ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

แบบฟอร์มกำหนดโครงการวิศวกรรมคอมพิวเตอร์

ประจำปีการศึกษา 2560

ชื่อหัวข้อโครงงาน

(ภาษาไทย)

ฐานข้อมูลออนไลน์ Short Tandem Repeat เพื่อการวิจัยและทดสอบทางนิติเวช

(ภาษาอังกฤษ)

Short tandem repeat online database for research and test in forensic medicine

ชื่อ – นามสกุลนิสิต เลขประจำตัว ลายมือชื่อนิสิต

1. นาย วศิน ปณิธานศิริกุล 5730540521 ……………………………

Mister Wasin Panitansirikul

อาจารย์ที่ปรึกษาโครงการ ลายมือชื่อ

(หลัก) อ.ดร. ดวงดาว วิชาดากุล …………………………

**สารบัญ**

หัวข้อ หน้า

ชื่อหัวข้อโครงงาน 2

ปัญหาและความสำคัญของปัญหา 2

ทฤษฏีที่เกี่ยวข้อง 3

งานวิจัยที่เกี่ยวข้อง 8

เครื่องมือที่ใช้ในการพัฒนา 9

วัตถุประสงค์ 9

ขอบเขตของโครงงาน 9

ขั้นตอนการดำเนินการ 12

ประโยชน์ที่คาดว่าจะได้รับ 12

เอกสารอ้างอิง 13

**ชื่อหัวข้อโครงงาน**

(ภาษาไทย)

ฐานข้อมูลออนไลน์ Short Tandem Repeat เพื่อการวิจัยและทดสอบทางนิติเวช

(ภาษาอังกฤษ)

Short tandem repeat online database for research and test in forensic medicine

**ปัญหาและความสำคัญของปัญหา**

การพิสูจน์อัตลักษณ์นั้นสามารถทำได้โดยการเปรียบเทียบรหัสพันธุกรรมของตัวอย่างทดสอบว่ามีความคล้ายคลึงกับชุดอ้างอิงมากน้อยเพียงใด โดยวิธีการหลักที่ได้รับการยอมรับเป็นมาตรฐานสากลคือการใช้ส่วน Short Tandem Repeat (STR) ในการจำแนกตัวบุคคลนั้นๆ โดย Short Tandem Repeat นั้นคือส่วนของ ลำดับเบสสั้นๆทีเกิดขึ้นซ้ำๆกันในบริเวณหนึ่งของสาย DNA เราเรียกบริเวณที่ใช้ตรวจสอบดังกล่าวว่า Locus โดยคนแต่ละคนอาจจะมีจำนวนการซ้ำของ STR เท่ากันบ้างในบาง Locus แต่ไม่มีทางที่จะมีจำนวนการซ้ำเท่ากันในทุก Locus อย่างแน่นอน ทำให้เราสามารถระบุตัวตนของบุคคลนั้นๆได้ หากมีข้อมูล Short Tandem Repeat ของบุคคลนั้นๆ

ภาควิชานิติเวชศาสตร์ คณะแพทย์ศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ได้ทำการเก็บข้อมูล STR ดังกล่าวย้อนหลังไว้เป็นเวลากว่า 10 ปีทำให้มีข้อมูลอยู่ปริมาณมาก แต่การจัดเก็บข้อมูลนั้นเก็บเป็น ไฟล์ Excel ทั่วไปซึ่งเป็นอุปสรรคต่อการนำข้อมูลเหล่านั้นมาวิเคราะห์รวมกัน และในส่วนของการวิเคราะห์เบื้องต้นนั้น เจ้าหน้าที่ห้องปฏิบัติการจะเป็นผู้ลงมือทำแบบอนาล็อก ซึ่งสร้างความยุ่งยากและใช้เวลาอย่างมาก ข้าพเจ้าจึงอาสามาช่วยสร้างระบบจัดเก็บข้อมูลและวิเคราะห์ข้อมูลเบื้องต้นเพื่ออำนวยความสะดวกเจ้าหน้าที่ห้องปฏิบัติการต่อไป

**ทฤษฏีที่เกี่ยวข้อง**

1. **Sanger Sequencing (CE)** [11] [12]

เป็นวิธีการในการสกัดสารพันธุกรรมแบบหนึ่ง โดยใช้เทคนิค Capillary Electrophoresis ซึ่งมีขั้นตอนต่างๆดังนี้

* 1. **Amplified DNA**

คือการสังเคราะห์ส่วนของ DNA ที่สนใจเพิ่มเนื่องจากการตรวจสอบจำเป็นต้องใช้ DNA มากปริมาณหนึ่ง จากนั้นให้ความร้อนส่วนของ DNA เหล่านั้นเพื่อให้ DNA แยกออกจากกันเป็นแบบขาเดี่ยว

* 1. **Attach the primer**

เติมส่วนของ primer ที่ขาเดี่ยวของ DNA เหล่านั้นเพื่อเป็นจุดเริ่มต้นสำหรับ DNA polymerase

* 1. **Add the fluorescent base**

เติมเบสชนิดพิเศษที่จะทำให้ DNA polymerase หยุดทำงานเนื่องจากเบสเหล่านี้ไม่มีส่วนของ Hydroxyl-Group ที่ปลาย Carbon 3’ โดยเบสชนิดพิเศษเหล่านี้มีคุณสมบัติเรืองแสงได้

