Automata for DNA Splicing Languages with Two Restriction Enzymes

Nurul Izzaty Ismail[1], Wan Heng Fong[1]* and Nor Haniza Sarmin[1]

[1] *Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia.*

DNA splicing system is modelled by the process of recombinant DNA that generates a language which is known as a splicing language. In DNA splicing system, molecules are cut and recombined with the existence of restriction enzymes and a ligase to produce new molecules. The molecules resulting from the splicing system depict the splicing language using formal language theory. In nature, the sequences of restriction enzyme can be recognised as palindromic or non-palindromic, where palindrome is a sequence of string which reads the same forwards and backwards. Research on splicing languages from DNA splicing systems has been done based on palindromic and non-palindromic restriction enzymes. In this research, the relation between DNA splicing language and concepts in automata theory is discussed. The splicing languages from the splicing system with one cutting site each of two palindromic or two non-palindromic restriction enzymes are deduced from the grammars. Then, the graphical representations of the languages generated by the grammars depicting the splicing languages are visualised as automata diagrams.

Keywords: Automata; DNA; Splicing language; Splicing system

I.   INTRODUCTION

In 1987, (Head, 1987) introduced a study relating between formal language theory and informational macromolecules which is called a Deoxyribonucleic acid (DNA) splicing system. The splicing system is simulated by the process of recombinant DNA where restriction enzymes and a ligase cut and recombine DNA molecules to produce further molecules. The molecules resulting from the splicing system generate a language using formal language theory which is named as DNA splicing language (Head, 1987). In formal language, a set of strings consists of symbols derived from an alphabet. Some notations of regular expression in formal languages are applied in this research such as empty string λ, empty set ∅, union +, star-closure * and parentheses {} or () (Linz, 2006). A grammar G is a set of production rules for strings in

formal language (Linz, 2006). The grammar generates strings by arranging the production rules in sequential order, known as a language, generated by the grammar.

The splicing system is made up of three sets via formal language theory. The first set consists of double stranded DNA (dsDNA) symbols formed by nitrogenous base pairings: adenine (A) pairs with thymine (T), while cytosine (C) pairs with guanine (G) (Paun et al., 1998). Next, the second set is the set containing initial DNA molecules or strings which is obtained from some fragments of protein or nucleotide sequences (Brejová et al., 2000). The third set involves the cleavage of restriction enzymes which acts as rules in the form of a triple (Head, 1998).

From splicing systems, the resulting splicing languages can be generalised based on the sequences of restriction enzymes. Palindrome refers to the sequence of

_____

*Corresponding author's e-mail: fwh@utm.my

restriction enzyme that reads the same forwards and backwards (Tomohiro, 2013). Previously, research on the generalised splicing languages resulting from DNA splicing systems involving palindromic and non-palindromic restriction enzymes has been done (Ismail et al., 2019, Fong, 2019). The name, sequence and cleavage for all restriction enzymes used in this research are obtained from the New England Biolabs (NEB) catalogue (New England Biolabs Inc., 2017).

This research focuses on the relation between DNA splicing systems and the concepts of automata. An automaton is constructed with a set of states where the current state can either accept or reject the input when moving to the next state through the transition functions (Linz, 2006). The inputs accepted from every transition depict the language generated by the automaton.

In this research, the graphical representations for DNA splicing systems with two palindromic or two non-palindromic rules for restriction enzymes with the same and different crossings are illustrated as deterministic finite automata using grammars.

II. PRELIMINARIES

In this research, the modelling of DNA splicing systems with palindromic and non-palindromic rules is done using Head's splicing model. The definitions of Head's splicing model and splicing language are stated in the following.

Definition 1 (Head, 1987) Splicing System and Splicing Language

A splicing system S = (A, I, B, C) consists of a finite alphabet A, a finite set I of initial strings in A^*, and finite sets B and C of triples (c, x, d) with c, x and d in A^*. Each such triple in B or C is called a pattern. For each such triple the string cxd is called a site and the string x is called a crossing. Patterns in B are called left patterns and patterns in C are called right patterns. The language L = $L(S)$ generated by S consists of the strings in I and all strings that can be obtained by adjoining to $ucxfq$ and $pexdv$ whenever $ucxdv$ and $pexfq$ are in L and (c, x, d) and (e, x, f) are patterns of the same hand. A language, L is a splicing language if there exists a splicing system S for

which L = $L(S)$.

Next, the definition of a palindromic string is presented.

Definition 2 (Yusof, 2012) Palindromic String

A string I of a dsDNA is said to be palindromic if the sequence from the left to the right side of the upper single strand is equal to the sequence from the right to the left side of the lower single strand.

The definition of a deterministic finite automaton is presented next.

