Malaysian Journal of Mathematical Sciences 19(1): 325–342 (2025)
https://doi.org/10.47836/mjms.19.1.17

Malaysian Journal of Mathematical Sciences

Journal homepage: https://mjms.upm.edu.my

A Study on Mathematical Model of n-th Order Limit Language from a Biological Perspective via Wet-Lab Experiment

Mohd Khairuddin, S. H.* [1], Ahmad, M. A. [1], Ramli, A. N. M. [2], Sarmin, N. H. [3], and Fong, W. H. [3]

[1]*Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300 Kuantan, Pahang, Malaysia*
[2]*Faculty of Industrial Sciences and Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300 Kuantan, Pahang, Malaysia*
[3]*Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia*

E-mail: sitihajarmohdkhairuddin96@gmail.com
**Corresponding author*

Received: 1 April 2024
Accepted: 23 October 2024

Abstract

The analysis of double-stranded DNA (dsDNA) recombinant behaviour led to the mathematical modelling of the deoxyribonucleic acid (DNA) splicing system. This multidisciplinary study is based on the fundamentals of formal language theory and informational macromolecules. The splicing system's number of rules previously specified the n-th order limit language. A previous experiment in the lab established the existence of a second-order limit language. Nevertheless, based on the quantity of rules employed in the splicing system, no laboratory experiments have been carried out to verify the existence of the n-th order limit language. This paper presents experimental evidence supporting the theoretical results of n-th order limit languages, which were previously proven using double induction. Laboratory experiments involving DNA digestion and ligation were conducted to validate these theoretical findings. This investigation has led to the validation of the model, indicating that research on third and fourth order limit language supports the notion of n-th order limit language biologically. Furthermore, it shows that the mathematical model of the n-th order limit language was empirically confirmed if the dsDNA molecules generated in the experiment match those anticipated by the model.This research advances the understanding of DNA splicing systems by empirically validating the n-th order limit language, bridging formal language theory and molecular biology, and paving the way for future studies and technological applications in DNA computing, synthetic biology, and bioinformatics.

Keywords: splicing system; formal language theory; splicing language; limit language; n-th limit language.

1 Introduction

Deoxyribonucleic acid (DNA) is a polymer made up of monomers, an essential cell found within every living thing [3]. DNA comprises thousands of nucleotides pair formed by a phosphate, sugar, and base. The base, which is composed of four separate chemical bases, is where the information in DNA is stored, given by adenine (A), guanine (G), thymine (T), as well as cytosine (C). Moreover, guanine with cytosine (C – G), adenine with thymine (A – T), and vice versa are the only conceivable pairings based on Watson-Crick complementarity [16]. It is important to note that restrictase, restriction endonuclease, or restriction enzyme are the names for the enzyme that fragments DNA at the recognition site. Following from there, restriction enzymes identify a specific sequence where each has a specific target site, which aids in binding to molecules and splicing them [2].

Head [5] was the first to use a formal presentation to demonstrate the recombination of DNA molecules in 1987. As time passed, numerous researchers conducted research on splicing systems, resulting in the development of enhanced or extended splicing systems, for instance, Pixton [6], Paun [15], Fuzzy splicing systems [8], and Yusof-Goode (Y – G) [17]. In order to model the splicing system biologically, dsDNA is cut and pasted in the occurrence of restriction enzymes as well as ligase in a test tube to create new hybrid DNAs called splicing language [7].

Splicing language results from analyzing DNA molecules after the splicing process through formal language theory and is divided into three categories: transient, limit, and inert/adult languages [7]. This research focuses on the limit language, initially proposed by Goode [7] after studying DNA behavior in the final stages of splicing. Several studies have explored limit languages. First, Goode and Pixton investigated the n-th order limit language, suggesting that a new language can be formed by removing transient words from the previous order limit language [7]. In 2022, a subsequent study refined this concept, establishing that the rules governing the splicing system determine the order of the limit language [12]. Further research applied automata theory to DNA splicing, transforming the limit language into a transition graph [12]. The n-th order limit language was derived from a grammar represented as an automaton, with transition graphs illustrating the language of transition labels corresponding to DNA molecules produced by the splicing system. Recently, the factors limiting the formation of the n-th order limit language were identified, including unequal rule lengths and the repeated application of the same rules to multiple crossing sites of the initial strings [13].

The integration of mathematics and biology has gained attention, particularly through mathematical models and simulations to understand complex biological systems. Studies like [4] have used statistical analysis to combine these fields. However, the rise of statistical packages that generate instant results has sometimes led to less meaningful conclusions. As mathematical modeling becomes more common, simulations, such as those by [9], assess the impact of environmental stressors like temperature and toxicants on ecosystems. This study aims to show how modeling and simulations can effectively study biological and ecological systems.

However, in this study few experiments were conducted to validate the existence of different splicing languages. As a result, it was proven that the limit language did exist through validation performed by Goode in [10]. In general, the experiment is carried out in one or two stages. Moreover, in one stage experiment, one or two restriction enzymes were used simultaneously by the researcher. In contrast, just one restriction enzyme is utilised at a given moment in the two-stage experiment [14]. The process is then continued by adding another enzyme. Previously, most experiments were performed in one step with one or two restriction enzymes being utilised simultaneously, called digestion and double digestion, respectively. However, multiple digestions

are used in this research because there are three and four restriction enzymes used in the wet lab experiments.

Formerly, the n-th order limit language's improvised version was examined theoretically throughout the research [12]. Rules utilised in the splicing system, according to the researchers, reveals the order of the limit language. Additionally, it was mentioned that the combination of the string generated by the splicing language would depend on the number of rules utilised in the system. The restriction enzymes involved in the splicing process are a number of regulations affecting the splicing system.