* 1. **Polyacrylamide Gel Electrophoresis**

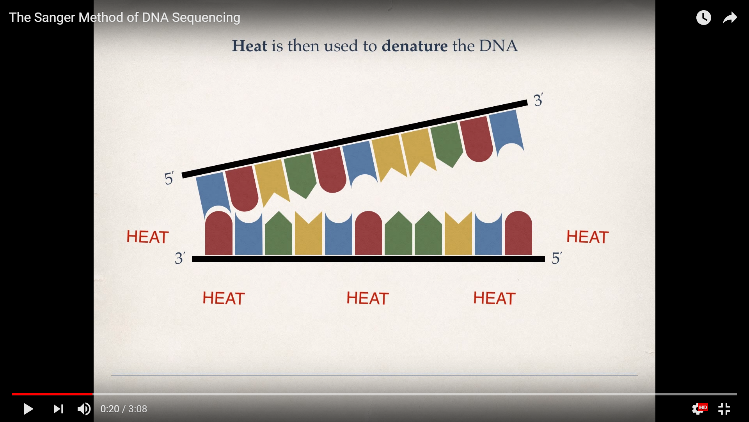
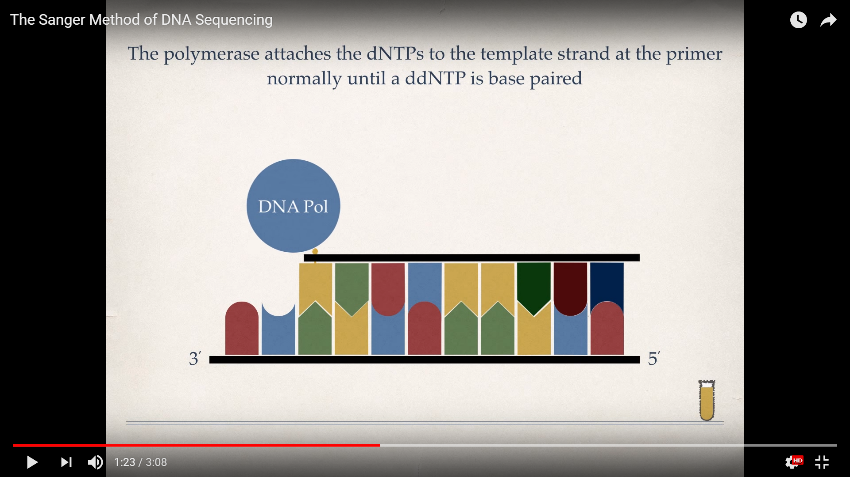
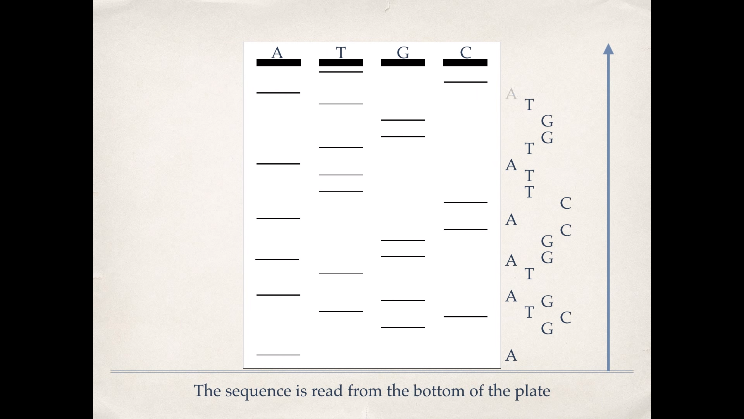
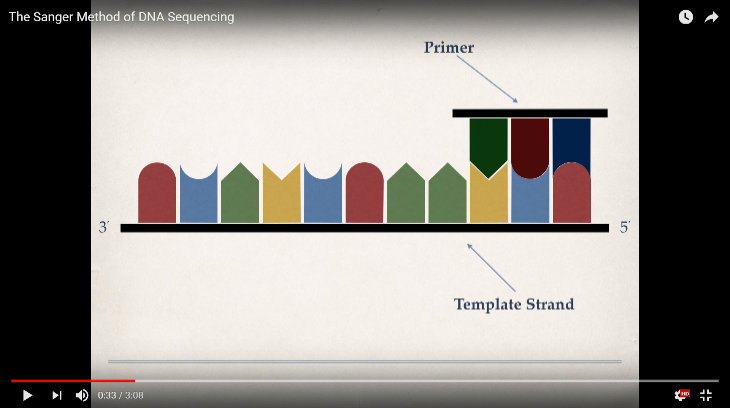
นำ DNA ที่ได้จากกระบวนการก่อนหน้ามาผ่าน Polyacrylamide Gel Electrophoresis Plate ผ่าน Sensor ซึ่งจะอ่านค่าออกมาเป็นแถบสีตามความยาวของ DNA ก่อนหน้าทำให้สามารถแสดงผลเป็นลำดับเบสของ DNA ได้

Figure กระบวนการ Sanger Sequencing

Reference: https://www.youtube.com/watch?v=FvHRio1yyhQ

1. **Next Generation Sequencing (NGS)** [10]

เป็นวิธีการใหม่ในการสกัดสารพันธุกรรม ซึ่งห้องทดลองของภาคนิติเวช คณะแพทย์ศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยนั้นใช้ชุดทดลอง Forenseq เป็นเครื่องมือในการทำ NGS จากบริษัท Illumina โดยสามารถแสดงรายละเอียดของลำดับเบสอย่างชัดเจนทำให้การเปรียบเทียบอัตลักษณ์มีความละเอีนดมากขึ้น กระบวนการทำ NGS สามารถแบ่งออกเป็น 4 ช่วงหลักๆ ดังนี้

* 1. **Library Preparation**

เป็นส่วนของการเตรียม DNA ที่จะใช้ในการตรวจสอบ โดยทำการตัด DNA เป็นส่วนเล็กๆ แล้วเพิ่มส่วนจำเพาะไว้บริเวณหัว และท้ายของสาย DNA จากนั้นจึงส่ง DNA เหล่านี้ไปเข้ากระบวนการ Polymerase Chain Reaction และ gel purified