Definition 3 (Linz, 2006) Deterministic Finite Automaton

A deterministic finite automaton M is a 5-tuple, (Q, Σ, δ, q_0, F) consisting of a finite set of states Q, a finite set of input symbols called the alphabet Σ, a transition function (δ: $Q \times \Sigma \rightarrow Q$), an initial state $q_0 \in Q$ and a set of final states $F \subseteq Q$.

Lastly, the definition of a grammar is shown below.

Definition 4 (Linz, 2006) Grammar

A grammar G is defined as a quadruple G = (V, T, S, P), where V a is a finite set of objects called variables, T is a finite set of objects called terminal symbols, $S \in V$ is a special symbol called the start variable and P is a finite set of productions.

The set $L(G)$ = $\{w \in T^* : S \stackrel{*}{\Rightarrow} w\}$ is the language generated by G, where $\stackrel{*}{\Rightarrow}$ denotes zero or more steps of sequence of productions.

Figure 1 shows an example of a deterministic finite automaton that accepts the language $L((a+b) \cdot c^*)$ generated by the grammar with P consisting of the productions

$$S_0 \rightarrow aS_1 \mid bS_1,$$

$$S_1 \rightarrow cS_1 \mid \lambda,$$

where Q = $\{S_0, S_1\}$, Σ = $\{a, b, c\}$, S_0 is the initial state, F = $\{S_1\}$ and δ is given by

$$\delta (S_0, a) = S_1,$$

$$\delta\,(S_0, b) = S_1,$$
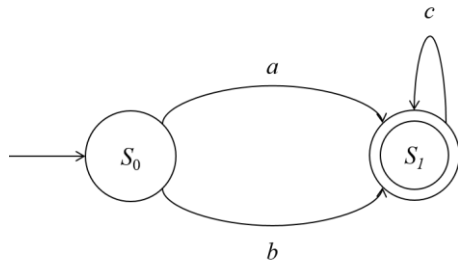
$$\delta\,(S_1, c) = S_1.$$

Figure 1.   A deterministic finite automaton

The automaton makes a move from the current state to another state based on the transition function and accepts the input symbols. In the automaton, single circle, double circle, arrows and arrow labels denote as the state, final state, transitions, and input symbols respectively.

III.   RESEARCH METHODOLOGY

The graphical representations for DNA splicing systems with one cutting site each of two palindromic or two non-palindromic rules for the same and different crossings are visualised as automata diagrams using the generalised splicing languages obtained from the respective DNA splicing systems.

Table 1 shows the generalised splicing languages from DNA splicing systems with two palindromic rules where Theorem 1 involves the same crossing and Theorem 2 involves different crossings of palindromic rules. While Table 2 shows the generalised splicing languages from DNA splicing systems with two non-palindromic rules where Theorem 3 involves the same crossing and Theorem 4 involves different crossings of non-palindromic rules.

Table 1. Generalised Splicing Languages from DNA Splicing Systems with Two Palindromic Rules

Theorem	Theorem 1 (Ismail et al., 2019)	Theorem 2 (Ismail et al., 2019)
Initial String	$N_1 N_1 \ldots N_1 X_1\, Y\, X_2 M\, M \ldots M W_1\, Y W_2 N_2 N_2 \ldots N_2$ $N_1' N_1' \ldots N_1' X_1' Y' X_2' M' M' \ldots M' W_1' Y' W_2' N_2' N_2' \ldots N_2'$	$N_1 N_1 \ldots N_1 X_1\, Y\, X_2 M\, M \ldots M W_1\, Z W_2 N_2 N_2 \ldots N_2$ $N_1' N_1' \ldots N_1' X_1' Y' X_2' M' M' \ldots M' W_1' Z' W_2' N_2' N_2' \ldots N_2'$
First Restriction Enzyme	$X_1 Y X_2$ $X_1' Y' X_2'$	$X_1 Y X_2$ $X_1' Y' X_2'$
Second Restriction Enzyme	$W_1\, Y W_2$ $W_1' Y' W_2'$	$W_1\, Z W_2$ $W_1' Z' W_2'$
Splicing Language	$\left(\begin{matrix} N_1 N_1 \ldots N_1 X_1 \\ N_1' N_1' \ldots N_1' X_1' \end{matrix} + \begin{matrix} N_2' N_2' \ldots N_2' W_1 \\ N_2 N_2 \ldots N_2\, W_1' \end{matrix} \right)$ $\left\{ \begin{matrix} Y \\ Y' \end{matrix} \left(\begin{matrix} X_2 M\, M \ldots M\, W_1 \\ X_2' M' M' \ldots M' W_1' \end{matrix} + \begin{matrix} W_2 M' M' \ldots M' X_1 \\ W_2'\, M\, M \ldots M\, X_1' \end{matrix} \right) \right\}^{*}$ $\begin{matrix} Y \\ Y' \end{matrix} \left(\begin{matrix} W_2 N_2 N_2 \ldots N_2 \\ W_2' N_2' N_2' \ldots N_2' \end{matrix} + \begin{matrix} X_2 N_1' N_1' \ldots N_1' \\ X_2' N_1 N_1 \ldots N_1 \end{matrix} \right)$	$\left\{ \begin{matrix} N_1 N_1 \ldots N_1 X_1\, Y\, X_2 \\ N_1' N_1' \ldots N_1' X_1' Y' X_2' \\ \left(\begin{matrix} M\, M \ldots M\, W_1\, Z W_2 M' M' \ldots M'\, X_1 Y X_2 \\ M' M' \ldots M' W_1' Z' W_2' M\, M \ldots M\, X_1' Y' X_2' \end{matrix} \right)^{*} \\ \left(\begin{matrix} N_1' N_1' \ldots N_1' \\ N_1 N_1 \ldots N_1 \end{matrix} + \begin{matrix} M\, M \ldots M\, W_1\, Z W_2 N_2 N_2 \ldots N_2 \\ M' M' \ldots M' W_1' Z' W_2' N_2' N_2' \ldots N_2' \end{matrix} \right) \end{matrix} \right\} +$ $\left\{ \begin{matrix} N_2' N_2' \ldots N_2'\, W_1\, Z W_2 \\ N_2 N_2 \ldots N_2\, W_1' Z' W_2' \\ \left(\begin{matrix} M' M' \ldots M'\, X_1 Y X_2\, M\, M \ldots M\, W_1\, Z W_2 \\ M\, M \ldots M\, X_1' Y' X_2' M' M' \ldots M' W_1' Z' W_2' \end{matrix} \right)^{*} \\ \left(\begin{matrix} N_2 N_2 \ldots N_2 \\ N_2' N_2' \ldots N_2' \end{matrix} + \begin{matrix} M' M' \ldots M'\, X_1 Y X_2\, N_1 N_1 \ldots N_1 \\ M\, M \ldots M\, X_1' Y' X_2' N_1' N_1' \ldots N_1' \end{matrix} \right) \end{matrix} \right\}$