The occurrence of n-th order limit language in this study is validated from the biological point of view through laboratory experiment. Section 2 provides all preliminaries, which are useful for analysing the results. After that, Section 3 presents mathematical or dry models of splicing systems that produce third- and fourth-order limit language. In addition, Section 4 gives the wet model of the n-th order limit language, which explains how initial strings of dsDNA represent a set of initial strings and the type of restriction enzymes that represent a set of rules are selected. Subsequently, Section 5 describes all procedures involved in the laboratory experiment to validate the results produced in Section 3 and 4.

In order to generalise the n-th order limit language, two models that produce third- and fourth-order limit language are generated. Second-order limit language is not selected since it has been validated through laboratory experiments by Ahmad et al. [1]. By validating these models, it is sufficient to prove the occurrence of the n-th order limit language.

The key definitions employed in this research are given in the following section.

2 Preliminaries

In this section, splicing language concepts [11] are explained. The formal definition also mentions the Head Splicing System [5].

Definition 2.1. *Alphabet, A,* [11]
A denotes an alphabet, a finite non-empty set of symbols.

Definition 2.2. *String,* [11]
A string represents a finite sequence of symbols deriving out of the alphabet.

Definition 2.3. *Language, L,* [11]
Language symbolises a set of strings selected from A^* in which A acts as a particular alphabet.

The Head splicing system is used throughout this research due to its straightforward notation for the rules used in the splicing process, making it the simplest splicing system. Previously, it was carefully developed using a biological example involving the molecular cut-and-paste process of DNA digestion and ligation. Additionally, this system is particularly suitable for our research because it is limited to finite cases. This means it can handle many initial strings and many strings used in the splicing system, reflecting wet lab experiments involving multiple enzymes and the PCR process to form many strands of DNA. Inspired by the processes of DNA digestion and ligation, the Head splicing system's mathematical model was verified through experiments involving multiple digestion and ligation. Then, the definition of Head splicing system is given as follows,

Definition 2.4. *Head Splicing System* [5]
The spicing system is made up of four unique groups of elements such as A, I, B as well as C, which are explained below:

- A *represents a set of alphabet.*
- I *denotes a set of initial strings.*
- B *symbolises a set of rules, which represents $5'-$overhang or blunt end.*
- C *denotes a set of rules, which represents $3'-$overhang.*

Next, the definition of the limit language [7] and n-th order limit language are elaborated.

Definition 2.5. *Limit Language,* [10]
Limit language, often known as as a first-order limit language, is a splicing language produced by the molecules that remain once the splicing system is completed or has attained equilibrium.

Definition 2.6. *n-th order limit language from rule viewpoints,* [12]
The order of the limit language is specified by the rules used in the splicing system, according to the original definition as innovated by Khairuddin et al. [12]*. The previous order of the limit language is different from the present order limit language, as per Goode's definition* [6] *in of the n-th order limit language.*

We denote the splicing language generated by a splicing system S as $L_n(S)$. Now, let's introduce the concept of the n-th order limit language, $L_n(S)$. In $L_n(S)$, n indicates the order of the language. Here is a key point: the initial strings in S are of the form cxd, where c and d represent the left and right contexts, respectively. The variable represents the crossing site, where splicing rules are applied. The $L_n(S)$ depends on the number of distinct splicing rules acting on each crossing site x. Importantly, these rules must be unique (no duplicates) and possess the same length for the crossing site portion of the rule. Note that a splicing language is referred to as $L_n(S)$, if the set of string generated in $L_n(S)$ is distinct from the set of strings of $L_1(S), L_2(S), \ldots, L_n(S)$ given by $\bigcap_{n=1}^{n} L_n(S) = \emptyset$ and $L_1(S) \not\subset L_2(S) \not\subset \ldots \not\subset L_n(S)$.

3 Mathematical Model of n-th order limit language

Firstly, a theorem on the formation of n-th order limit language has been discussed in [12].

Theorem 3.1. *If a splicing system contains x number of initial strings and y number of rules where $x, y \in \mathbb{Z}^+$, then the splicing system generates the n-th order limit language.*

The theorem above leads to the following summarisation in Table 1.

Table 1: Summarisation of cases.

Initial string(s)	Rule(s)	Order of limit language
1	1	1^{st}
1	2	2^{nd}
1	3	3^{rd}
1	y	n^{th}
2	y	n^{th}
x	y	n^{th}

The findings then lead to the following lemmas,

Lemma 3.1. *If $y \geq 1$ rule is used in the splicing system, then $2(y-1)$ different combination of the string in the language is produced.*

Table 2 shows $2(y-1)$ number of the different combination of the string in the language is produced based on the string and rule involve in the splicing system.

Table 2: Combination of the strings in the language.

Initial string(s)	Rule(s)	Combination of the string
1	1	0
1	2	2
1	3	4
2	1	0
2	2	2
2	3	4
3	1	0
3	2	2
3	3	4

Then, double induction method is used to proof the lemma where $f(x, y)$ is a function that presents the number of combinations of the strings that consist of x strings and y rules of the splicing system.

By the principle of double induction, the ensuing prerequisite has been validated. Double induction is a valuable mathematical proof technique, particularly suited for statements involving two interconnected variables. In this context, the variables x and y are interdependent, making double induction an appropriate method for proving the theorem:

1. $f(x, y)$ is true for $x = 1$ or $y = 1$.
2. For all $y \geq 1$, if $f(x, k)$ for some $k \leq y$ is true, then $f(x, k + 1)$ is true.
3. For all $x \geq 1$, if $f(k, y)$ is true for all $y \geq 1$, and some $k \leq x$, then $f(k + 1), p$ is true for all $y \geq 1$.

