

ILP formulation

- q is an upper bound of the total quantity of MDS ($length/2$)
- $=$ is string equivalence
- $MIC[i, j]$ ($MAC[i, j]$) is the substring starting at i and finishing at j (i, j being positions) of the $MIC(MAC)$. Can be trivially defined using string concatenation and $MIC(i, c)$ ($MAC(i, c)$).
- $Inverse(String)$ is the Watson-Crick reverse complement of $String$
- Size of the Oxytricha Input genome: MIC is fragmented into $\sim 750\ 000$ MDSs, MAC into $300\ 000$.

$$objective\ function : \quad \min \sum_{i,j} MDS_{MACstart}(i,j)$$

$$Possible_{MDSMAC}(i,a,b) = \begin{cases} 0 \\ 1, & \text{if MDS } i \text{ can start at } a \text{ and finish at } b \text{ in the MAC} \end{cases}$$

$$Possible_{MDSMIC}(i,a,b) = \begin{cases} 0 \\ 1, & \text{if MDS } i \text{ can start at } a \text{ and finish at } b \text{ in the MIC} \end{cases}$$

$$Possible_{assignment}(a,b,c,d) = \begin{cases} 0 \\ 1, & \text{if } MIC[a,b] = MAC[c,d] \end{cases}$$

$$MDS_{MICstart}(i,j) = \begin{cases} 0 \\ 1, & \text{if MDS } i \text{ starts at position } j \text{ in the MIC} \end{cases}$$

$$MDS_{MICend}(i,j) = \begin{cases} 0 \\ 1, & \text{if MDS } i \text{ ends at position } j \text{ in the MIC} \end{cases}$$

$$MDS_{MACstart}(i,j) = \begin{cases} 0 \\ 1, & \text{if MDS } i \text{ starts at position } j \text{ in the MAC} \end{cases}$$

$$MDS_{MACend}(i,j) = \begin{cases} 0 \\ 1, & \text{if MDS } i \text{ ends at position } j \text{ in the MAC} \end{cases}$$

$$Inv(i) = \begin{cases} 0 \\ 1, & \text{if MDS } i \text{ is inverted in the MAC} \end{cases}$$

$$cwc(i,j,h,l) = \begin{cases} 0 \\ 1, & \text{if } MIC[i:j] \text{ is the reverse complement of } MAC[h:l] \end{cases}$$

$$MAC(i,c) = \begin{cases} 0 \\ 1, & \text{if } c \text{ is the character at position } i \text{ in the MAC} \end{cases}$$

$$MIC(i,c) = \begin{cases} 0 \\ 1, & \text{if } c \text{ is the character at position } i \text{ in the MIC} \end{cases}$$

$$IES(i) = \begin{cases} 0 \\ 1, & \text{if } i \text{ is part of an IES: } \sum_{j \leq i \leq k; 1 \leq a \leq q} MDS_{MICstart}(a,j) + MDS_{MICend}(a,k) = 0 \end{cases}$$

$$MDS_{MICstart}(i,a) + MDS_{MICend}(i,b) + MDS_{MACstart}(i,c) + MDS_{MACend}(i,d) + IES(i) = 5 \Rightarrow \\ MIC[a,b] = \text{Inverse}(\text{MAC}[c,d])$$

$$MDS_{MICstart}(i,a) + MDS_{MICend}(i,b) + MDS_{MACstart}(i,c) + MDS_{MACend}(i,d) = 4, IES(i) = 0 \Rightarrow \\ MIC[a,b] = \text{MAC}[c,d]$$

$$\sum_j MDS_{MICstart}(i,j) \leq 1$$

$$\sum_j MDS_{MICend}(i,j) = \sum_j MDS_{MICstart}(i,j)$$

$$Eq(i,j,h,l) = \begin{cases} 0 \\ 1, & \text{if } MIC[i:j] = MAC[h:l] \end{cases}$$

$$P_{start}(i,j) = \begin{cases} 0 \\ 1, & \text{if } MDS_{MACstart}(i,j) = 1, \text{ Pointer } i \text{ starts at position } j \text{ in the MAC} \end{cases}$$

$$P_{end}(i,j) = \begin{cases} 0 \\ 1, & \text{if } MDS_{MACend}(i-1,j) = 1, \text{ Pointer } i \text{ ends at position } j \text{ in the MAC} \end{cases}$$

$$Cov_{MIC}(i,j) = \begin{cases} 0 \\ 1, & \text{if MDS } i \text{ covers the position } j \text{ in the MIC} \end{cases}$$

$$Cov_{MIC}(i,j) = \begin{cases} 0 \\ 1, & \text{if } \sum_{l \leq i} MDS_{MICstart}(i,l) + \sum_{l < i} MDS_{MICend}(i,l) = 1 \end{cases}$$

$$Cov_{MAC}(i,j) = \begin{cases} 0 \\ 1, & \text{if MDS } i \text{ covers the position } j \text{ in the MAC} \end{cases}$$

$$Cov_{MAC}(i,j) = \begin{cases} 0 \\ 1, & \text{if } \sum_{l \leq i} MDS_{MACstart}(i,l) + \sum_{l < i} MDS_{MACend}(i,l) = 1 \end{cases}$$