Table 2. Generalised Splicing Languages from DNA Splicing Systems with Two Non-Palindromic Rules

Theorem	Theorem 3 (Fong et al.,2019)	Theorem 4 (Fong et al.,2019)
Initial String	$N_1 N_1 \ldots N_1 X_1\, Y\, X_2 M\, M \ldots M W_1\, Y W_2 N_2 N_2 \ldots N_2$ $N_1' N_1' \ldots N_1' X_1' Y' X_2' M' M' \ldots M' W_1' Y' W_2' N_2' N_2' \ldots N_2'$	$N_1 N_1 \ldots N_1 X_1\, Y\, X_2 M\, M \ldots M W_1\, Z W_2 N_2 N_2 \ldots N_2$ $N_1' N_1' \ldots N_1' X_1' Y' X_2' M' M' \ldots M' W_1' Z' W_2' N_2' N_2' \ldots N_2'$
First Restriction Enzyme	$X_1 Y X_2$ $X_1' Y' X_2'$	$X_1 Y X_2$ $X_1' Y' X_2'$
Second Restriction Enzyme	$W_1\, Y W_2$ $W_1' Y' W_2'$	$W_1\, Z W_2$ $W_1' Z' W_2'$

Splicing Language	$N_1 N_1 \dots N_1 X_1 \begin{pmatrix} Y X_2 \ M \ M \ \dots \ M \ W_1 \\ Y' X_2' M' M' \dots M' W_1' \end{pmatrix}^* \begin{matrix} Y W_2 \ N_2 N_2 \dots N_2 \\ Y' W_2' N_2' N_2' \dots N_2' \end{matrix}$	$\begin{matrix} N_1 N_1 \ \dots N_1 X_1 \ Y \ X_2 M \ M \dots M W_1 \ Z W_2 N_2 N_2 \dots N_2 \\ N_1' N_1' \dots N_1' X_1' Y' X_2' M' M' \dots M' W_1' Z' W_2' N_2' N_2' \dots N_2' \end{matrix}$

The symbols $\begin{matrix} N_1 & X_1 & Y & X_2 & M & W_1 & Z & W_2 \\ N_1'' & X_1'' & Y'' & X_2'' & M'' & W_1'' & Z'' & W_2'' \end{matrix}$ and $\begin{matrix} N_2 \\ N_2' \end{matrix}$ denote arbitrary dsDNA symbol(s), where N_1', X_1', Y', X_2', M', W_1', Z', W_2' and N_2' are complementaries for N_1, X_1, Y, X_2, M, W_1, Z, W_2, and N_2 respectively, and $\frac{Y}{Y'}$ and $\frac{Z}{Z'}$ are the crossings.

IV.   RESULTS AND DISCUSSIONS

In this research, the automata diagrams for DNA splicing systems with one cutting site each of two palindromic or two non-palindromic rules for the same and different crossings are constructed using grammars where the splicing languages from the splicing system are deduced from the grammars.