Then, $f(x, y)$ is true for all $x \geq 1$ and $y \geq 1$.

Lemma 3.2. *If x number of initial strings are used in the splicing system, then $x(2x+1)$ pattern of strings of the language is obtained.*

Table 3 shows number of the pattern of strings of the language is obtained based on the string and rule involve in the splicing system.

Table 3: Pattern of the strings in the language.

Initial string(s)	Rule(s)	Pattern of the string
1	1	
1	2	3
1	3	
2	1	
2	2	10
2	3	
3	1	
3	2	21
3	3	

Then, double induction method, again, is used to proof the lemma where $g(x, y)$ is a function that presents the number of the pattern of the strings that consist of x strings and y rules of the splicing system.

By the principle of double induction, the ensuing prerequisite has been validated:

1. $g(x, y)$ is true for $x = 1$ or $y = 1$.
2. For all $y \geq 1$, if $g(x, k)$ for some $k \leq y$ is true, then $g(x, k + 1)$ is true.
3. For all $x \geq 1$, if $g(k, y)$ is true for all $y \geq 1$, and some $k \leq m$, then $g(k + 1), y$ is true for all $y \geq 1$.

Then, $g(x, y)$ is true for all $x \geq 1$ and $y \geq 1$.

Now, a mathematical model that produces n-th order limit language is developed where the result will be based on the aforementioned theorem and lemmas. Let,

$$S = \left(\{a, c, g, t\}, \{\mu wxzzwxxy \ldots \gamma\}, \{(w, xz, z), (w, xx, y), \ldots, (e_p, x_p, f_p)\}, \emptyset\right),$$

in which w and z, as well as x and y, accomplish one another while $\mu, \gamma, w, x, y, z \in A^*$. Moreover, the initial string is explained below.

$$5'-\mu wxzzwxxywxyz \ldots \gamma - 3',$$
$$3'-\mu' zywwzywxzyxw \ldots \gamma' - 5'.$$

After the splicing process, the splicing language is formed as below:

$$L(S) = \left\{ \begin{array}{c} \mu wxyz\gamma, \mu wwyz\mu', \gamma' wxyz\gamma, \mu wxzzwxyz\gamma, \mu wxzzwwyz\mu', \\ \gamma' wxzzwxyz\gamma, \mu wxzzwxxywxyz\gamma, \mu wxzzwxxywwyz\mu', \\ \gamma' wxzzwxxywxyz\gamma, \mu wxzzwxxywxyzxxyz\gamma, \\ \mu wxzzwxxywxyzxwyz\mu', \gamma' wxzzwxxywxyzxxyz\gamma, \\ \mu wxzzwxxywxyzxyyzwxyz\gamma, \mu wxzzwxxywxyzxyyzwwyz\mu', \\ \gamma' wxzzwxxywxyzxyyzwxyz\gamma, \dots \end{array} \right\},$$

where $\mu', \gamma' \in A^*$.

Provided that the order is determined via the rules used in splicing system is shown below,

$$L_n(S) = \left\{ \begin{array}{c} \mu w \left(xzzw \cup xxyw \cup xyzx \cup yyzw \cup xyz \dots \cup \dots\right) * \gamma, \\ \mu w \left(xzzw \cup xxyw \cup xyzx \cup yyzw \cup wyz \dots \cup \dots\right) * \mu', \\ \gamma' w \left(xzzw \cup xxyw \cup xyzx \cup yyzw \cup xyz \dots \cup \dots\right) * \gamma \end{array} \right\}.$$

Based on the above example, the combination of the language is the amount of strings for example $xzzw, xxyyw, \dots$ in $(\dots)$ and the number of pattern of the string in the language is the pattern $\mu \dots \gamma, \mu \dots \mu', \gamma' \dots \gamma$. The following section focuses on the wet model of third- and fourth-order limit language, which is modelled as per the n-th order limit language's mathematical model above.

4 Wet Model of n-th order limit language

In this study, two wet-lab experiments are conducted to validate the third- and fourth-order limit language, respectively.

Two types of models exist. Model 1 is a splicing system that produces third-order limit language, whereas Model 2 is a splicing system that produces fourth-order limit language. In order to validate the n-th order limit language, these two models are developed.

Model 1 uses three restriction enzymes, which are *Msp*I, *Aci*I and *Mse*I. The splicing language is produced according to the given rule $B = \big(\{c, cg, g\}, \{c, cg, c\}, \{t, ta, a\}\big)$. Here, an initial strand of dsDNA with three crossing sites and three rules is chosen for the third-order limit language to be produced. Subsequently, the third-order limit language is illustrated in the following general form.

Let,

$$\begin{aligned} S &= (A, I, B, C) \text{ consisting } A = \{a, c, g, t\} \text{ is set of alphabets,} \\ I &= \{\mu ccgg\eta ccgc\sigma ttaa\gamma\}, \\ B &= \big(\{c, cg, g\}, \{c, cg, c\}, \{t, ta, a\}\big), \\ C &= \{\oslash\}, \text{ where } \mu, \gamma, \eta, \mu', \gamma', \eta' \in A^*. \end{aligned}$$

The string is divided into four parts since the rules which act on the crossing site of the initial strings are $ccgg$, $ccgc$ and $ttaa$. Thus, there are eight different parts: A, B, C, D, E, F, G as well as H. All generated strings, as well as the initial string, are given below in Table 4.

Table 4: All possible generated string for Model 1.

