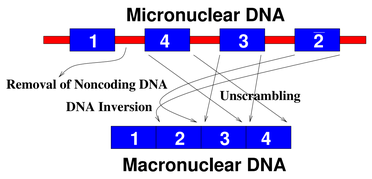
A Mathematical Model for Processes of Nested DNA Recombination

Ryan Arredondo - Jonathan Burns - University of South Florida

Biological Motivation

- DNA recombination occurs throughout many species, including humans. We study the ciliate species *O. Nova* as a model organism for DNA recombinant processes [1,5].

(a) Ciliates mating



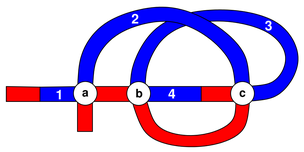
(b) Ciliate DNA recombination during mating

- Ciliates contain two types of nuclei: macronucleus (somatic) and micronucleus (germline).
- The micronuclear DNA contains segments of functional DNA that are interrupted by non-coding DNA and arranged in an order that is permuted relative to the somatic macronuclear DNA.
- After conjugation one of the micronuclei develops into a macronucleus through massive rearrangements of DNA involving thousands of genes.

Mathematical Model

The assembly graph model is a model for DNA recombination. The model suggests that micronuclear DNA is aligned at certain guiding segments called pointers [1].

- The DNA recombination process is represented by a set of edges (DNA segments) and vertices (points of alignment) called an assembly graph (Figure (c)).



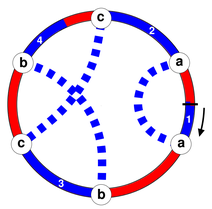
(c) Assembly graph modeling recombination of DNA from Figure (b)

- Assembly graphs satisfying certain properties can be represented by double occurrence words (Figure (d)), i.e. words in which each symbol appears exactly twice.

abcbca

(d) Double occurrence word representing the graph in Figure (c)

- Another way to visualize double occurrence words is with a chord diagram (Figure (e)) which consists of a circle with points labeled by symbols of the word and a set of chords connecting like symbols.



(e) Chord diagram for double occurrence word abcbca

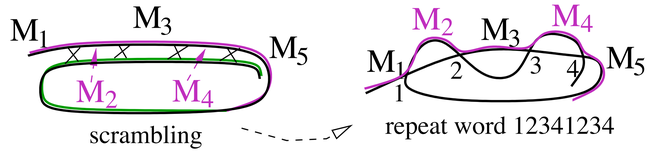
Nested Recombination

It is believed that some sequences of micronuclear DNA undergo recombination before others. These sequences appear in the scrambled DNA of certain ciliate species. There are two rearrangement patterns that appear most frequently in experimental data.

- The first sequence is of the form

$$(i)(i+1)(i+2)\cdots(i+n)(i)(i+1)(i+2)\cdots(i+n)$$

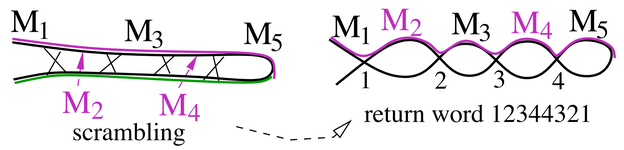
and is called a *repeat word*.



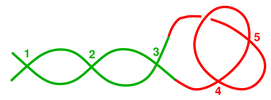
- The second sequence is of the form

$$(i)(i+1)(i+2)\cdots(i+n)(i+n)\cdots(i+2)(i+1)(i)$$

and is called a *return word*.



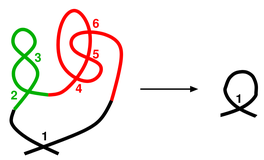
- Often these patterns associated with micronuclear DNA are nested within one another. For instance, the word **1234545321** has the repeat word **4545** nested within the return word **123321**.



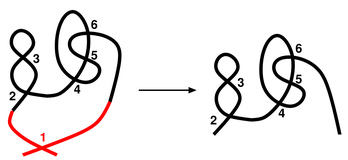
Measuring Complexity of Scrambled Genes

Since the scrambled micronluclear DNA corresponds to double occurrence words, we use operations defined on double occurrence words to model the steps in recombination of the nested sequences seen above.