The automaton for DNA splicing system with one cutting site each of two palindromic restriction enzymes with the same crossing is presented as Theorem 5.

Theorem 5

Given
$$S =$$
$$\left(\begin{Bmatrix} A, C, G, T \\ T \ \ G \ \ C \ \ A \end{Bmatrix}, \begin{matrix} N_1 N_1 \dots N_1 X_1 \ Y \ X_2 M \ \ M \dots M W_1 \ Y W_2 N_2 N_2 \dots N_2 \\ N_1' N_1' \dots N_1' X_1' Y' X_2' M' \ M' \dots M' W_1' Y' W_2' N_2' N_2' \dots N_2' \end{matrix}, \right.$$
$$\left. \begin{Bmatrix} \begin{pmatrix} X_1 & Y & X_2 \\ X_1' & Y'' & X_2' \end{pmatrix}, \begin{pmatrix} W_1 & Y & W_2 \\ W_1'' & Y'' & W_2' \end{pmatrix} \end{Bmatrix}, \emptyset \right)$$

is a splicing system involving one cutting site each of two palindromic rules $\frac{X_1 Y X_2}{X_1' Y' X_2'}$ and $\frac{W_1 Y W_2}{W_1' Y' W_2'}$ with the same crossing $\frac{Y}{Y'}$ where $\begin{matrix} N_1 & X_1 & Y & X_2 & M & W_1 & W_2 \\ N_1'' & X_1'' & Y' & X_2'' & M' & W_1'' & W_2' \end{matrix}$ and $\frac{N_2}{N_2'}$ are variables used to denote any arbitrary dsDNA and N_1', X_1', Y', X_2', M', W_1', W_2' and N_2' are complementaries for N_1, X_1, Y, X_2, M, W_1, W_2, and N_2 respectively, $M_1 = (Q, \Sigma, \delta, q_0, F)$ is a deterministic finite automaton for the DNA splicing system that accepts the language $L(S)$, in which $Q = \{q_0, q_1, q_2, q_3, q_4, q_5\}$ is the set of states where q_0 is the initial state and $F = \{q_4, q_5\}$ is the set of final states, $\Sigma =$
$$\begin{Bmatrix} N_1 N_1 \ \dots N_1 X_1 & N_2' N_2' \dots N_2' W_1 & X_2 \ M \ M \dots \ M \ W_1 & Y \\ N_1' N_1' \dots N_1' X_1' & N_2 N_2 \dots N_2 \ W_1' & X_2' M' M' \dots M' W_1' & Y'' \\ W_2 M' M' \dots M' X_1 & W_2 N_2 N_2 \dots N_2 & X_2 N_1' N_1' \dots N_1' & \\ W_2' M \ M \ \dots M \ X_1' & W_2' N_2' N_2' \dots N_2'' & X_2' N_1 N_1 \dots N_1 & \end{Bmatrix}$$
is the set of inputs and δ are given by illustrating the automata diagram in Figure 2.
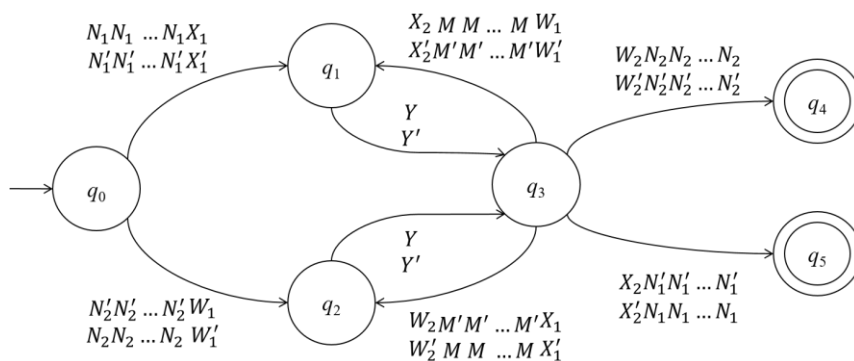

Figure 2. An automaton diagram for M_1

Proof

The splicing language $L(S)$ from the splicing system can be written as a language generated by a grammar G_1 where $G_1 = (\{S_0, S_1, S_2, S_3\}, \begin{Bmatrix} N_1 & X_1 & Y & X_2 \\ N_1'' & X_1'' & Y'' & X_2'' \end{Bmatrix}$ $\begin{matrix} M & W_1 & W_2 & N_2 \\ M'' & W_1'' & W_2'' & N_2'' \end{matrix}\}, S_0, P_1)$ with P_1 consisting of the

productions,

$$S_0 \to \begin{matrix} N_1 N_1 \ \dots N_1 X_1 \\ N_1' N_1' \dots N_1' X_1' \end{matrix} S_1 \mid \begin{matrix} N_2' N_2' \dots N_2' W_1 \\ N_2 N_2 \dots N_2 \ W_1' \end{matrix} S_2,$$

