0) Files description

Target file

The group you want to find candidate regions

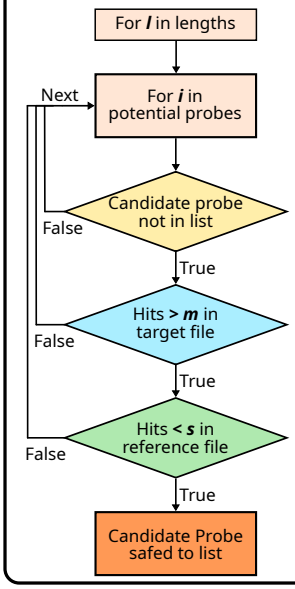
Reference file

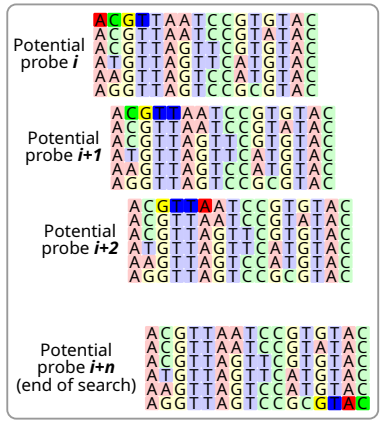
The sequences you want to exclude (or the rest of the diversity)

1) findPrimer

Searches candidate primers/probes by sliding window

findPrimer -t target **-r** reference **-l** length **-s** specificity **-m** minMatch



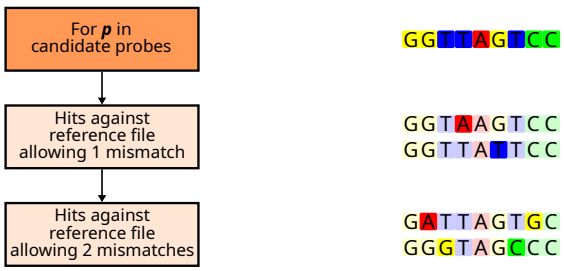


- Sequence
- Reverse complement sequence
- GC content
- Basic melting temperature
- Hits in target file
- Hits in reference file

2) testPrimer

Test primers/probes by allowing mismatches

testPrimer -f primers **-r** reference

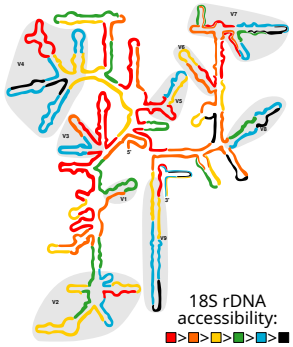


3) rateAccess

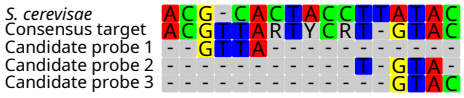
Compare candidate probes to the accessibility map of *Saccharomyces cerevisae* 18S rDNA (Behrens et al., 2003):

- Create consensus sequence of target group
- Align to the *S. cerevisae* 18S rDNA sequence
- Add candidate probes to the alignment

alignPrimers -t target.fasta **-p** probes.fasta **-o** probes_align.fasta



Behrens et al. (2003)



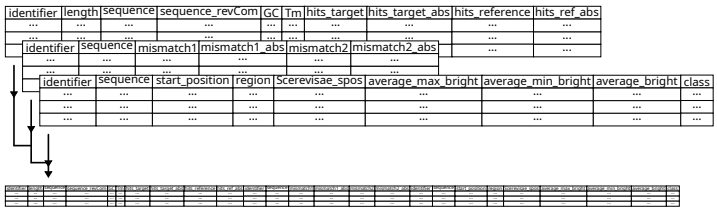
rateAccess -f probes_align.fasta **-o** probes_access.tsv

- Accessibility score
- Region in the 18S rDNA gene

4) Selecting probes

Merge all log files:

bindLogs -f probes.tsv probes_tested.tsv probes_access.tsv **-o** probes_log.tsv -r



Selecting the best probes:

- covering most of the targeted diversity
- with a high GC content
- with similar theoretical melting temperature
- with low hits to the reference file allowing mismatches
- highly accessible

For example:

filterLog -l probes_log.tsv **-s** "0.4" **-m** "0.001" **-M** "0.0001" **-c** "III"

selectLog -l probes_log_filtered.tsv **-n** "4"

5) Concluding remarks

Complementary softwares:

- **ARB** (SILVA; arb-home.de/)
- **BLAST** (blast.ncbi.nlm.nih.gov/Blast.cgi)
- **PR2-primers** (app.pr2-primers.org/)

Future implementations:

- **R2DT** (rnacentral.org/r2dt):
  Automatic visualization of 2D structure with the probes
- Check for **self-binding**: (e.g.) ACGTnnnnACGT

Remember that probe design is a tedious work that requires a final empirical test for its completion.
Therefore, bioinformatic pipelines will only provide theoretical candidate probes, that have to be tested in the laboratory