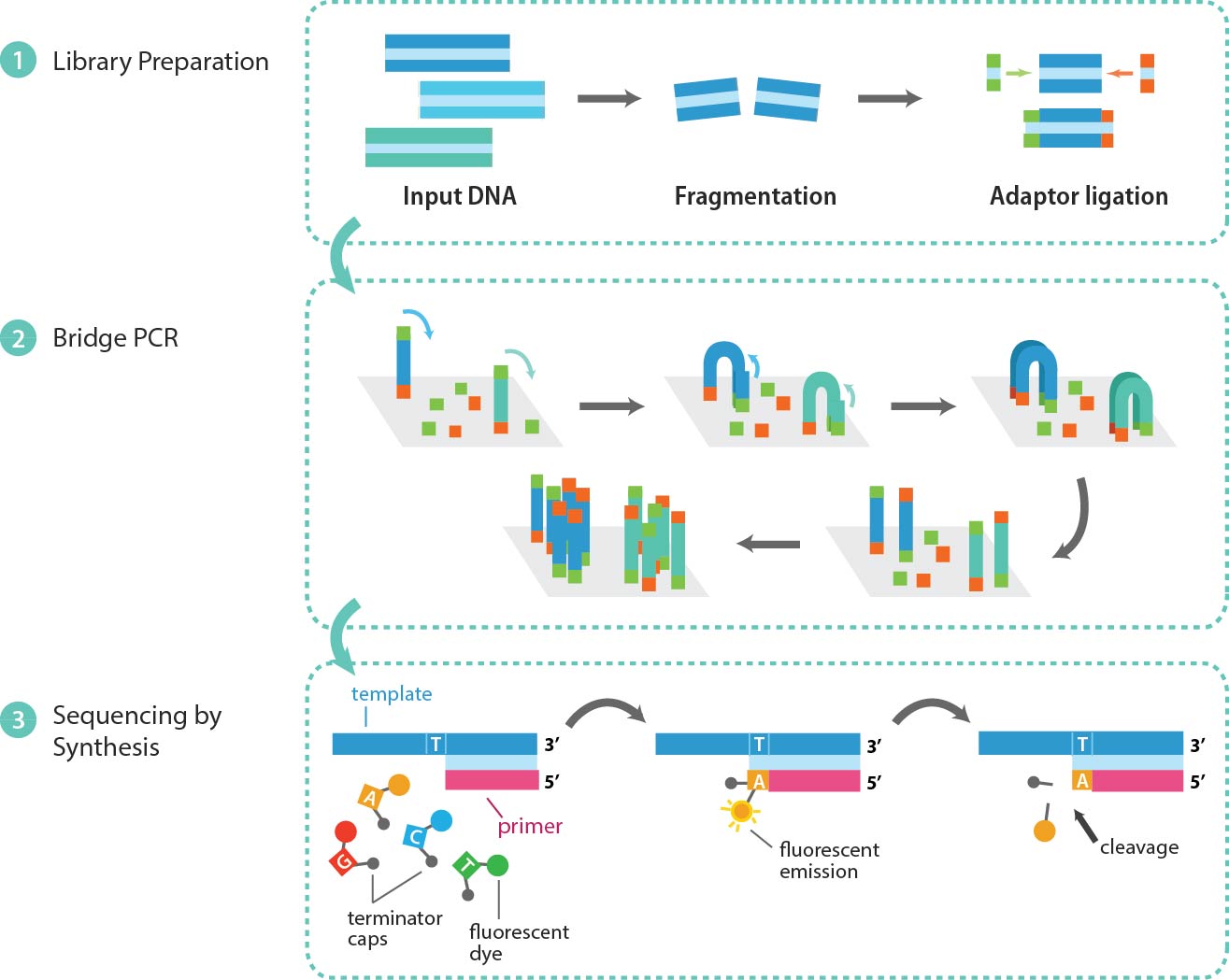


Figure ขั้นตอน Library Preparation

Referece: https://www.abmgood.com/Enzymes/images/NGS-Process.jpg

* 1. **Cluster Generation**

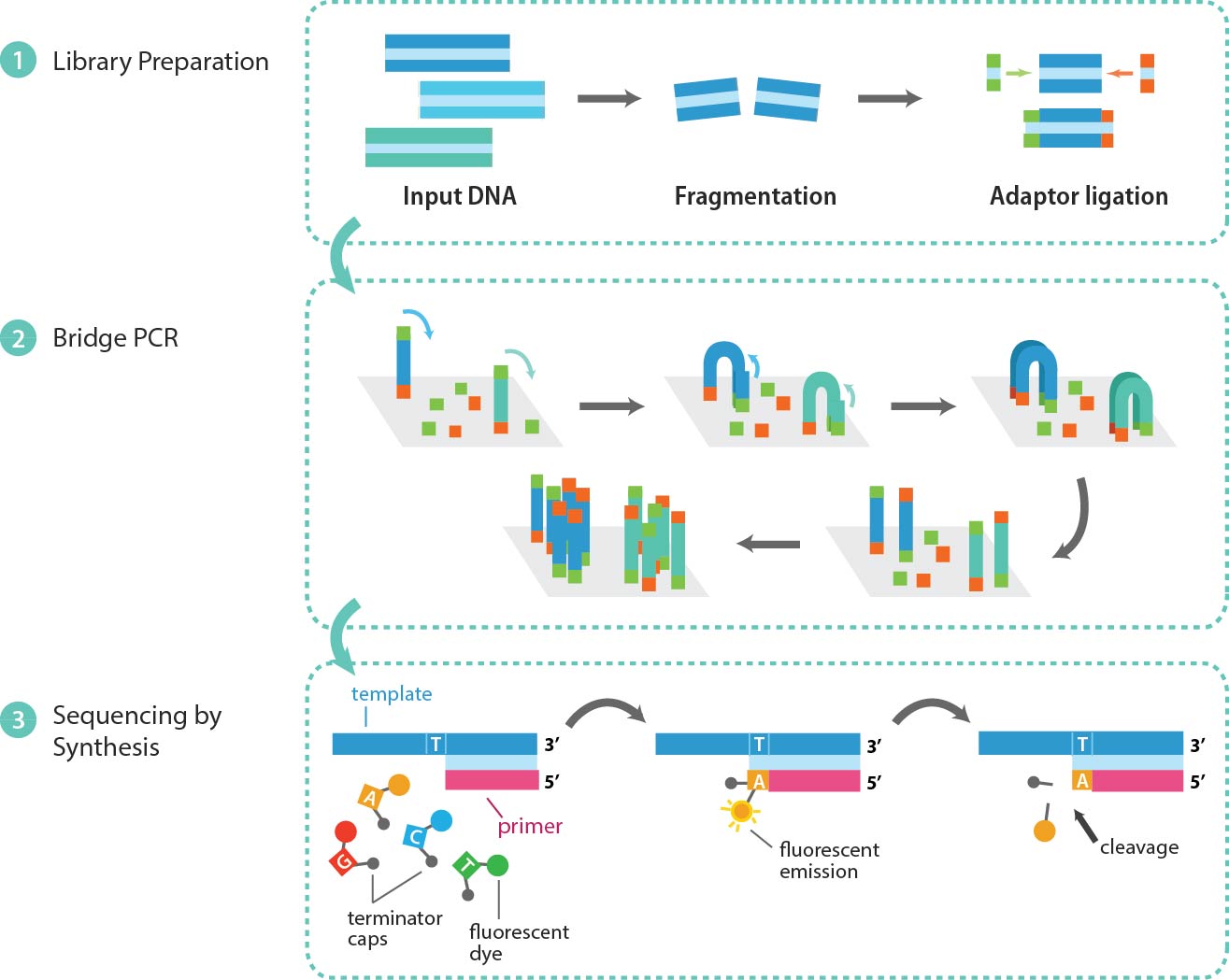
นำ DNA ที่ผ่านกระบวนการ Library Preparation แล้วมาติดกับแผ่น flow cell และเข้าสู่กระบวนการ Bridge Amplification cycle