- We define two reduction operations that act on double occurrence words:
  (Operation 1): Removal of all repeat words and return words,
  (Operation 2): Removal of a letter.



123324564561 $\xrightarrow{\text{Op. 1}}$ **11**



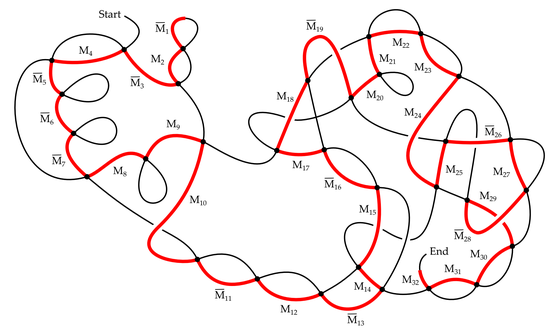
123324564561 $\xrightarrow{\text{Op. 2}}$ **2332456456**

- Given a double occurrence word w, the *nesting index of* w, $\mathrm{NI}(w)$, is defined to be the least number of reduction operations to be applied to w to obtain the empty word, i.e., the word with no symbols.
- Given a gene represented by a word w, $\mathrm{NI}(w)$ provides a measure of complexity for the rearrangement of that gene.

Nesting Index Algorithm

An algorithm to compute the nesting index of double occurrence words was implemented in C and in JavaScript.

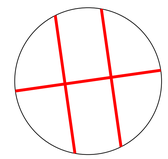
- For JavaScript with web interface please visit:
  `http://rarredon.myweb.usf.edu/reductions_ui.html`
- For greater computational power the C source code is readily available for download at:
  `http://knot.math.usf.edu/software/NI/nest_index.c`
- The assembly graph below represents a complex scrambled gene obtained from sequence analysis [4] by Landweber's Lab at Princeton University. Our program computed $\mathrm{NI}(w) = 15$ for this gene.



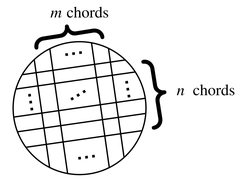
Results

- Note that some words do not contain repeat words or return words, for example, 123213. For this word we are then forced to apply reduction operation 2 in any reduction of w.
- We say a double occurrence word w is 1-*reducible*, if applying only reduction operation 1 to w some number of times will reduce w to the empty word.
- The following theorem characterizes when a gene should require step-wise rearrangement through reduction operations 1 and 2.

Theorem[2]**:** A double occurrence word w is 1-reducible if and only if the chord diagram of w does not contain the following as a sub-chord diagram.



Corollary[2]**:** Let $2 \leq n \leq m$ be integers. If w is a double occurrence word and the chord diagram of w has the following sub-chord diagram, then $\mathrm{NI}(w) \geq n+1$.



References

[1] A. Angeleska, N. Jonoska, M. Saito, L.F. Landweber, RNA-Guided DNA Assembly, J. of Theoretical Biology (2007) 248:706-720.

[2] R. Arredondo, Reductions on Double Occurrence Words, to appear in Congressus Numerantium.

[3] W.J. Chang, P.D. Bryson, H. Liang, M.K. Shin, L.F. Landweber, The evolutionary origin of a complex scrambled gene. Proc. Natl. Acad. Sci. (2005) 102:15149-15154

[4] X. Chen, et al., The architecture of a scrambled genome reveals a massive level of DNA rearrangement, submitted for publication.

[5] D.M. Prescott and A.F. Greslin, Scrambled actin I gene in the micronucleus of *Oxytricha nova*. Dev. Genet. (1992), 13: 66-74.

Acknowledgments

Many thanks to Dr. Egor Dolzhenko, Dr. Nataša Jonoska, Dr. Masahico Saito, and Tim Yeatman. For related work visit `http://knot.math.usf.edu`.
This work was supported in part by NSF grant DMS-0900671 and NIH grant 1R01GM109459-01

Ryan Arredondo - Graduate Student - Department of Mathematics and Statistics - University of South Florida Mail: `rarredon@mail.usf.edu` Web: `http://rarredon.myweb.usf.edu`