

ระบบสนับสนุนการจัดเก็บและวิเคราะห์ข้อมูลลำดับซ้ำเรียงต่อเนื่องแบบสั้นสำหรับนิติวิทยาศาสตร์



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2565

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

A Support System for Collecting and Analyzing Short Tandem Repeats for Forensic
Science



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering in Computer Engineering

Department of Computer Engineering

FACULTY OF ENGINEERING

Chulalongkorn University

Academic Year 2022

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	ระบบสนับสนุนการจัดเก็บและวิเคราะห์ข้อมูลลำดับซ้ำเรียง ต่อเนื่องแบบสั้นสำหรับนิติวิทยาศาสตร์
โดย	นายณัฐรัชย์ กุลธรรมนิตย์
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร.ดวงดาว วิชาตากุล

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่ง
ของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

.....	คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)	
คณะกรรมการสอบวิทยานิพนธ์	
.....	ประธานกรรมการ
(รองศาสตราจารย์ ดร.วิวัฒน์ วัฒนาวุฒิ)	
.....	อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร.ดวงดาว วิชาตากุล)	
.....	กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.พิตติพล คันธวัฒน์)	
.....	กรรมการภายนอกมหาวิทยาลัย
(ผู้ช่วยศาสตราจารย์ ดร.นุรีย์ วิวัฒน์วัฒนา)	

ณัฐชัย กุลธรรมนิตย์ : ระบบสนับสนุนการจัดเก็บและวิเคราะห์ข้อมูลลำดับซ้ำเรียง
ต่อเนื่องแบบสั้นสำหรับนิติวิทยาศาสตร์. (A Support System for Collecting and
Analyzing Short Tandem Repeats for Forensic Science) อ.ที่ปรึกษาหลัก : รศ.
ดร.ดวงดาว วิชาดากุล

ลำดับซ้ำเรียงต่อเนื่องแบบสั้น (Short Tandem Repeat) หรือเอสทีอาร์ (STR) เป็น
ลำดับที่ซ้ำกันเป็นชุด ๆ ที่พบได้ในจีโนม (Genome) ของมนุษย์และมีประโยชน์มากในนิติ
วิทยาศาสตร์ เช่น การยืนยันตัวบุคคล การหาความสัมพันธ์ทางเครือญาติ เทคโนโลยีการลำดับเบส
ยุคใหม่ (Next-Generaton Sequencing: NGS) เช่น ForenSeq Signature Prep สามารถหา
ลำดับ STRs และให้ข้อมูลเชิงลึกเกี่ยวกับโครงสร้างประชากรได้ ถึงแม้ว่าเอสทีอาร์ที่ได้จาก
เทคโนโลยีนี้จะมีประโยชน์มากมาย แต่ไม่มีแพลตฟอร์มซอฟต์แวร์โอเพ่นซอร์สใดที่รวมการจัดการ
และการวิเคราะห์ข้อมูลของเอสทีอาร์ไว้ในแพลตฟอร์มเดียว ผู้ใช้งานอาจต้องใช้หลายโปรแกรมใน
การวิเคราะห์ข้อมูลเอสทีอาร์ จากนั้นรวบรวมผลลัพธ์ลงในฐานข้อมูลแยกหรือโพลเดอร์ระบบไฟล์
เพื่อแก้ไขปัญหาดังกล่าว ระบบที่นำเสนอ STRategy เป็นเว็บแอปพลิเคชันที่มีระบบการจัดการ
และวิเคราะห์ข้อมูลเอสทีอาร์ โดย STRategy อนุญาตให้ผู้ใช้งานเก็บข้อมูลลงฐานข้อมูลหลังจาก
นั้นระบบจะวิเคราะห์และแสดงผลข้อมูลโดยอัตโนมัติ ระบบนี้ถูกออกแบบให้ใช้ในองค์กรหรือ
ห้องปฏิบัติการ จึงมีระบบการกำหนดสิทธิผู้ใช้ระบบตามบทบาท (Role Based Access Control)
เพื่อให้ผู้ใช้งานเข้าถึงข้อมูลตามสิทธิของแต่ละบุคคลเท่านั้น ระบบถูกออกแบบให้มีความยืดหยุ่นสูง
และได้ปฏิบัติตามแนวคิดสถาปัตยกรรม 3-เลเยอร์ หลักการออกแบบโปรแกรมเชิงวัตถุ รวมถึงมี
การใช้แบบรูป (Design patterns) ต่าง ๆ เพื่อรองรับการเพิ่มโมดูลการวิเคราะห์และโมดูลการ
แสดงผล รวมถึงการพัฒนาต่อยอดได้ง่ายในอนาคต