Figure ขั้นตอน Cluster Generation

Reference: https://www.abmgood.com/Enzymes/images/NGS-Process.jpg

* 1. **Sequencing**

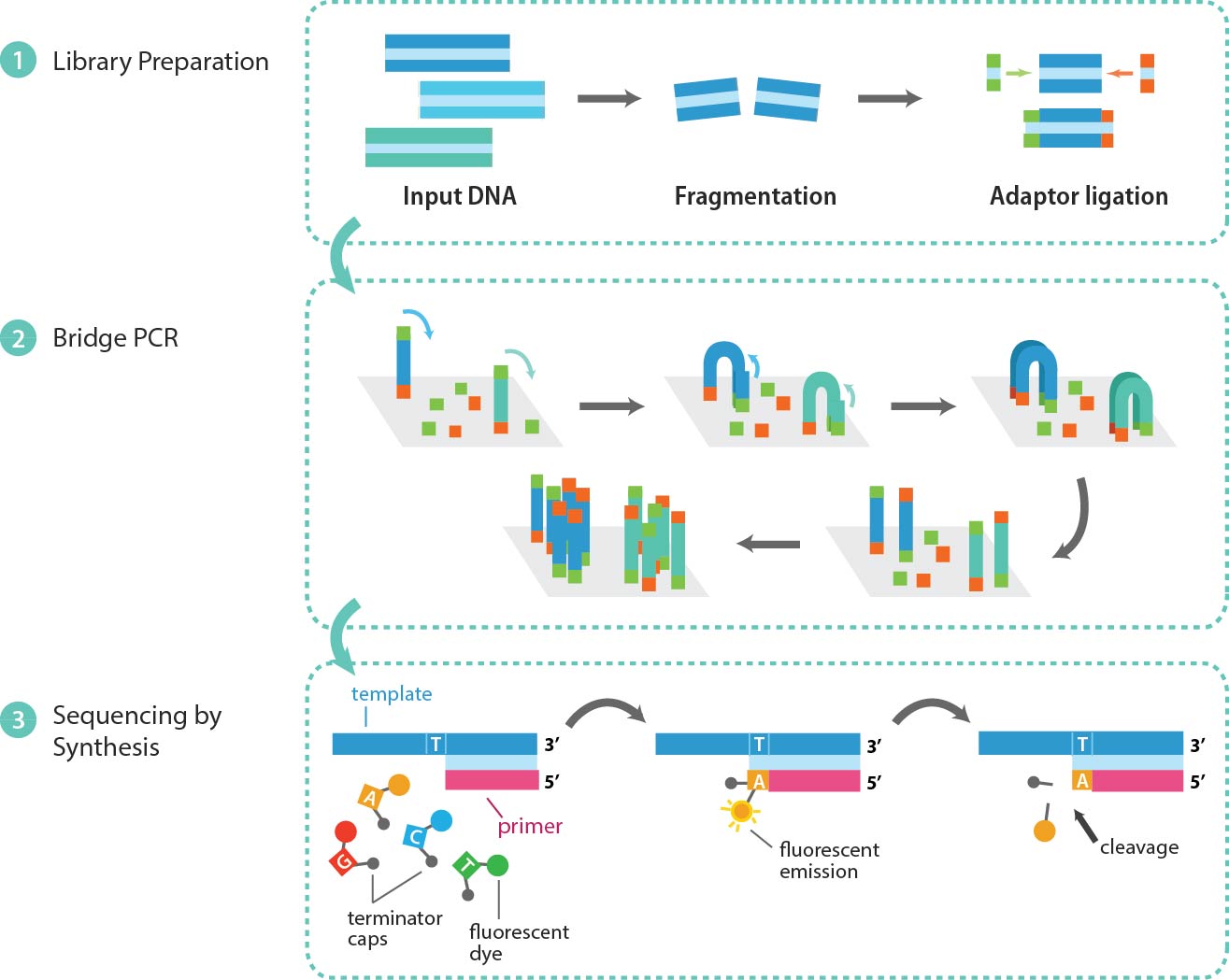
เพิ่ม nucleotide ลงไปใน DNA ที่เตรียมไว้ โดย nucleotide เหล่านี้จะมีลักษณะพิเศษคือถูกกำกับไว้อย่างชัดเจนโดยการเรืองแสงเพื่อให้สามารถตรวจสอบคู่เบสได้อย่างชัดเจน

Figure ขั้นตอน Sequencing

Reference: https://www.abmgood.com/Enzymes/images/NGS-Process.jpg

* 1. **Data Analysis**

นำชิ้นส่วน DNA เหล่านั้นมาเปรียบเทียบกับ Reference genome เพื่อตรวจสอบหาความแตกต่างและสรุปผลเพื่อสร้างลำดับ DNA ที่ถูกต้อง