No.	String	Possible combination of string
1.	$\mu \dots \gamma$	$A + (B \cup C \cup F \cup G)^* + D$
2.	$\mu \dots \mu'$	$A + (B \cup C \cup F \cup G)^* + E$
3.	$\gamma' \dots \gamma$	$H + (B \cup C \cup F \cup G)^* + D$

For the symbol $(\cup)$, the part of the string can occur alternately as either B, C, F, G in the string, while the symbol $(^*)$ in the string means it can occur recursively. In Table 5, the resulted strings of the DNA Splicing System are listed. They are the generated strings where the same molecules are eliminated.

Table 5: Output strings generated by DNA splicing system (Model 1).

No.	String	General form
1.	$\mu \dots \gamma$	$\mu c\big(cgg\eta c \cup cgc\sigma a \cup tta\sigma' g \cup cgg\eta' c\big)^* taa\gamma$
2.	$\mu \dots \mu'$	$\mu c\big(cgg\eta c \cup cgc\sigma a \cup tta\sigma' g \cup cgg\eta' c\big)^* cgg\mu'$
3.	$\gamma' \dots \gamma$	$\gamma' a\big(cgg\eta c \cup cgc\gamma a \cup tta\sigma' g \cup cgg\eta' c\big)^* taa\gamma$

Hence, the third-order limit language produced from the splicing language above is given by,

$$L_3(S) = \left\{ \begin{matrix} \mu c\big(cgg\eta c \cup cgc\sigma a \cup tta\sigma' g \cup cgg\eta' c\big)^* taa\gamma, \\ \mu c\big(cgg\eta c \cup cgc\sigma a \cup tta\sigma' g \cup cgg\eta' c\big)^* cgg\mu', \\ \gamma' a\big(cgg\eta c \cup cgc\gamma a \cup tta\sigma' g \cup cgg\eta' c\big)^* taa\gamma \end{matrix} \right\}.$$

Next, Model 2 uses four restriction enzymes, which are *Age*I, *Eag*I, *BspE*I and *AvrI*I. The splicing language is created in reference to the given rule,

$$B = \big(\{g, gtac, c\}, \{a, ccgg, t\}, \{c, ggcc, g\}, \{a, agtc, t\}\big).$$

Here, an initial strand of dsDNA with four crossing sites and four rules is chosen for the fourth-order limit language to be produced. Then, the fourth-order limit language is presented in the general form.

Assume $S = (A, I, B, C)$ comprising:

$$\begin{aligned} A &= \{a, t, c, g\} \text{ as a set of alphabets,} \\ I &= \{\mu ggtacc\eta accggt\sigma cggccg\varphi aagctt\gamma\} \text{ represents a set of initial strings,} \\ B &= (\{g, gtac, c\}, \{a, ccgg, t\}, \{c, ggcc, g\}, \{a, agtc, t\}) \text{ is a set of rules, and} \\ C &= \{\oslash\}, \text{ where } \mu, \gamma, \eta, \varphi, \mu', \gamma', \eta', \varphi' \in A^*. \end{aligned}$$

The string is divided into five parts since the rules which act on the crossing site of the initial strings are $ggtacc$, $accggt$, $cggccg$ and $aagctt$. Thus, there are ten different element: A, B, C, D, E, F, G, H, I, J. All generated strings, including the initial string, are given in the following Table 6.

Table 6: All possible generated string for Model 2.

No.	String	Possible combination of string
1.	$\mu \dots \gamma$	$A + (B \cup C \cup D \cup J \cup H \cup I)^* + E$
2.	$\mu \dots \mu'$	$A + (B \cup C \cup D \cup J \cup H \cup I)^* + F$
3.	$\gamma' \dots \gamma$	$J + (B \cup C \cup D \cup J \cup H \cup I)^* + E$

In Table 7, the resulted strings of the DNA Splicing System are listed. They are the generated strings where the same molecules are eliminated.

Table 7: Output strings generated by DNA splicing system (Model 2).

No.	String	General form
1.	$\mu \dots \gamma$	$\mu a(ccggt\eta c \cup ggccg\sigma t \cup ccgga\varphi c \cup ggccg\eta' a \cup ccgga\sigma' c \cup ctagg\varphi' t)^* ctagg\gamma$
2.	$\mu \dots \mu'$	$\mu g(ccggt\eta c \cup ggccg\sigma t \cup ccgga\varphi c \cup ggccg\eta' a \cup ccgga\sigma' c \cup ctagg\varphi' t)^* gtacc\mu'$
3.	$\gamma' \dots \gamma$	$\gamma' a(ccggt\eta c \cup ggccg\sigma t \cup ccgga\varphi c \cup ggccg\eta' a \cup ccgga\sigma' c \cup ctagg\varphi' t)^* agctt\gamma$

Hence, the fourth-order limit language is produced from the splicing language above.

$$L_4(S) = \begin{Bmatrix} \gamma' a\big(gtacc\eta a \cup ccggt\sigma g \cup ccggt\sigma' g \cup ggccg\eta' a \cup ggccg\varphi a \cup agctt\varphi' c\big) * agctt\gamma, \\ \mu g\big(gtacc\eta a \cup ccggt\sigma g \cup ccggt\sigma' g \cup ggccg\eta' a \cup ggccg\varphi a \cup agctt\varphi' c\big) * gtacc\mu', \\ \gamma' a\big(gtacc\eta a \cup ccggt\sigma g \cup ccggt\sigma' g \cup ggccg\eta' a \cup ggccg\varphi a \cup agctt\varphi' c\big) * agctt\gamma, \end{Bmatrix}.$$

The methods involved in the laboratory experiment are described in more detail in the next section.

