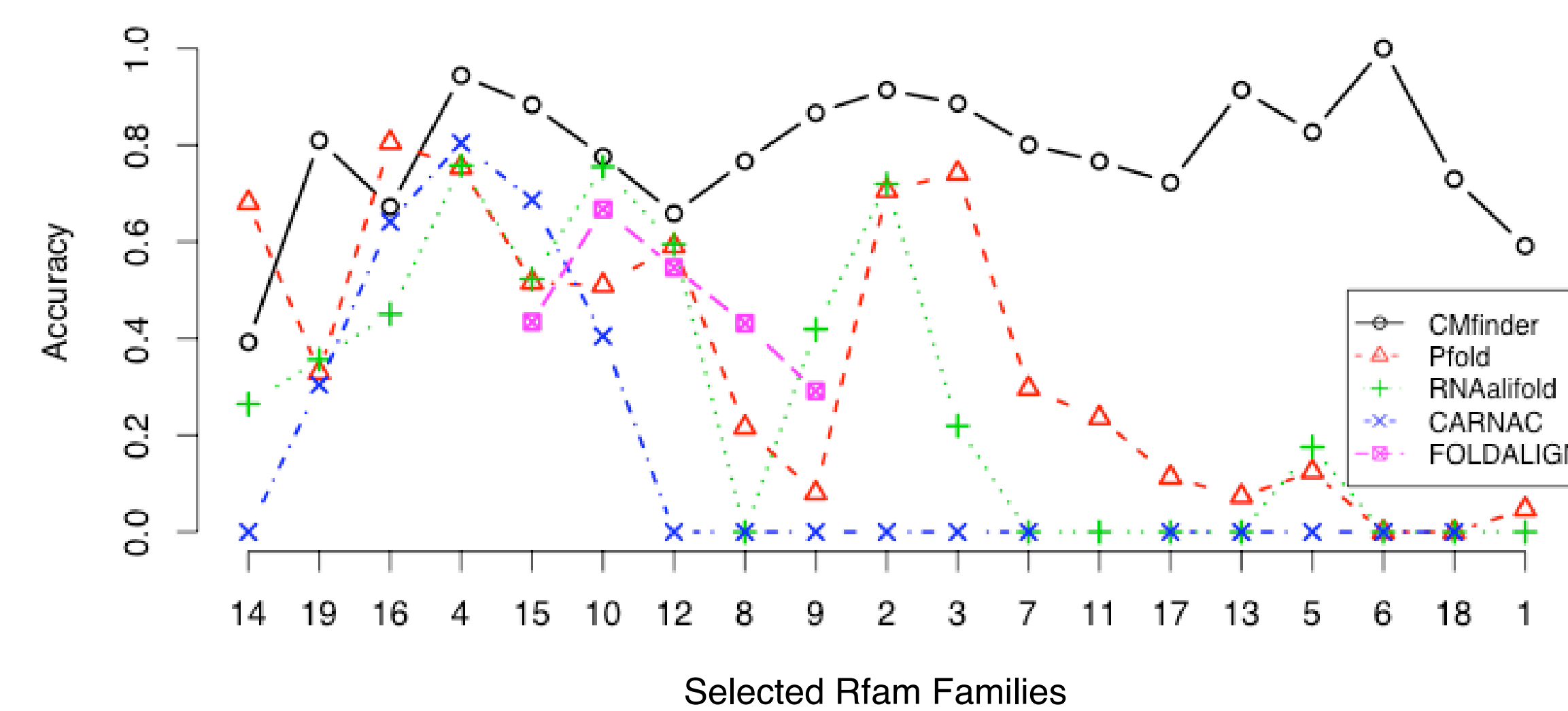
All are important, all need to be integrated.

Results: successful, automated pipeline attending to all this.

- 5/7 new likely riboswitch families experimentally verified
- 20 other elements in a wide variety of bacteria.
- involved in diverse but individually specific cellular processes, (citric acid cycle, ribosome, molybdenum cofactor, ...)
- more surprising: one widespread riboswitch in disparate processes: natural competence in *Vibrio cholerae* to electron transport in *Geobacter sulfurreducens*.