1. **Chromosome**

คือส่วนเก็บสารพันธุกรรมของมนุษย์และสิ่งมีชีวิต โดยในมนุษย์นั้นถูกแบ่งออกเป็น 2 ส่วนได้แก่

* Autosomal Chromosome

Chromosome ที่เก็บสารพันธุกรรมที่แสดงลักษณะต่างๆของร่างกาย เช่น แขน ขา สีตา จมูก เป็นต้น

* Sex Chromosome

Chromosome ที่เก็บสารพันธุกรรมที่เป็นส่วนกำหนดเพศของสิ่งมีชีวิตได้แก่ Y – Chromosome และ X - Chromosome

1. **Sequence Alignment**

การนำสาย DNA หลายๆเส้นมาเรียงเปรียบเทียบความแตกต่างที่เกิดขึ้นในแต่ละสาย เพื่อนำไปวิเคราะห์ต่อในกระบวนการทางพันธุศาสตร์

1. **ประเภทของ Short Tandem Repeat Sequence**
   1. **Autosomal STR**

ข้อมูลที่ได้จากการถอดรหัส Autosomal Chromosome ซึ่งจะแสดงถึงลักษณะที่แตกต่างกันตามแต่ละ Locus ซึ่ง Locus นี้ก็คือการอ้างอิงส่วนพื้นที่หนึ่งบนสาย DNA โดยจะแสดงส่วนของ Locus ไหนบ้างนั้น ก็จะแตกต่างกันไปตามชนิดชุดตรวจสอบที่ใช้ เช่นกรณีที่ห้องทดลองของคณะแพทย์ศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยใช้ชุดตรวจสอบ Forenseq นั้นจะแสดงค่าออกมาทั้งหมด 27 Loci หรือก็คือ ตรวจสอบ 27 ตำแหน่งบนสาย DNA นั่นเอง

* 1. **Y-STR**

ข้อมูลที่ได้จากการถอดรหัส Y – Chromosome ซึ่งจะแสดงถึงลักษณะที่แตกต่างกันของแต่ละ Locus ในแต่ละบุคคล ซึ่งในส่วนของ Locus นั้น ก็จะแตกต่างกันไปตามชนิดชุดตรวจสอบที่ใช้ โดยสำหรับข้อมูลชนิดนี้จะมีเฉพาะกลุ่มตัวอย่างที่เป็นเพศชายเท่านั้น

* 1. **X-STR**

ข้อมูลที่ได้จากการถอดรหัส X – Chromosome ซึ่งจะแสดงถึงลักษณะที่แตกต่างกันของแต่ละ Locus ในแต่ละบุคคล ซึ่งในส่วนของ Locus นั้น ก็จะแตกต่างกันไปตามชนิดชุดตรวจสอบที่ใช้

* 1. **Single Nucleotide Polymorphism (SNP)**

เป็นการตรวจสอบโดยอาศัยข้อมูลที่แตกต่างจาก STR โดย SNP นั้นจะตรวจสอบชนิดของคู่เบสที่พบใน Locus นั้นๆว่าเป็นชนิดใด ( A C T G ) ซึ่งข้อมูล SNP นั้นสามารถใช้เพื่อระบุลักษณะสำคัญต่างๆของตัวอย่างทดสอบได้ เช่น

* Identity SNP สำหรับระบุตัวตนของบุคคลนั้นๆ
* Phenotypic SNP สำหรับระบุลักษณะต่างๆของตัวอย่างทดสอบ เช่น สีตา
* Ancestral SNP เพื่อระบุความเกี่ยวพันทางสายเลือดของตัวอย่างทดสอบ (ความเป็นพ่อลูก)

1. **ค่าทางสถิติต่างๆ**
   1. **Allele**

คือจำนวนการเกิดซ้ำของช่วง STR ใน Locus ที่ทำการตรวจสอบ เช่น Locus ชื่อ TPOX มี Allele 8 หมายความว่ามีการซ้ำของชุดเบสในส่วนนี้ (AATG) 8 ครั้ง ดังนี้ AATGAATGAATGAATGAATGAATGAATGAATG เป็นต้น

* 1. **Allele frequency**

ปริมาณสัดส่วนของข้อมูลที่มีจำนวนการซ้ำของ Allele ชนิดนั้นๆเทียบกับจำนวนข้อมูลทั้งหมดใน Locus หนึ่ง

* 1. **Heterozygosity**

ปริมาณสัดส่วนของข้อมูลที่มี Type Allele หรือ Allele ที่ใช้ในการประเมินผลทั้ง 2 ตัว เป็นคนละค่ากันเทียบกับข้อมูลทั้งหมดใน Locus หนึ่ง

* 1. **Homozygosity**

ปริมาณสัดส่วนของข้อมูลที่มี Type Allele หรือ Allele ที่ใช้ในการประเมินผลทั้ง 2 ตัว เป็นค่าเดียวกันเทียบกับข้อมูลทั้งหมดใน Locus หนึ่ง

* 1. **Haplotype frequency**

จำนวนความซ้ำที่เกิดขึ้นบนรูปแบบของ Sequence แต่ละ Locus โดยจะเป็นการเก็บข้อมูลเป็นตารางไว้ใช้ในเชิงสถิติ เช่น

|  |  |  |
| --- | --- | --- |
| **จำนวนครั้งการซ้ำ** | **จำนวนรูปแบบของการซ้ำ** | **รูปแบบ Haplotype** |
| 2 | 2 | AATG |
|  |  | ACTG |
| 3 | 1 | TTG |

**งานวิจัยที่เกี่ยวข้อง**

1. **“STRidER STRs for identity ENFSI Reference database”** [5][6]

Strider 2.0 เป็นฐานข้อมูลส่วนกลางของกลุ่มศึกษาวิจัยด้าน DNA ในทวีปยุโรป ที่มีฟังก์ชันในการตรวจสอบข้อมูล forensic ตัวอย่างเทียบกับข้อมูล forensic ในฐานข้อมูลโดยเป็นการพัฒนาต่อยอดมาจาก Strider อีกทีหนึ่ง