5 Methodology

The initial strands of dsDNA in this experiment are taken from Enterobacteria lambda phage DNA (λDNA) acquired from New England Biolabs MD, USA [Research Biolabs Sdn. Bhd. New England Biolabs 2022-23 Catalog and Technical Reference. USA: Catalogue. 2022/23]

Model 1 of the λDNA identifies a region of interest compared to the initial string (I). This region includes cutting sites for restriction enzymes *Msp*I, *Aci*I, and *Mse*I. Similarly, Model 2 focuses on a different region of interest within the λDNA compared to the initial string (I). This region contains recognition sites for restriction enzymes *Age*I, *Eag*I, *BspE*I, and *Avr*II. The specific lengths of these fragments for each model will be detailed below in Table 8

Table 8: Length of each fragment for Models 1 and 2.

Model	Initial Strand	Length of each Fragment
1	$A-Msp\text{I}-B-Aci\text{I}-C-Mse\text{I}-D$	$\lvert A\rvert = 32\text{bp}$, $\lvert Msp\text{I}\rvert = 4\text{bp}$, $\lvert B\rvert = 247\text{bp}$, $\lvert Aci\text{I}\rvert = 4\text{bp}$, $\lvert C\rvert = 167\text{bp}$, $\lvert Mse\text{I}\rvert = 4\text{bp}$, $\lvert D\rvert = 12\text{bp}$
c	$F-Age\text{I}-G-Eag\text{I}-H-BspE\text{I}-I-AvrII-J$	$\lvert F\rvert = 13\text{bp}$, $\lvert Age\text{I}\rvert = 6\text{bp}$, $\lvert G\rvert = 244\text{bp}$, $\lvert Eag\text{I}\rvert = 6\text{bp}$, $\lvert H\rvert = 2393\text{bp}$, $\lvert BspE\text{I}\rvert = 6\text{bp}$, $\lvert I\rvert = 1972\text{bp}$, $\lvert AvrII\rvert = 6\text{bp}$, $\lvert J\rvert = 33\text{bp}$

The length of $(A-Msp\text{I}-B-Aci\text{I}-C-Mse\text{I}-D)$ is 610 bp. This strand has exactly three cutting sites of restriction enzyme *Msp*I, *Aci*I and *Mse*I, respectively. Therefore, the initial string is known as $A-B-C-D$. Moreover, the length of $(F-Age\text{I}-G-Eag\text{I}-HBspE\text{I}-I-AvrII-J)$ is 4760 bp. This strand has a precise four cutting sites of restriction enzyme *Age*I, *Eag*I, *BspE*I, and *Avr*II, respectively. Therefore, the initial string is denoted as $F-G-H-I-J$.

Both strands are produced by PCR via Bio-Rad MyCycler™ Thermal Cycler according to the following recipe. The PCR process is conducted for both strands separately. The PCR can form thousands of DNA strands' copies. OneTaq® Hot Start 2X Master Mix alongside standard buffer is utilised for $A-B-C-D$ and $F-G-H-I-J$ strands in Table 9.

Table 9: Solution for PCR aliquote.

Solution	Volume
λDNA (10 ng/μL)	4 μL
Forward primer (10μM)	1 μL
Reverse primer (10μM)	1 μL
OneTaq® hot start 2X master mix	25 μL
Nuclease free water	19 μL
Total volume	50 μL

The reverse and forward primers as below were selected depending on the kind of DNA polymerase and PCR concentration. The melting point of each primer, as well as the annealing temperature that works for both forward and reverse primers, may be calculated using the Tm calculator on the NEB website. Model 1 has an annealing temperature of 51°C in Table 10, whereas Model 2 has a temperature of 49°C in Table 11.

Table 10: Forward and reverse primers of $A-B-C-D$ strand.

No.	Primer	% GC	Tm (°C)	Annealing Temperature (°C)
	Forward Primer: 18bp			
1.	$5'-$ ACCCTTCGTCCGTCTTTC $-3'$	56	56	51
	$3'-$ TGGGAAGCAGGCAGAAAG $-5'$			
	Reverse Primer: 15bp			
2.	$5'-$ GGCTGACCATCCGGA $-3'$	67	56	
	$3'-$ CCGACTGGTAGGCCT $-5'$			

Table 11: Forward and reverse primers of $F-G-H-I-J$ strand.

No.	Primer	% GC	Tm (°C)	Annealing Temperature (°C)
	Forward Primer: 13bp			
1.	$5'-$ CCACCGGTTCCGG $-3'$	77	55	49
	$3'-$ GGTGGCCAAGGCC $-5'$			
	Reverse Primer: 20bp			
2.	$5'-$ TAACCAATTCCTAGGCAGGT $-3'$	45	54	
	$3'-$ ATTGGTTAAGGATCCGTCCA $-5'$			

The instrumental settings for the PCR for both models are shown in the Tables 12 and 13 where initial denaturation, denaturation, annealing, extension, and ultimate extension in Steps 1, 2, 3, 4, and 5, respectively.

Table 12: Conditions for thermocycling a standard PCR for Model 1.

Step	Temperature(°C)	Time	Number of Cycle
1	94	30s	1
2	94	30s	35
3	58	40s	35
4	72	30s	35
5	72	30s	1

Table 13: Conditions for thermocycling a standard PCR for Model 2.

Step	**Temperature**(°C)	**Time**	**Number of Cycle**
1	94	30s	1
2	94	30s	30
3	58	40s	30
4	72	30s	30
5	72	5 mins	1

The preparation for digestion and ligation follows the technique for DNA digestion, with three restriction enzymes for Model 1 and four restriction enzymes for Model 2. First, all the components are added to a clean tube in the order shown in Tables 14 and 15.