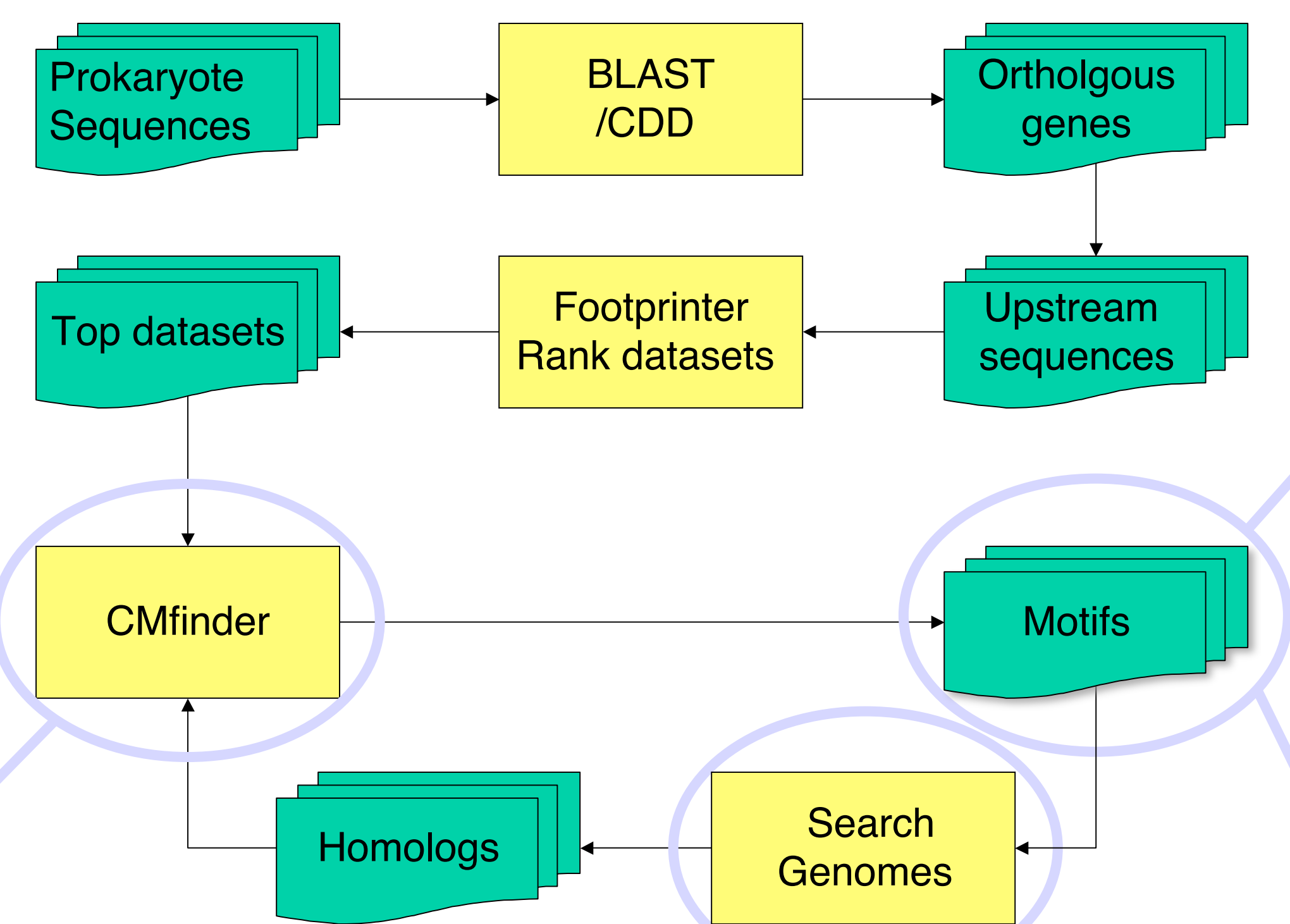
Motif Inference:

- Cmfinder–Motif inference in unaligned sequences with extraneous flanking regions (i.e., local alignment)
- CM as motif model
- Expectation-Maximization (EM) style iteration, with sophisticated initialization and structure heuristics



Yao, Weinberg and Ruzzo. Cmfinder--A Covariance Model Based RNA Motif Finding Algorithm. *Bioinformatics*, 2006, 22(4): 445-452.

