

DNA recombination

Biological background

Ciliates exhibit *nuclear dimorphism* through the presence of somatic macronuclei (**MAC**) and germline micronuclei (**MIC**).

The **MAC** controls metabolism, and works as a template for genes required for asexual growth. Its DNA is the one actively expressed and effectively results in the phenotype of the organism.

The **MIC** has reproductive functions, produces meiotic products during sexual reproduction (*conjugation*). Its DNA is passed during this phase.

The rearrangement process

In some ciliates - e.g. the species *Oxytricha trifallax* (genus *Oxytricha*, class Spirotrichea) - the MAC DNA is derived from the MIC DNA by extensive DNA recombination.

This process consists in:

- The MAC begins a copy of the MIC DNA. The chromosomes are fragmented and amplified. The result of this process is the *precursor*. ~90% of the complexity is lost.
 - Fragmentation
 - Amplification
- From the precursor the final MAC DNA is produced through these further operations:
 - Elimination
 - Inversion
 - Telomere Addition
 - Gene Scrambling / Unscrambling

The extent and the nature of these operations varies among ciliate species.

The precursor DNA pieces are called *Macronuclear-Destined Sequences* (MDSs). Precursor's MDSs are interrupted/separated by *Internally-Eliminated Sequences* (IESs): they mostly fall **between** genes in the *Tetrahymena* but in *Oxytricha* and *Paramecium* they interrupt them.

MDSs overlap slightly in the MAC (*pointers*).

Some complex IESs in *Oxytricha* can even contain MDSs for other genes or entire genes themselves (AV¹: how to handle overlap?).

Some MDSs rearrange during the MAC development (AV²: is this a different phase?).

This entire - error correcting - mechanism is guided by small RNAs and epigenetic chromatin marks (AV³: The resulting map is definitely correlated with this. What is the role of understanding this mechanism in our work?)

Instance

Input

- MIC (or precursor?) and MAC (AV⁴: should the sequences be already annotated with the MDSs, IESs and pointers?)

Respectively the input and the product genomes (Input and Output of the described DNA rearrangement process in the named species of Ciliates).

Admissible Solution

A rearrangement *map*: a set of disjoint substrings representing the building blocks (MDSs, IESs, pointers), eventual operations they will go through the process (scrambling, inversion/orientation, ..) and their “destination” on the MAC final produced genome.

An algorithm *g* should then be able to build the product genome having the input and the computed rearrangement map (and viceversa).

$f(\text{MIC}, \text{MAC}) = \text{Map}$
 $g(\text{MIC}, \text{Map}) = \text{MAC}$
 $g'(\text{MAC}, \text{Map}) = \text{MIC}$

AV⁵: Should the *precursor* building process be considered? i.e. the first input could be the MIC or the *precursor*:

MIC is fragmented and amplified to produce the *precursor*, should we accept the former, and include these operations in the solution (maybe in *phases*) or the latter, and accept directly the *precursor* as input?