1. **YHRD** [7]

ฐานข้อมูล Y-STR haplotypes สำหรับการตรวจสอบ forensic และกรณีตรวจหาความเป็นเครือญาติ

1. **USDR** [8]

ฐานข้อมูล Y-STR haplotypes โดย National Center for Forensic Science ร่วมกับ University of Central Florida

1. **The Allele FREquency Database ( ALFRED )** [9]

ฐานข้อมูลของ Gene Frequency สำหรับประชากรมนุษย์โดยกระทรวงวิทยาศาสตร์ ประเทศสหรัฐอเมริกา

1. **“STRBase: a short tandem repeat DNA database for the human testing community”** [1]

ฐานข้อมูลสำหรับ DNA typing community ที่ประกอบด้วยข้อมูลต่างๆดังนี้

- Short tandem repeat (STR) DNA markers

- STR multiplex kit ชนิดต่างๆ

- Polymerase chain reaction primer sequence ต่างๆ

1. **“A catalog of sequence diversity at human identification Short Tandem Repeat loci”** [3]

ข้อมูล Short Tandem Repeat sequence จากตัวอย่างทดสอบจะถูกเก็บไว้ในรูปแบบ GenBank ใช้สำหรับแลกเปลี่ยนและเปรียบเทียบกันในประเทศต่างๆ โดยในแต่ละ GenBank record นั้นประกอบไปด้วย autosomal STR, Y-chromosomal STRs, X-chromosomal STRs.

โดยจากตัวอย่างฐานข้อมูลดังกว่าวสามารถสรุป functionality ของ ฐานข้อมูลเหล่านั้นออกมาเป็นตารางได้ดังนี้

|  |  |
| --- | --- |
| **ฐานข้อมูลตัวอย่าง** | **ฟังก์ชันที่สามารถทำได้** |
| Strider 2.0 | - เปรียบเทียบชุดข้อมูลทดสอบกับฐานข้อมูลโดยมีชนิดชุดทดลองดังนี้  - SGMplus - ESSplex  - Identifier - ESSplex SE  - Powerplex 16, 18 21 - NGM  - Fusion - NGM-SE  - ESI-16, 17 - ESIX- 16, 17  - Globalfiler  - แสดงสถิติของข้อมูลในฐานข้อมูลในปัจจุบันว่าที่ Locus ใดมีความถี่มากเพียงใด  - แสดงรายละเอียดว่าข้อมูลมาจากประชากรประเทศใดสัดส่วนเท่าใด |
| YHRD | - เปรียบเทียบชุดข้อมูลทุสอบกับฐานข้อมูลโดยมีชนิดชุดทดลองดังนี้  - Minimal - Maximal  - Powerplex Y - Powerplex Y23  - Yfiler - Yfiler Plus  - แสดงสถิติของข้อมูลในฐานข้อมูลในปัจจุบันว่าที่ Locus หนึ่ง Allele ใดมีความถี่มากเพียงใด  - แสดงรายละเอียดว่าข้อมูลมาจากประชากรประเทศใดสัดส่วนเท่าใด |
| USDR | - เปรียบเทียบชุดข้อมูลทดสอบกับฐานข้อมูลโดยมีชนิดชุดทดลองดังนี้  - Powerplex Y - Powerplex Y23  - Yfiler - Yfiler Plus  - แสดงสถิติของข้อมูลในฐานข้อมูลในปัจจุบันว่าที่มีมาจากหน่วยงานใด |
| ALFRED | - การค้นหาโดย Loci, ประชากร  - รายละเอียดของตัวอย่างในแต่ละ Loci ที่มีในฐานข้อมูลได้แก่  - ทวีปที่อยู่ของตัวอย่าง - ข้อมูล Heterozygosity  - สัดส่วนของ gene ที่เจอใน Loci |