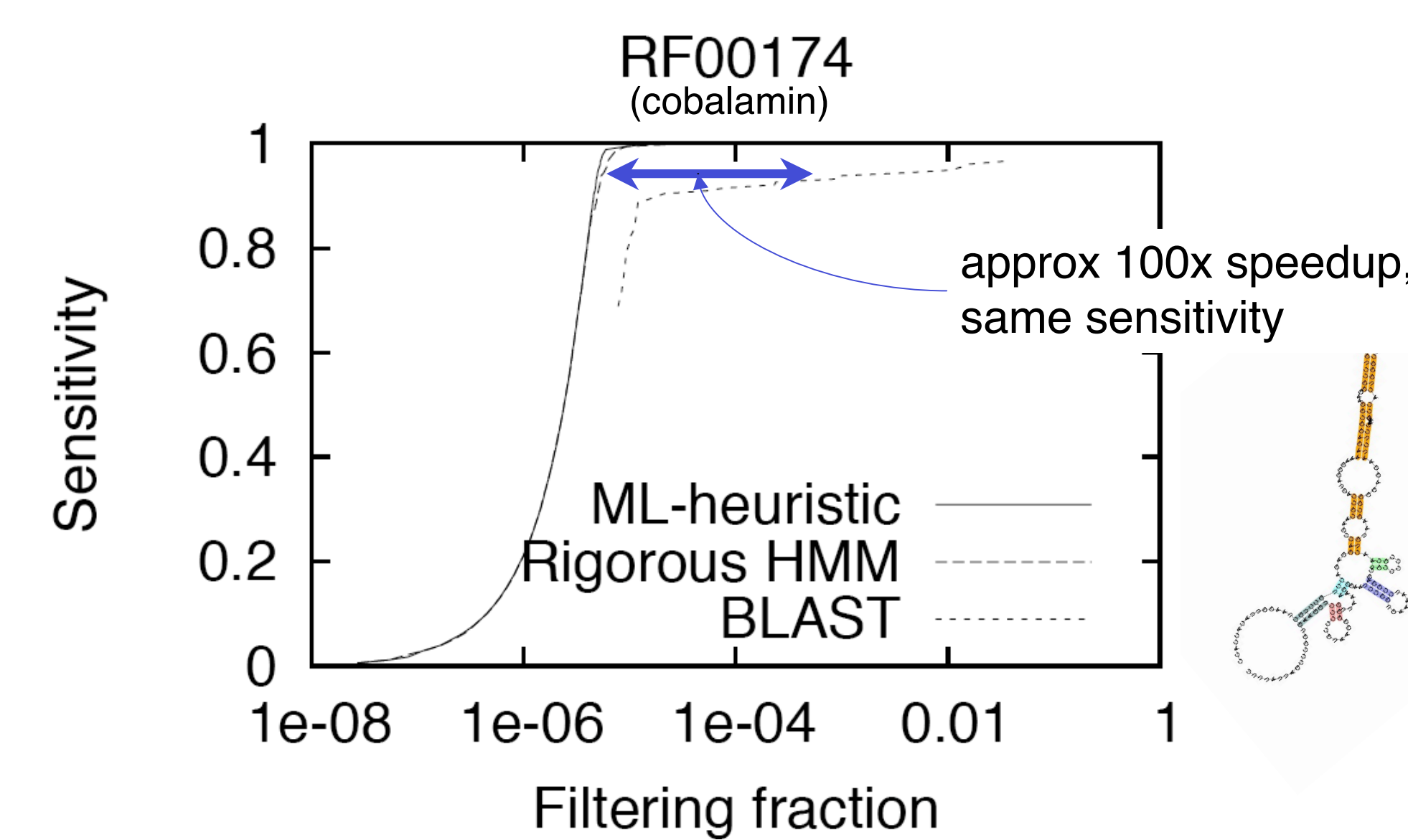
The Pipeline



Yao, Barrick, Weinberg, Neph, Breaker, Tompa and Ruzzo. A Computational Pipeline for High Throughput Discovery of *cis*-Regulatory Noncoding RNA in Prokaryotes. *PLoS Computational Biology*. 3(7): e126, July 6, 2007.