สาขาวิชา วิศวกรรมคอมพิวเตอร์

ปีการศึกษา 2565

ลายมือชื่อนิสิต

ลายมือชื่อ อ.ที่ปรึกษาหลัก

6370086021 : MAJOR COMPUTER ENGINEERING

KEYWORD: short tandem repeats, pattern alignment, forensic science, support system for collecting and analyzing, str

Nuttachai Kulthammanit : A Support System for Collecting and Analyzing Short Tandem Repeats for Forensic Science. Advisor: Assoc. Prof. Duangdao Wichadakul

Short tandem repeats (STRs) are short repeated sequences commonly found in the human genome and valuable in forensic science, used for human identity and relatedness markers. Next-generation sequencing (NGS) technologies, e.g., ForenSeq Signature Prep, can sequence STRs, providing valuable insights into population and sub-population structures. Despite the potential benefits of NGS for STRs, no open-source software platform integrates the collection, management, and analysis of STR data from NGS into one place. Users must use multiple programs to process their STR data and then collect the results into a separate database or a file system folder. To address this gap, we introduce STRategy, a standalone web-based application supporting essential STR data management and analysis capabilities. STRategy allows users to collect their data into its database, automatically analyze data, and visualizes the data in various forms. STRategy was designed for internal use in a laboratory or an organization. Hence, our system includes role-based access control that allows users to search for or access specific data based on their responsibilities. This system was designed based on the 3-layer architecture, the principles of object-oriented design, and design patterns, which enable the flexible extension of new data formats and add-on modules for various analyses and visualization in the future.

Field of Study: Computer Engineering

Student's Signature

Academic Year: 2022

Advisor's Signature

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความอนุเคราะห์ของ รศ. ดร.ดวงดาว วิชาดากุล อาจารย์ที่ปรึกษาวิทยานิพนธ์ซึ่งสละเวลาให้คำปรึกษา ช่วยตรวจสอบแก้ไขข้อบกพร่องต่าง ๆ จนทำให้การวิจัยครั้งนี้สำเร็จลุล่วงไปได้ด้วยดี

ขอขอบพระคุณกรรมการสอบวิทยานิพนธ์ รศ. ดร.วิวัฒน์ วัฒนาวุฒิ ผศ. ดร.พิติพล คันธวัฒน์ และ ผศ. ดร.นุวิทย์ วิวัฒน์วัฒนา ที่กรุณาสละเวลาให้คำแนะนำ ตรวจสอบและแก้ไขวิทยานิพนธ์ ซึ่งเป็นประโยชน์ในการทำวิทยานิพนธ์ฉบับนี้อย่างยิ่ง

ขอขอบพระคุณ รศ. นพ.กรเกียรติ วงศ์ไพศาลสิน นางสาวปุณยภัทร สุขวุฒิมิยา นางสาวทิฆัมพร สติรแพทย์ และ นางสาวอัสนีย์ โนะ ที่สนับสนุนตัวอย่างข้อมูลลำดับซ้ำเรียงต่อเนื่องแบบสั้น พร้อมแนวทางการวิเคราะห์ข้อมูล ตรวจสอบความถูกต้องและแปลผลข้อมูล

ขอขอบพระคุณอาจารย์ทุกท่านที่ได้ประสิทธิ์ประสาทความรู้ตลอดจนให้ข้อคิดต่าง ๆ ซึ่งเป็นแรงผลักดันสำคัญ ให้มีความมุ่งมั่นตั้งใจทำงานวิจัยนี้ให้ดีที่สุด