$$S_1 \to \begin{matrix} Y \\ Y' \end{matrix} S_3,$$

$$S_2 \to \begin{matrix} Y \\ Y' \end{matrix} S_3 \text{ and}$$

$$S_3 \to \begin{matrix} X_2 \ M \ M \dots \ M \ W_1 \\ X_2' M' M' \dots M' W_1' \end{matrix} S_1 \mid \begin{matrix} W_2 M' M' \dots M' X_1 \\ W_2' M \ M \ \dots M \ X_1' \end{matrix} S_2 \mid$$

$$\left.\begin{matrix} W_2N_2N_2\dots N_2 \\ W_2'N_2'N_2'\dots N_2' \end{matrix}\right| \begin{matrix} X_2N_1'N_1'\dots N_1' \\ X_2'N_1N_1\dots N_1 \end{matrix}.$$

Then, a sequence for the language generated by the grammar G_1, $L(G_1)$ is

$$S_0 \Rightarrow \begin{pmatrix} N_1N_1\ \dots N_1X_1 \\ N_1'N_1'\dots N_1'X_1' \end{pmatrix} S_1 + \begin{matrix} N_2'N_2'\dots N_2'\,W_1 \\ N_2N_2\dots N_2\ W_1' \end{matrix} S_2 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} N_1N_1\ \dots N_1X_1 \\ N_1'N_1'\dots N_1'X_1' \end{pmatrix} + \begin{matrix} N_2'N_2'\dots N_2'\,W_1 \\ N_2N_2\dots N_2\ W_1' \end{matrix} \begin{pmatrix} Y \\ Y' \end{pmatrix} S_3$$

$$\Rightarrow \begin{pmatrix} N_1N_1\ \dots N_1X_1 \\ N_1'N_1'\dots N_1'X_1' \end{pmatrix} + \begin{matrix} N_2'N_2'\dots N_2'\,W_1 \\ N_2N_2\dots N_2\ W_1' \end{matrix} \end{pmatrix}$$

$$\left\{ \begin{matrix} Y \\ Y' \end{matrix} \begin{pmatrix} X_2\,M\,M\ \dots\ M\,W_1 \\ X_2'M'M'\dots M'W_1' \end{pmatrix} + \begin{matrix} W_2M'M'\dots M'X_1 \\ W_2'M\,M\ \dots M\,X_1' \end{matrix} \right\}^*$$

$$\begin{matrix} Y \\ Y' \end{matrix} \begin{pmatrix} W_2N_2N_2\dots N_2 \\ W_2'N_2'N_2'\dots N_2' \end{pmatrix} + \begin{matrix} X_2N_1'N_1'\dots N_1' \\ X_2'N_1N_1\dots N_1 \end{matrix} \end{pmatrix},$$

which depicts the splicing language $L(S)$ from Theorem 1 in Table 1.

Based on G_1, the automaton for the splicing system is constructed using productions in G_1. The relation between productions in G_1 and transition functions, δ for M_1 from Figure 2 is given in Table 3.

Table 3. Productions in G_1 and transition functions for M_1

Production in G_1	Transition Function, δ
$S_0 \rightarrow \begin{matrix} N_1N_1\ \dots N_1X_1 \\ N_1'N_1'\dots N_1'X_1' \end{matrix} S_1$	$\delta\left(q_0,\ \begin{matrix} N_1N_1\ \dots N_1X_1 \\ N_1'N_1'\dots N_1'X_1' \end{matrix}\right) = q_1$
$S_0 \rightarrow \begin{matrix} N_2'N_2'\dots N_2'\,W_1 \\ N_2N_2\dots N_2\ W_1' \end{matrix} S_2$	$\delta\left(q_0,\ \begin{matrix} N_2'N_2'\dots N_2'\,W_1 \\ N_2N_2\dots N_2\ W_1' \end{matrix}\right) = q_2$
$S_1 \rightarrow \begin{matrix} Y \\ Y' \end{matrix} S_3$	$\delta\left(q_1,\ \begin{matrix} Y \\ Y' \end{matrix}\right) = q_3$
$S_2 \rightarrow \begin{matrix} Y \\ Y' \end{matrix} S_3$	$\delta\left(q_2,\ \begin{matrix} Y \\ Y' \end{matrix}\right) = q_3$
$S_3 \rightarrow \begin{matrix} X_2\,M\,M\ \dots\ M\,W_1 \\ X_2'M'M'\dots M'W_1' \end{matrix} S_1$	$\delta\left(q_3,\ \begin{matrix} X_2\,M\,M\ \dots\ M\,W_1 \\ X_2'M'M'\dots M'W_1' \end{matrix}\right) = q_1$
$S_3 \rightarrow \begin{matrix} W_2M'M'\dots M'X_1 \\ W_2'M\,M\ \dots M\,X_1' \end{matrix} S_2$	$\delta\left(q_3,\ \begin{matrix} W_2M'M'\dots M'X_1 \\ W_2'M\,M\ \dots M\,X_1' \end{matrix}\right) = q_2$
$S_3 \rightarrow \begin{matrix} W_2N_2N_2\dots N_2 \\ W_2'N_2'N_2'\dots N_2' \end{matrix}$	$\delta\left(q_3,\ \begin{matrix} W_2N_2N_2\dots N_2 \\ W_2'N_2'N_2'\dots N_2' \end{matrix}\right) = q_4$
$S_3 \rightarrow \left|\begin{matrix} X_2N_1'N_1'\dots N_1' \\ X_2'N_1N_1\dots N_1 \end{matrix}\right.$	$\delta\left(q_3,\ \begin{matrix} X_2N_1'N_1'\dots N_1' \\ X_2'N_1N_1\dots N_1 \end{matrix}\right) = q_5$