**เครื่องมือที่ใช้ในการพัฒนา**

1. **ส่วน Server (Back-end)**

- implement โดย Framework Node JS

1. **ส่วนติดต่อกับผู้ใช้ (Front-end)**

- implement โดย library React JS

1. **ฐานข้อมูล**

- implement โดย MySQL Database

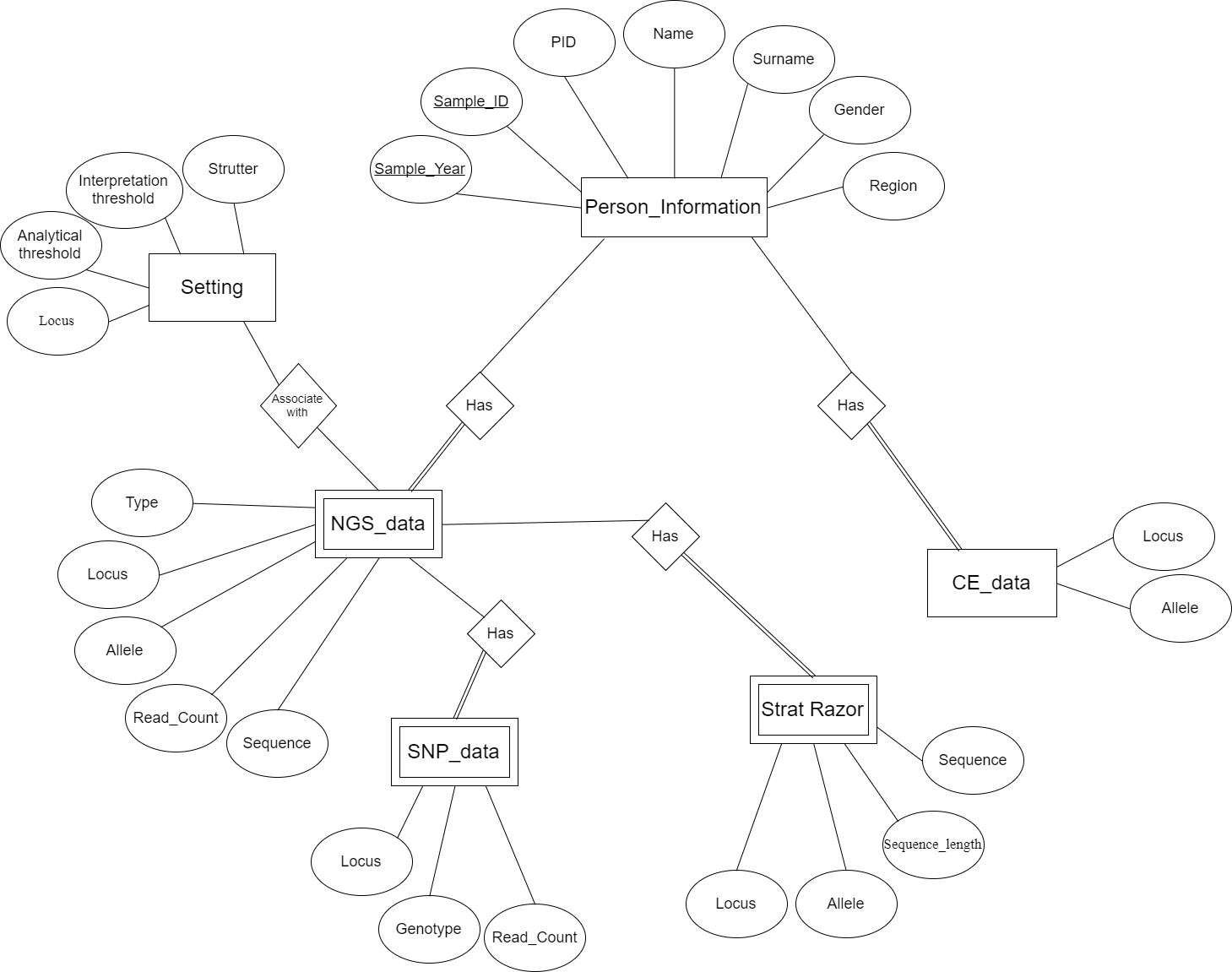
1. **การจัดการเกี่ยวกับการวิเคราะห์ข้อมูล**

- implement โดยใช้ Python

**วัตถุประสงค์**

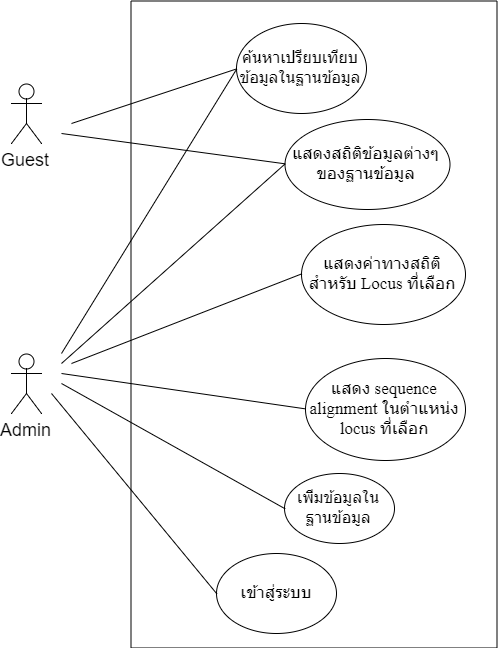
1. เพื่อสร้างระบบฐานข้อมูลที่ช่วยอำนวยความสะดวกให้กระบวนการตรวจสอบและการวิจัยทางนิติเวช

**ขอบเขตของโครงงาน**

1. ศึกษาฐานข้อมูลที่เกี่ยวข้องกับ STRs ที่มีอยู่ในปัจจุบัน
2. วิเคราะห์ความแตกต่างของแต่ละฐานข้อมูลที่ได้ไปศึกษามา
3. ออกแบบและพัฒนาระบบฐานข้อมูลใหม่ที่มีความเหมาะสมกับข้อมูลที่ต้องการจัดการโดยมีความสามารถดังนี้
   * เก็บและเพิ่มข้อมูลในรูปแบบฐานข้อมูล
   * ค้นหาเปรียบเทียบข้อมูลตัวอย่างกับข้อมูลในฐานข้อมูล
   * ทำ Sequence Alignment สำหรับ Locus ที่ต้องการ
   * แสดงสถิติข้อมูลที่มีอยู่ในฐานข้อมูล
   * คำนวณค่า Allele Frequency, Heterozygosity, Homozegosity และ Haplotype Frequency ได้