Table 14: Digestion and ligation recipe for Model 1.

Without ligase	**With ligase**
10 μL DNA	10 μL DNA
2 μL buffer of each enzymes	2 μL buffer of each enzymes
1 μL *Msp*I	1 μL *Msp*I
1 μL *Aci*I	1 μL *Aci*I
1 μL *Mse*I	1 μL *Mse*I
11 μL sterile water	4 μL sterile water
	2 μL T4 ligase
	5 μL buffer for T4 ligase

Table 15: Digestion and ligation recipe for Model 2.

Without ligase	**With ligase**
10 μL DNA	10 μL DNA
2 μL buffer of each enzymes	2 μL buffer of each enzymes
1 μL *Age*I	1 μL *Age*I
1 μL *Eag*I	1 μL *Eag*I
1 μL *BspE*I	1 μL *BspE*I
1 μL *Avr*I	1 μL *Avr*I
18 μL sterile water	11 μL sterile water
	2 μL T4 ligase
	5 μL buffer for T4 ligase

Due to the repeated digestions, the samples (aliquots) required a longer incubation time (2 hours) at $37°C$. After this extended incubation, the enzymatic activity was ceased by heat treatment ($65°C$ for 15 minutes). This inactivated enzymes and halted the digestion process. The digested DNA was then ready for gel electrophoresis. To further halt any remaining enzymatic activity and prepare for gel loading, all samples were frozen at $-20°C$. Following electrophoresis,

the gel was visualized using a UV transilluminator to reveal the DNA fragment sizes. The predicted outcomes for both models are presented in the following Tables 16 and 17 and Figures 1 and 2.

Table 16: The Fragment Size (bp) of Predicted Molecules for Model 1.

No.	**Fragment**	**Size** (bp)	**No.**	**Fragment**	**Size** (bp)
1.	A	63	9.	A−B−D	379
2.	B	251	10.	A−C−D	359
3.	C	231	11.	A−B−A′	377
4.	D	65	12.	A−C−A′	357
5.	A−B−C−D	610	13.	D′−B−D	361
6.	A−D	379	14.	D′−C−D	381
7.	A−A′	126	15.	A−B−C−A′	608
8.	D′−D	130	16.	D′−B−C−D	612

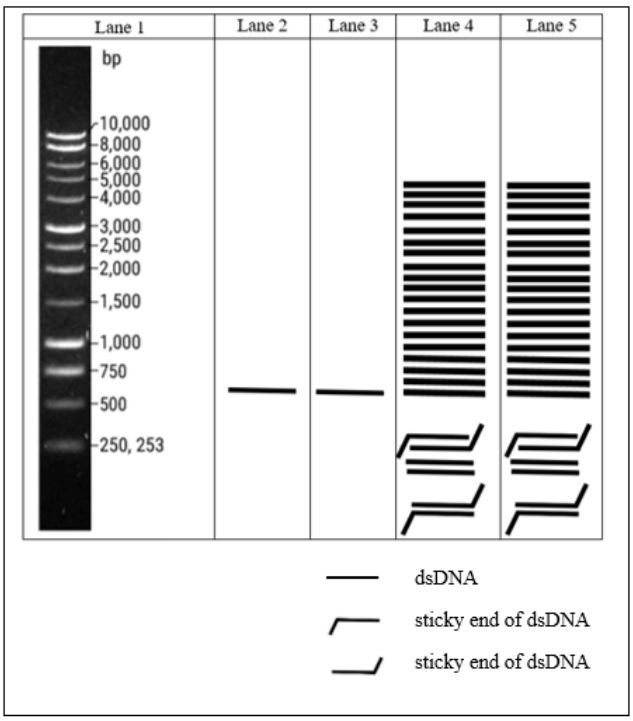

Figure 1: Predicted gel of digestion and ligation towards strand for Model 1.

The following explains what has been filled in each lane that appears in the Figure 1. Lane 1: 1 kb ladder; Lane 2: PCR-processed DNA; Lane 3: Pre-ligated DNA; Lane 4: Post-ligation processed DNA; Lane 5: Overnight incubation of DNA at 4°C.

Table 17: The fragment size (bp) of predicted molecules for Model 2.

No.	**Fragment**	**Size** (bp)	**No.**	**Fragment**	**Size** (bp)	**No.**	**Fragment**	**Size** (bp)
1.	F	64	11.	F–H–J	2531	21.	F–H–I–J	4510
2.	G	250	12.	F–I–J	2111	22.	F–G–H–F′	2785
3.	H	2399	13.	F–G–F′	378	23.	F–G–I–F′	2357
4.	I	1979	14.	F–H–F′	2527	24.	F– H–I–F′	4514
5.	J	68	15.	F–I–F′	2107	25.	J′–G–H–J	2785
6.	F–G–H–I–J	4760	16.	J′–G–J	386	26.	J′–G–I –J	2365
7.	F–J	132	17.	J′–H–J	2535	27.	J′–H–I–J	4514
8.	F–F′	128	18.	J′–I–J	2115	28.	F– G–H–I–F′	4756
9.	J′–J	136	19.	F–G–H–J	2781	29.	J′–G–H–I–J	4764
10.	F–G–J	382	20.	F–G–I–J	2361			