Thus, Theorem 5 is proved. □

Automaton for DNA splicing system with one cutting site each of two palindromic restriction enzymes with different crossings is presented as Theorem 6.

Theorem 6

Given

$$S = \begin{pmatrix} \left\{\begin{matrix} A,\ C,\ G,\ T \\ T\ \ G\ \ C\ \ A \end{matrix}\right\}, \begin{matrix} N_1N_1\ \dots N_1X_1\ Y\ X_2M\ \ M\ \dots MW_1\ ZW_2N_2N_2\ \dots N_2 \\ N_1'N_1'\dots N_1'X_1'Y'X_2'M'\ M'\ \dots M'W_1'Z'\,W_2'N_2'N_2'\dots N_2' \end{matrix}, \\ \left\{\begin{pmatrix} X_1\ \ Y\ \ X_2 \\ X_1'\ \ Y'\ \ X_2' \end{pmatrix}, \begin{pmatrix} W_1\ \ Z\ \ W_2 \\ W_1'\ \ Z'\ \ W_2' \end{pmatrix}\right\}, \emptyset \end{pmatrix}$$

is a splicing system involving one cutting site each of two palindromic rules $\begin{matrix} X_1YX_2 \\ X_1'Y'X_2' \end{matrix}$ and $\begin{matrix} W_1\,ZW_2 \\ W_1'Z'\,W_2' \end{matrix}$ with different crossings $\begin{matrix} Y \\ Y' \end{matrix}$ and $\begin{matrix} Z \\ Z' \end{matrix}$, where $\begin{matrix} N_1\ \ X_1\ \ Y\ \ X_2\ \ M\ W_1\ \ Z\ \ W_2 \\ N_1'\ \ X_1'\ \ Y'\ \ X_2'\ \ M'\ W_1'\ \ Z'\ \ W_2' \end{matrix}$ and $\begin{matrix} N_2 \\ N_2' \end{matrix}$ are variables used to denote any arbitrary dsDNA and N_1', X_1', Y', X_2', M', W_1', Z', W_2' and N_2' are complementaries for N_1, X_1, Y, X_2, M, W_1, Z, W_2, and N_2 respectively, $M_2 = (Q, \Sigma, \delta, q_0, F)$ is a deterministic finite automaton for the DNA splicing system that accepts the language $L(S)$, in which $Q = \{q_0, q_1, q_2, q_3, q_4\}$ is the set of states where q_0 is the initial state and $F = \{q_3, q_4\}$ is the set of final states, $\Sigma =$

$$\left\{\begin{matrix} N_1N_1\ \dots N_1X_1\ Y\ X_2 & N_2'N_2'\dots N_2'\,W_1\ ZW_2 & M\ M\ \dots\ M\ W_1\ ZW_2 \\ N_1'N_1'\dots N_1'X_1'Y'X_2', & N_2N_2\ \dots N_2\ W_1'Z'\,W_2', & M'M'\ \dots M'W_1'Z'\,W_2', \\ M'M'\ \dots M'\,X_1YX_2 & N_2N_2\dots N_2 & N_1'N_1'\dots N_1' \\ M\ M\ \dots M\ X_1'Y'X_2', & N_2'N_2'\dots N_2', & N_1N_1\dots N_1 \end{matrix}\right\}$$

is the set of inputs and δ are given by illustrating the automata diagram in Figure 3.
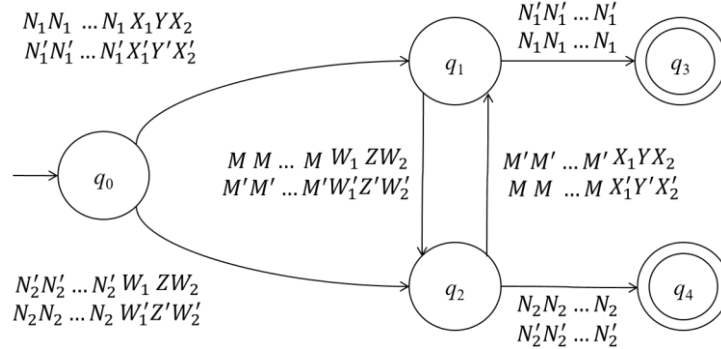

Figure 3. An automaton diagram for M_2

Next, the automaton for DNA splicing system with one cutting site each of two non-palindromic restriction enzymes with the same crossing is presented as Theorem 7.