โดย ER ข้างต้นมี Data Dictionary ดังนี้

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Name | Description | Attribut Name | Attribute description | Data Type | Length | Key | Nullable |
| Personal\_  Information | เก็บข้อมูลประวัติส่วนบุคคลต่างๆ | Sample\_Year | ปี พศ. ที่ตัวอย่างทำการตรวจสอบ | STRING | 2 | Y | N |
|  |  | Sample\_ID | หมายเลขกำกับของตัวอย่าง | STRING | 6 | Y | N |
|  |  | PID | รหัสประจำตัวประชาชน | STRING | 13 | N | N |
|  |  | Name | ชื่อเจ้าของข้อมูล | STRING | 20 | N | N |
|  |  | Surname | นามสกุลเจ้าของข้อมูล | STRING | 40 | N | N |
|  |  | Gender | เพศของเจ้าของข้อมูล | STRING | 1 | N | N |
|  |  | Region | ภูมิลำเนาเจ้าของข้อมูล | STRING | 10 | N | N |
| CE\_data | ข้อมูลที่ได้มาจากกระบวนการ Sanger Sequencing | Locus | ชื่อตำแหน่งอ้างอิงบนสาย DNA | STRING | 12 | N | N |
|  |  | Allele | จำนวนการซ้ำของส่วน STR | STRING | 2 | N | N |
| NGS\_data | ข้อมูลที่ได้มาจากกระบวนการ Next Generation Sequencing | Type | แสดงว่าเป็น STR จาก chromosome ประเภทใด | STRING | 1 | N | N |
|  |  | Locus | ชื่อตำแหน่งอ้างอิงบนสาย DNA | STRING | 12 | N | N |
|  |  | Allele | จำนวนการซ้ำของส่วน STR | STRING | 2 | N | N |
|  |  | Read\_Count | จำนวนครั้งของการปรากฏ ในขั้นตอนการถอดรหัส | Number | - | N | N |
|  |  | Sequence | ลำดับเบสที่เกิดขึ้น | STRING | 255 | N | N |
| SNP\_data | ข้อมูลประเภท Single Nucleotide Polymorphism | Locus | ชื่อตำแหน่งอ้างอิงบนสาย DNA | STRING | 12 | N | N |
|  |  | Genotype | ชนิดของคู่เบสที่พบใน Locus นี้ | STRING | 1 | N | N |
|  |  | Read\_Count | จำนวนครั้งของการปรากฏ ในขั้นตอนการถอดรหัส | Number | - | N | N |
| Strat Razor | ข้อมูลที่ได้มาจากกระบวนการ Strat Razor Analysis | Locus | ชื่อตำแหน่งอ้างอิงบนสาย DNA | STRING | 12 | N | N |
|  |  | Allele | จำนวนการซ้ำของส่วน STR | STRING | 2 | N | N |
|  |  | Sequence | ลำดับเบสที่เกิดขึ้น | STRING | 255 | N | N |
|  |  | Sequence\_length | ความยาวของลำดับเบส | Number | - | N | N |
| Setting | เงื่อนไขที่ใช้ในการแปรผล | Locus | ชื่อตำแหน่งอ้างอิงบนสาย DNA | STRING | 12 | N | N |
|  |  | Strutter |  | STRING | - | N | N |
|  |  | Interpretation\_threshold | ค่าขั้นต่ำที่จะให้การยอมรับผลการทดสอบนั้น | Number | - | N | N |
|  |  | Analytical\_threshold | ต่าขั้นต่ำที่จะนำผลการทดสอบนั้นไปใช้ได้จริง | Number | - | N | N |

1. ****ทดสอบและประเมินผลการทำงานของระบบใหม่โดยจะต้องมี use case diagram ดังนี้

**ขั้นตอนการดำเนินงาน**

1. กำหนดวัตถุประสงค์ เป้าหมาย และขอบเขตของโครงการ
2. ศึกษาค้นคว้าทฤษฏีที่เกี่ยวข้อง
3. ศึกษาการทำงานของระบบฐานข้อมูลตัวอย่างที่มีอยู่แล้ว
4. ทำการพัฒนาระบบฐานข้อมูลสำหรับข้อมูลต้องการจัดเก็บ
5. ทดสอบประสิทธิภาพของระบบฐานข้อมูลใหม่
6. ศึกษาค้นคว้าเพิ่มเติมและแก้ไขสิ่งที่อาจสามารถเพิ่มประสิทธิภาพให้กับระบบได้
7. รวบรวมข้อมูลเพื่อจัดทำรายงาน
8. จัดทำรายงานฉบับสมบูรณ์
9. ตรวจทานและแก้ไขรายงานฉบับสมบูรณ์

**ประโยชน์ที่คาดว่าจะได้รับ**

1. ได้ระบบที่ช่วยเหลืองานวิจัย และการทำงานด้านนิติเวช
2. ได้รับความรู้จากการศึกษาเรื่องพันธุกรรมของมนุษย์ และส่วนที่เกี่ยวข้อง
3. ได้รับความชำนาญจากการ implement web application

**เอกสารอ้างอิง**

[1] Christian M. Ruitberg, Dennis J. Reeder and John M. Butler, 2001, STRBase: a short tandem repeat DNA database for the human identity testing community.: Nucleic Acids Research, v.29, p. 320 – 322.

[2] “a short tandem repeat DNA database for the human identity testing community”, [Online]. Available at: http://strbase.nist.gov/

[3] Gettings, K. B., L. A. Borsuk, D. Ballard, M. Bodner, B. Budowle, L. Devesse, J. King, W. Parson, C. Phillips, and P. M. Vallone, 2017, STRSeq: A catalog of sequence diversity at human identification Short Tandem Repeat loci.: Forensic Sci Int Genet, v. 31, p. 111-117.

[4] “How does Illumina NGS work”, [Online]. Available at: <https://www.illumina.com/science/technology/next-generation-sequencing.html>

[5] “STRidER STRs for identity ENFSI Reference database”, [Online]. Available at: <https://strider.online/>

[6] Bodner M, Bastisch I, Butler JM, Fimmers R, Gill P, Gusmão L, Morling N, Phillips C, Prinz M, Schneider PM, Parson W (2016) Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal Short Tandem Repeat allele frequency databasing (STRidER); [Forensic Sci Int Gen 24:97-102](https://www.isfg.org/Publication;Bodner2016)

[7] “YHRD international Y-STR haplotypes database”, [Online]. Available at: <https://yhrd.org/>

[8] “USDR USA Y-STR haplotypes database”, [Online]. Available at: [https://www.usystrdatabase.org/](https://l.facebook.com/l.php?u=https%3A%2F%2Fwww.usystrdatabase.org%2F&h=ATNGmzc7Ry5oh7ILLbHxkak35YMVkOyrui9Y2ZDq3p5TrhVCPj80Ldp8_AtRQHzHNwLf3ZvRhJB8B1e1UVAKuuXW-FHK1XtAToQblv1XrHbZ8qvnAmVUJ41PTZVpCQUjPnykwW8F4Q7zkw)

[9] Cheung KH, Miller PL, Kidd JR, Kidd KK, Osier MV, Pakstis AJ. "ALFRED: a Web-accessible allele frequency database".*Pac Symp Biocomput 2000*.:639-50, [Online]. Available at: <https://alfred.med.yale.edu/>

[10] YaranYang ,BingbingXie, JiangweiYan, 2014, Application of Next-generation Sequencing Technology in Forensic Science.: Genomics, Proteomics & Bioinformatics, v12 p 190-197

[11] McCord, B. Encyclopedia of Forensic Sciences Capillary Electrophoresis in Forensic Genetics, 2013, 394-401.

[12] Oorschot R ; Ballantyne K. Capillary Electrophoresis in Forensic Biology, 2013, 560-566.