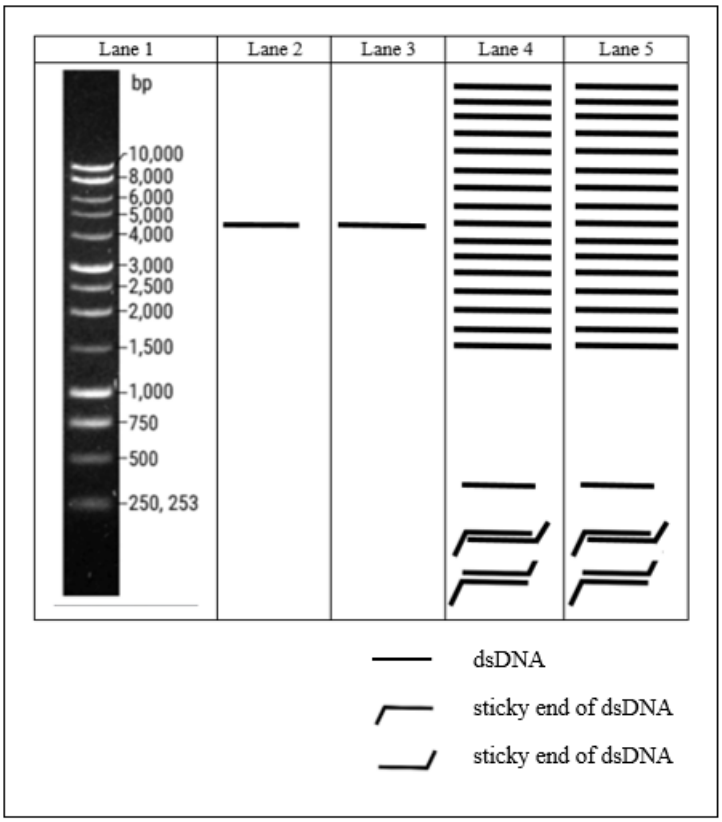

Figure 2: Predicted gel of digestion and ligation towards strand for Model 2.

6 Discussion

In this section, the result which was obtained in the laboratory is presented and analysed together with the dry model, which has been elaborated on earlier.

The experiment's results are provided, in which the final product of gel electrophoresis is subjected to UV rays. The agarose gel was visualised with a UV transilluminator and photographed using a mobile phone. Figure 3 and 4 illustrate the third- and fourth-order limit language, respectively.

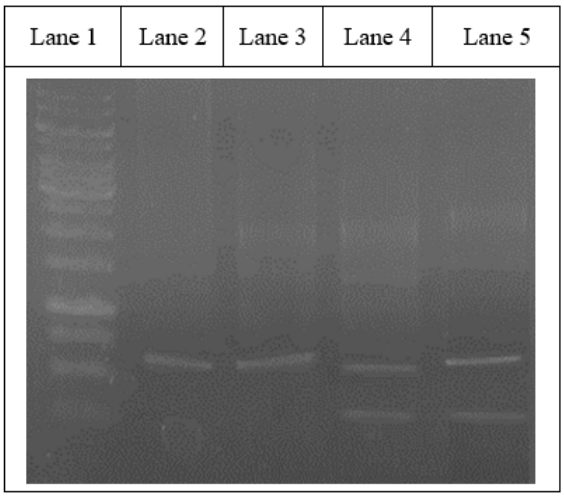

Figure 3: Gel photo of Model 1.

Based on Figure 3, Lane 1 depicts a 1kb DNA ladder capable of measuring $250-10000$bp. Lane 2 displays a single band that emerged between $500-750$bp, suggesting the 610 bp starting strand of dsDNA. Lane 3 also displays a single band that emerged between $500-750$bp, showing the 610 bp initial strand of dsDNA. Since the DNA solution has not been incubated, the digestion process has not yet begun. Lanes 4 and 5 exhibit various bands, representing a few range lengths: 126bp and 130bp, 251bp and 231bp, 357bp, and 381bp. However, many bands over the 610bp range may also be visible on the gel. The reason for the band observed above 610bp is when the fragments B, B', C and C' bind recursively. Therefore, we can conclude that the third-order limit language exists since numerous bands over 610bp are seen in Lanes 4 and 5.

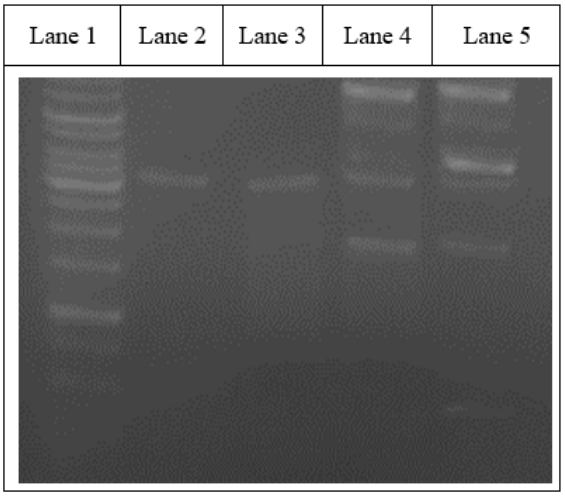

Figure 4: Gel photo of Model 2.

Based on Figure 4, Lane 1 depicts a 1kb DNA ladder capable of measuring $250 - 10000$bp. Meanwhile, Lane 2 reveals a single band that emerged between $4000-5000$bp, showing the 4760bp starting strand of dsDNA. On the other hand, Lane 3 also displays a single band that emerged between $4000-5000$bp, representing the 4760bp initial strand of dsDNA. Since the DNA solution has not been incubated, the digestion process has not yet begun. Lanes 4 and 5 exhibit various bands, representing a few range lengths of $128 - 136$bp, $378 - 386$bp, and $2107 - 2781$bp. Furthermore, several bands emerged in the aforementioned range of 4760bp bands, which are also displayed on the gel. The reason for the band observable above 4760bp is when the fragments and G, G', H, H', I and I' bind recursively. Therefore, we may deduce that the fourth-order limit language exists since numerous bands over 4760bp are seen in Lanes 4 and 5.