Theorem 7

Given $S =$

$$\left(\begin{matrix}A, C, G, T\\T\ G\ C\ A\end{matrix}\right\}, \begin{matrix}N_1 N_1 \dots N_1 X_1\ Y\ X_2 M\quad M \dots M W_1\ Y W_2 N_2 N_2 \dots N_2\\N_1' N_1' \dots N_1' X_1' Y' X_2' M'\quad M' \dots M' W_1' Y' W_2' N_2' N_2' \dots N_2',\end{matrix}\right.$$
$$\left.\left\{\begin{pmatrix}X_1\ \ Y\ \ X_2\\X_1',\ Y'',\ X_2'\end{pmatrix}, \begin{pmatrix}W_1\ \ Y\ \ W_2\\W_1',\ Y'',\ W_2'\end{pmatrix}\right\}, \emptyset\right)$$

is a splicing system involving one cutting site each of two

non-palindromic rules $\dfrac{X_1 Y X_2}{X_1' Y' X_2'}$ and $\dfrac{W_1\ Y W_2}{W_1' Y' W_2'}$ with the same

crossing $\dfrac{Y}{Y'}$, where $\dfrac{N_1}{N_1'}, \dfrac{X_1}{X_1'}, \dfrac{Y}{Y'}, \dfrac{X_2}{X_2'}, \dfrac{M}{M'}, \dfrac{W_1}{W_1'}, \dfrac{W_2}{W_2'}$ and $\dfrac{N_2}{N_2'}$ are

variables used to denote any arbitrary dsDNA and $N_1', X_1',$

Y', X_2', M', W_1', W_2' and N_2' are complementaries for $N_1, X_1,$
$Y, X_2, M, W_1, W_2,$ and N_2 respectively, $M_3 = (Q, \Sigma, \delta, q_0, F)$
is a deterministic finite automaton for the DNA splicing
system that accepts the language $L(S)$, in which $Q = \{q_0,$
$q_1, q_2, q_3\}$ is the set of states where q_0 is the initial state

and $F = \{q_3\}$ is the set of final state, $\Sigma = \left\{\begin{matrix}N_1 N_1 \dots N_1 X_1\\N_1' N_1' \dots N_1' X_1',\end{matrix}\right.$

$\left.\begin{matrix}X_2\ M\ M \dots\ M\ W_1\ \ W_2 N_2 N_2 \dots N_2\ \ Y\\X_2' M' M' \dots M' W_1'\ \ W_2' N_2' N_2' \dots N_2',\ Y'\end{matrix}\right\}$ is the set of inputs and

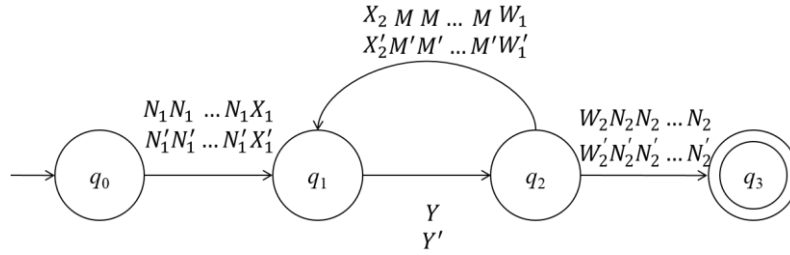δ are given by illustrating the automata diagram in Figure
4.

Figure 4. An automaton diagram for M_3

Lastly, the automaton for DNA splicing system with
one cutting site each of two non-palindromic restriction
enzymes with different crossings is presented as Theorem
8.

Theorem 8

Given $S = \left(\begin{matrix}A, C, G, T\\T\ G\ C\ A\end{matrix}\right\}, \begin{matrix}N_1 N_1 \dots N_1 X_1\ Y\ X_2 M\\N_1' N_1' \dots N_1' X_1' Y' X_2' M'\end{matrix}\right.$

$\begin{matrix}M \dots M W_1\ Z W_2 N_2 N_2 \dots N_2\\M' \dots M' W_1' Z' W_2' N_2' N_2' \dots N_2',\end{matrix} \left\{\begin{pmatrix}X_1\ \ Y\ \ X_2\\X_1',\ Y'',\ X_2'\end{pmatrix}, \begin{pmatrix}W_1\ \ Z\\W_1',\ Z'',\end{pmatrix}\right.$