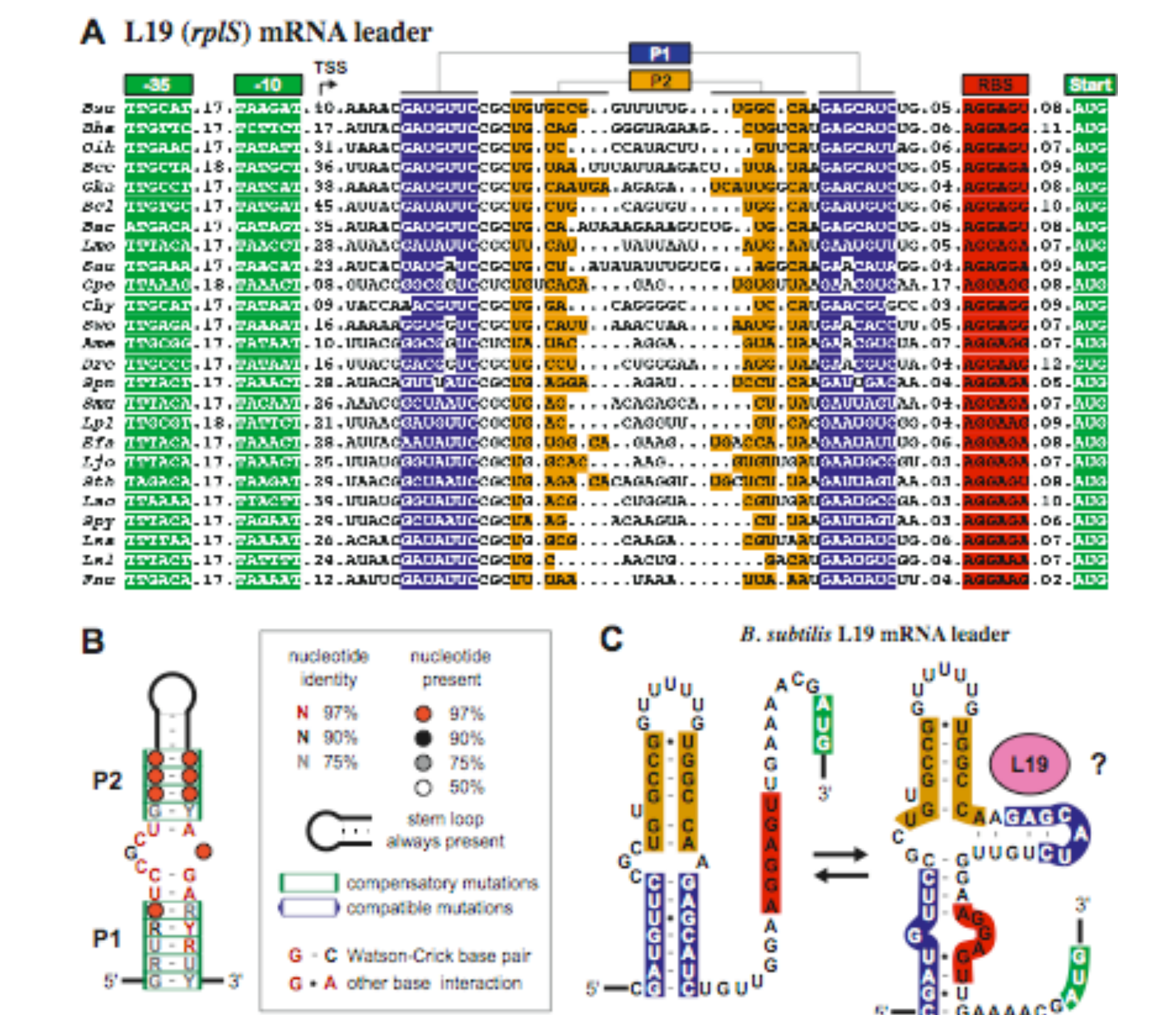
Motif Search:

- one RNA family often regulates multiple genes in a pathway, not just homologs
- More instances => better model
- RaveNnA–sensitive CM search, 100x faster
 - rigorous filtering–no sensitivity loss, provably
 - heuristic filtering–faster still, BLAST-LIKE sensitivity