7 Conclusion

The existence of third- and fourth-order limit languages has been demonstrated in a laboratory experiment. With the presence of enzymes for Model 1 namely *Msp*I, *Aci*I, and *Mse*I, and Model 2 namely *Age*I, *Eag*I, *BspE*I, and *AvrI*I, the action of 'cut and paste' in a splicing system was projected to converge to a specific set of third- and fourth-order limit language. Moreover, the new definition states that the number of enzymes used in the splicing process determines the order of limit language. Furthermore, Model 1 and 2 demonstrate the third- and fourth-order limit languages accordingly. This has been established as the additional bands appear above the length of the deoxyribonucleic acid (DNA) strand. The n-th order limit language in the dry model is generalised using these two models. In conclusion, given that the model's predictions for the double-stranded DNA (dsDNA) molecules produced in the experiment were accurate, the mathematical representation of the third- and fourth-order limit language has been empirically supported. The theoretical definition of n-th order limit language has been biologically demonstrated in this experiment.

Acknowledgement The authors express their gratitude to the Universiti Malaysia Pahang Al-Sultan Abdullah for its financial support through RDU223229 for this research.

Conflicts of Interest The authors declare no conflict of interest.

References

[1] M. A. Ahmad, N. H. Sarmin, M. F. Abdul-Wahab, W. H. Fong & Y. Yusof (2018). Biomolecular aspects of second-order limit language. *Malaysian Journal of Fundamental and Applied Sciences*, *14*(1), 15–19.

[2] B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts & P. Walter (2014). *Molecular Biology of the Cell*. Garland Science, New York.

[3] M. M. El-Sharkasy (2018). Topological model for recombination of DNA and RNA. *International Journal of Biomathematics*, *11*(8), Article ID: 1850097. https://doi.org/10.1142/S1793524518500973.

[4] H. Hasan (2008). Rekabentuk ukuran ulangan dan proses bercabang dalam biologi. *Malaysian Journal of Mathematical Sciences*, *2*(2), 93–102.

[5] T. Head (1987). Formal language theory and DNA: An analysis of the generative capacity of specific recombinant behaviors. *Bulletin of Mathematical Biology*, *49*(6), 737–759. https://doi.org/10.1016/S0092-8240(87)90018-8.

[6] T. Head, D. Pixton & E. Goode (2003). Splicing systems: Regularity and below. In M. Hagiya & A. Ohuchi (Eds.), *DNA Computing*, volume 2568 of *International Workshop on DNA-Based Computers. Lecture Notes in Computer Science* pp. 262–268. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-36440-4_23.

[7] N. Jonoska & G. Păun (2003). *Aspects of Molecular Computing*. Essays Dedicated to Tom Head on the Occasion of His 70th Birthday. Springer, Verlag Berlin. https://doi.org/10.1007/b94864.

[8] F. Karimi, S. Turaev, N. H. Sarmin & W. H. Fong (2014). Fuzzy splicing systems. In *Computational Collective Intelligence. Technologies and Application*, volume 8733 of *6th International Conference, ICCCI 2014. Lecture Notes in Computer Science* pp. 20–29. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-11289-3_3.

[9] H. L. Koh, H. L. Lee, S. Y. Teh, S. S. Keong & A. I. Md Ismail (2008). Biomathematics: From daphnia to dengue. *Malaysian Journal of Mathematical Sciences*, *2*, 1–10.

[10] T. E. G. Laun (1999). *Constants and Splicing Systems*. PhD thesis, State University of New York, New York.

[11] P. Linz (2015). *An Introduction to Formal Languages and Automata*. Jones & Bartlett Learning, Massachusetts.

[12] S. H. Mohd Khairuddin (2022). *Modelling of n-th Order Limit Language using Automata Theory in DNA Splicing System*. PhD thesis, Universiti Malaysia Pahang Al-Sultan Abdullah, Gambang, Pahang.

[13] S. H. Mohd Khairuddin, M. A. Ahmad & N. Adzhar (2022). Various regulations on the existence of n-th order limit language. *Jurnal Teknologi*, *84*(3), 17–23. https://doi.org/10.11113/jurnalteknologi.v84.17845.

[14] M. H. Mudaber, Y. Yusof & M. S. Mohamad (2017). Differentiating the persistency and permanency of some two stages DNA splicing language via Yusof-Goode (YG) approach. In *Journal of Physics: Conference Series*, volume 890 of *1st International Conference on Applied & Industrial Mathematics and Statistics 2017* pp. 012081. Kuantan, Malaysia. IOP Publishing. https://doi.org/10.1088/1742-6596/890/1/012081.

[15] G. Păun (1996). On the splicing operation. *Discrete Applied Mathematics*, *70*(1), 57–79. https://doi.org/10.1016/0166-218X(96)00101-1.

[16] J. D. Watson & F. H. C. Crick (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, *171*(4356), 737–738. https://doi.org/10.1038/171737a0.

[17] Y. Yusof, N. H. Sarmin, W. H. Fong, T. E. Goode & M. A. Ahmad (2013). An analysis of four variants of splicing system. In *AIP Conference Proceedings*, volume 1522 of *Proceedings of the 20th National Symposium on Mathematical Sciences: Research in Mathematical Sciences: A Catalyst for Creativity and Innovation* pp. 888–895. Putrajaya. American Institute of Physics. https://doi.org/10.1063/1.4801223.