

Cassette Array Grammar

The following are the grammar rules used to annotate cassette arrays and discover cassettes. Terminal tokens matched by name are in quotes, otherwise token symbols refer to semantic type. Directions on all left hand side tokens have to agree except for λ tokens that have no directionality. If direction is not noted, either direction can be used as long as the direction for all tokens is the same. Square brackets denote a definite clause range for the length of a λ token.

1 Structure Rules

Cassette	::=	"aadA \mathcal{L} " ... 214 cassettes ...
NonCassetteInsertion	::=	"attI1 L-spacer" "IS1" ... 38 features ...
ArrayEnd	::=	"3'-CS" "class 2 3'" "class 3 3'" "ybeA"
ArrayStart	::=	"5'-CS" "class 2 5'" "class 3 5'"
ArrayMid	::=	Cassette
ArrayEnd	::=	ArrayEnd ArrayEnd
ArrayStart	::=	ArrayStart ArrayStart
ArrayMid	::=	ArrayMid ArrayMid
$\overrightarrow{\text{Array}}$::=	$\overrightarrow{\text{Array}}$ $\overrightarrow{\text{ArrayEnd}}$
$\overleftarrow{\text{Array}}$::=	$\overleftarrow{\text{Array}}$ $\overleftarrow{\text{ArrayStart}}$
$\overrightarrow{\text{Array}}$::=	$\overrightarrow{\text{ArrayStart}}$ $\overrightarrow{\text{ArrayEnd}}$
$\overleftarrow{\text{Array}}$::=	$\overleftarrow{\text{ArrayEnd}}$ $\overleftarrow{\text{ArrayStart}}$
$\overrightarrow{\text{ArrayStart}}$ $\overrightarrow{\text{ArrayMid}}$::=	$\overrightarrow{\text{Array}}$ $\overrightarrow{\text{ArrayMid}}$
ArrayMid $\overrightarrow{\text{ArrayMid}}$::=	NonCassetteInsertion $\overrightarrow{\text{ArrayMid}}$
ArrayMid $\overleftarrow{\text{ArrayMid}}$::=	ArrayMid NonCassetteInsertion
$\overleftarrow{\text{ArrayEnd}}$ $\overleftarrow{\text{ArrayMid}}$::=	$\overleftarrow{\text{"tni"}}$ $\overleftarrow{\text{ArrayMid}}$
$\overrightarrow{\text{ArrayEnd}}$ $\overrightarrow{\text{ArrayMid}}$::=	$\overrightarrow{\text{ArrayMid}}$ $\overrightarrow{\text{"tni"}}$
$\overrightarrow{\text{Array}}$::=	$\overrightarrow{\text{ArrayStart}}$ $\overrightarrow{\text{ArrayMid}}$ $\overrightarrow{\text{ArrayEnd}}$
$\overleftarrow{\text{Array}}$::=	$\overleftarrow{\text{ArrayEnd}}$ $\overleftarrow{\text{ArrayMid}}$ $\overleftarrow{\text{ArrayStart}}$
$\overrightarrow{\text{ArrayMid}}$ λ $\overrightarrow{\text{ArrayEnd}}$::=	$\overrightarrow{\text{ArrayMid}}$ λ $\overrightarrow{\text{"tni"}}$
$\overleftarrow{\text{ArrayEnd}}$ λ $\overleftarrow{\text{ArrayMid}}$::=	$\overleftarrow{\text{"tni"}}$ λ $\overleftarrow{\text{ArrayMid}}$

The cassette carrying the *qacE* gene is identical in its first 390 bp to a conserved sequence typically marking the end of an array in a class 1 integron ("3'-CS"). If a truncated version of the cassette with less than 390 bp is found, the cassette is incorrectly labelled as a truncated "3'-CS" by the annotator. The rule $\overrightarrow{\text{ArrayStart}}\overrightarrow{\text{ArrayMid}} ::= \overrightarrow{\text{Array}}\overrightarrow{\text{ArrayMid}}$ corrects the annotation by relabelling the truncated "3'-CS" to a truncated *qacE* gene cassette so that the cassette array is then correctly annotated.

2 Discovery Rules

$$\begin{array}{l}
 \overrightarrow{\text{ArrayStart}} \overrightarrow{\text{Cassette}} \overrightarrow{\text{ArrayEnd}} ::= \overrightarrow{\text{ArrayStart}} \lambda[300-1860] \overrightarrow{\text{ArrayEnd}} \\
 \overleftarrow{\text{ArrayStart}} \overleftarrow{\text{Cassette}} \overleftarrow{\text{ArrayEnd}} ::= \overleftarrow{\text{ArrayStart}} \lambda[300-1860] \overleftarrow{\text{ArrayEnd}} \\
 \overrightarrow{\text{ArrayStart}} \overrightarrow{\text{Cassette}} \overrightarrow{\text{ArrayMid}} ::= \overrightarrow{\text{ArrayStart}} \lambda[300-1860] \overrightarrow{\text{ArrayMid}} \\
 \overleftarrow{\text{ArrayMid}} \overleftarrow{\text{Cassette}} \overleftarrow{\text{ArrayStart}} ::= \overleftarrow{\text{ArrayMid}} \lambda[300-1860] \overleftarrow{\text{ArrayStart}} \\
 \overrightarrow{\text{ArrayMid}} \overrightarrow{\text{Cassette}} \overrightarrow{\text{ArrayEnd}} ::= \overrightarrow{\text{ArrayMid}} \lambda[300-1860] \overrightarrow{\text{ArrayEnd}} \\
 \overleftarrow{\text{ArrayEnd}} \overleftarrow{\text{Cassette}} \overleftarrow{\text{ArrayMid}} ::= \overleftarrow{\text{ArrayEnd}} \lambda[300-1860] \overleftarrow{\text{ArrayMid}} \\
 \text{ArrayMid} \text{Cassette} \text{ArrayMid} ::= \text{ArrayMid} \lambda[300-1860] \text{ArrayMid}
 \end{array}$$