$\left.\left.\begin{matrix}W_2\\W_2'\end{matrix}\right)\right\}, \emptyset\right)$ is a splicing system involving one cutting site

each of two palindromic rules $\dfrac{X_1 Y X_2}{X_1' Y' X_2'}$ and $\dfrac{W_1\ Z W_2}{W_1' Z' W_2'}$ with

different crossings $\dfrac{Y}{Y'}$ and $\dfrac{Z}{Z'}$ where $\dfrac{N_1}{N_1'}, \dfrac{X_1}{X_1'}, \dfrac{Y}{Y'}, \dfrac{X_2}{X_2'}, \dfrac{M}{M'}, \dfrac{W_1}{W_1'},$

$\dfrac{Z}{Z''}, \dfrac{W_2}{W_2'}$ and $\dfrac{N_2}{N_2'}$ are variables used to denote any arbitrary

dsDNA and $N_1', X_1', Y', X_2', M', W_1', Z', W_2'$ and N_2' are
complementaries for $N_1, X_1, Y, X_2, M, W_1, Z, W_2,$ and N_2
respectively, $M_4 = (Q, \Sigma, \delta, q_0, F)$ is a deterministic finite
automaton for the DNA splicing system that accepts the
language $L(S)$, in which $Q = \{q_0, q_1, q_2, q_3, q_4, q_5\}$ is the set
of states where q_0 is the initial state and $F = \{q_5\}$ is the set

of final state, $\Sigma = \left\{\begin{matrix}N_1 N_1 \dots N_1\ \ X_1 Y X_2\ \ W_1\ Z W_2\\N_1' N_1' \dots N_1'\ X_1' Y' X_2'\ W_1' Z' W_2',\end{matrix}\right.$

$\left.\begin{matrix}M\ M \dots\ M\ \ N_2 N_2 \dots N_2\\M' M' \dots M''\ N_2' N_2' \dots N_2'\end{matrix}\right\}$ is the set of inputs and δ are given

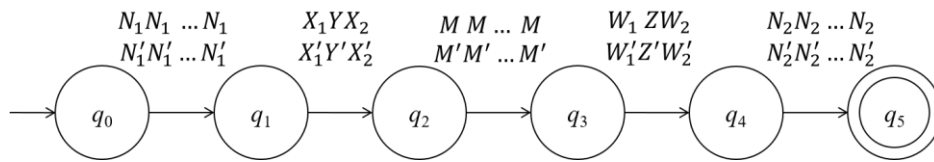by illustrating the automata diagram in Figure 5.

Figure 5. An automaton diagram for M_4

Theorems 6, 7 and 8 are proved in same manner as
Theorem 5 using grammars. Hence, the splicing
languages from the DNA splicing system involving two
palindromic and non-palindromic rules with the same
and different crossings are visualised as automata
diagrams.

V. CONCLUSION

In this research, the relation of DNA splicing systems and
automata theory is discussed. The automata diagrams for
the DNA splicing systems with one cutting site each of two
palindromic or two non-palindromic rules for the same

and different crossings of the restriction enzymes are modelled using deterministic finite automata. The languages resulting from the automata depict the languages generated by the grammars.

VII.   REFERENCES

Brejová, B., DiMarco, C., Vinar, T., Hidalgo, SR, Holguin, G. & Patten, C. (2000). Finding Patterns in Biological Sequences. Unpublished project report for CS798G, University of Waterloo, Fall.

Fong, WH, Ismail, NI & Sarmin, NH. (2019). The Modelling of DNA Splicing Systems with Two Non-Palindromic Restriction Enzymes. Indian Journal of Public Health Research & Development, In press.

Head, T. (1987). Formal Language Theory and DNA: An Analysis of the Generative Capacity of Specific Recombinant Behaviors. Bulletin of Mathematical Biology, 49, no. 6, 737-759.

Head, T. (1998). Splicing Representations of Strictly Locally Testable Languages. Discrete Applied Mathematics, 87, no. 1, 139-147.

Ismail, NI, Fong, WH & Sarmin, NH. (2019). Generalisations of Splicing Languages in DNA Splicing Systems Involving Two Palindromic Restriction Enzymes. Malaysian Journal of Fundamental and Applied Sciences, In press.

Linz, P. (2006). An Introduction to Formal Languages and Automata. 4th ed, Jones and Bartlett Publisher, USA.

New England Biolabs Inc. (2017). NEB 2017-18 Catalog & Technical Reference, Ipswich, United States, Catalogue.

Paun, G., Rozenberg, G. & Salomaa, A. (1998). DNA Computing: New Computing Paradigms, Springer -Verlag Berlin Heidelberg, Germany.

Tomohiro, I., Inenaga, S. & Takeda, M. (2013). Palindrome Pattern Matching. Theoretical Computer Science, 483, 162-170.

Yusof, Y. (2012). DNA Splicing System Inspired by Bio Molecular Operations. Ph.D. Thesis, Universiti Teknologi Malaysia.