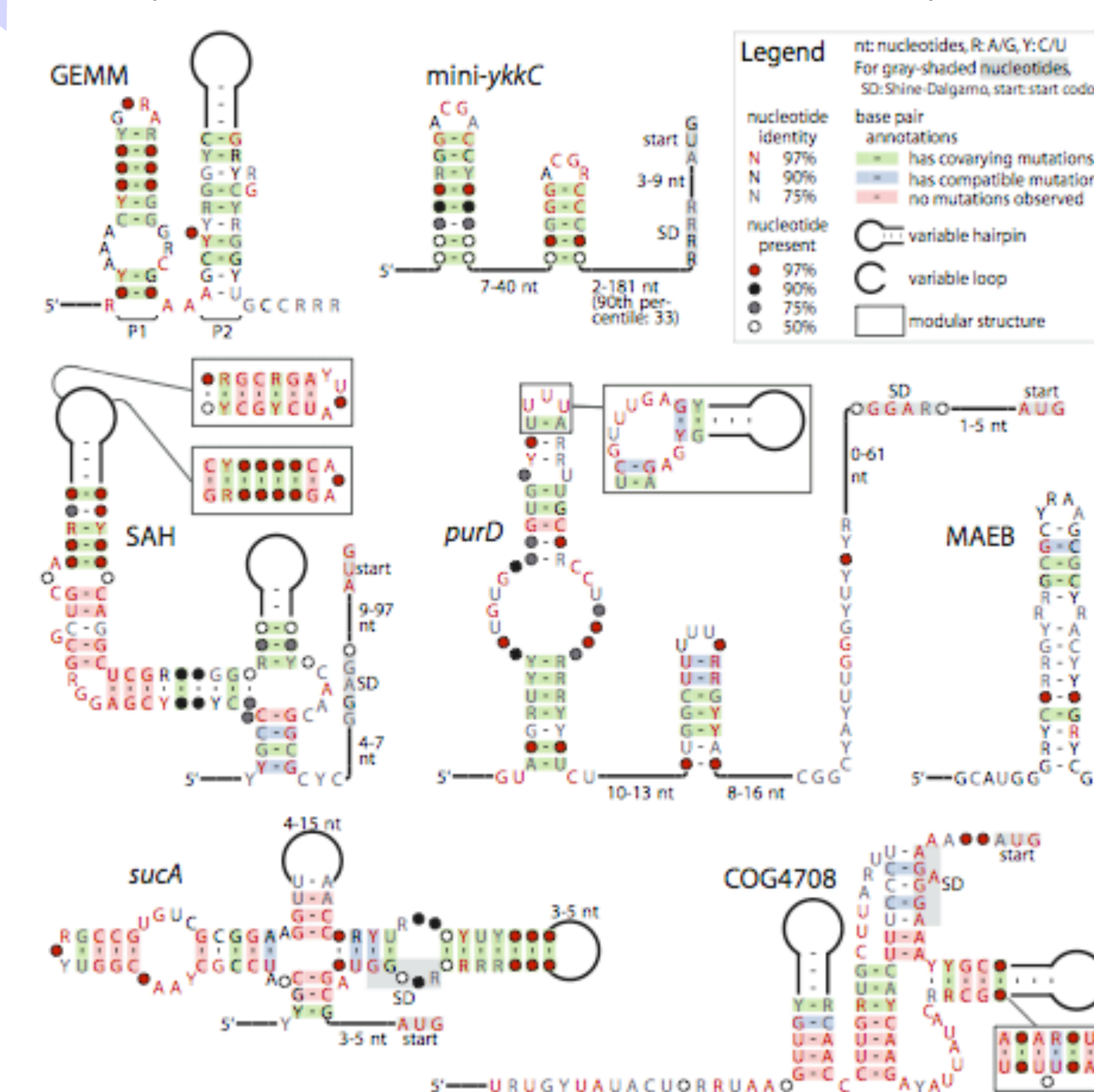


Weinberg and Ruzzo. Sequence-based heuristics for faster annotation of non-coding RNA families. *Bioinformatics*, 2006, 22(1):35-39

Example: Ribosomal Autoregulation: Excess L19 represses L19 (RF00556; 555-559 similar)



Examples: 7 Representative motifs (of 22; 5/22 confirmed riboswitches)



Weinberg, Barrick, Yao, Roth, Kim, Gore, Wang, Lee, Block, Sudarsan, Neph, Tompa, Ruzzo and Breaker. Identification of 22 candidate structured RNAs in bacteria using the Cmfinder comparative genomics pipeline. *Nucl. Acids Res.*, July 2007 35: 4809-4819.