ท้ายที่สุด ขอกราบขอบพระคุณ ครอบครัวสำหรับกำลังใจ รวมถึงขอขอบพระคุณผู้บังคับบัญชาในสายงาน เพื่อนร่วมงาน รุ่นพี่ รุ่นน้องและมิตรสหาย ที่คอยให้กำลังใจและช่วยเหลือในด้านต่าง ๆ

สุดท้ายนี้ข้าพเจ้าหวังเป็นอย่างยิ่งว่า เนื้อหาในวิทยานิพนธ์ฉบับนี้จะประโยชน์แก่ผู้อื่นไม่มากนัก

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ณัฐชัย กุลธรรมนิตย์

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ค
บทคัดย่อภาษาอังกฤษ	ง
กิตติกรรมประกาศ.....	จ
สารบัญ.....	ฉ
สารบัญตาราง.....	ญ
สารบัญรูป	ฎ
บทที่ 1	1
1.1. ความเป็นมาและความสำคัญของปัญหา.....	1
1.2. วัตถุประสงค์ของงานวิจัย	2
1.3. ขอบเขตงานวิจัย.....	3
1.4. ขั้นตอนและแผนการดำเนินการวิจัย.....	4
1.5. ประโยชน์ที่ได้รับ	5
1.6. การตีพิมพ์ในวารสารวิชาการนานาชาติ.....	5
บทที่ 2	6
2.1. ทฤษฎีที่เกี่ยวข้อง	6
2.1.1. โครโมโซม (Chromosome).....	6
2.1.2. ดีเอ็นเอ (DNA)	6
2.1.3. ยีน (Gene).....	7
2.1.4. แอลลีล (Allele).....	8
2.1.5. แฮพลอยด์ (Haploid) และ ดิพลอยด์ (Diploid)	9
2.1.6. การระบุบุคคลจากสารชีวโมเลกุล	9

2.1.7. ลำดับซ้ำเรียงต่อเนื่องแบบสั้น (Short Tandem Repeats).....	9
2.1.8. อิเล็กโทรโฟรีซิสหลอดเล็ก (Capillary Electrophoresis).....	10
2.1.9. การหาลำดับเบสยุคใหม่ (Next-Generation Sequencing)	11
2.1.10. พหุสัณฐานเดี่ยวของนิวคลีโอไทด์ (single nucleotide polymorphism).....	12
2.1.11. การสืบค้นแบบโปรไฟล์.....	14
2.1.12. กระบวนการรีเวิร์สคอมพลีเมนต์ (Reverse complement).....	16
2.1.13. GeoJSON.....	17
2.1.14. การเขียนโปรแกรมเชิงวัตถุ (Object Oriented Programming).....	17
2.1.15. หลักการ SOLID (SOLID principle).....	19
2.1.16. สถาปัตยกรรม 3-เลเยอร์ (3-layer architecture).....	20
2.1.17. สถาปัตยกรรมเชิงคอมโพเนนท์ (Component-based architecture)	21
2.1.18. แบบรูปการออกแบบ (Design patterns).....	21
2.1.19. หลักการอินเวอร์ชันออฟคอนโทรล (Inversion of control principle).....	22
2.1.20. หลักการดีเพนเดนซีอินเจกชัน (Dependency Injection principle).....	22
2.1.21. แบบรูปโมเดล-วิว-คอนโทรลเลอร์ (Model-view-controller design pattern) ...	23
2.1.22. Object Relational Mapping (ORM)	23
2.1.23. Java Persistence API (JPA)	24
2.2. งานวิจัยที่เกี่ยวข้อง.....	24
2.2.1. ฐานข้อมูล STRidER (STRs for Identity ENFSI Reference Database).....	24
2.2.2. NIST Standard Reference Database SRD 130	25
2.2.3. US Y-STR Database (YHRD).....	25
2.2.4. The ALlele FREquency Database (ALFRED)	26
2.2.5. STRait Razor 3.0.....	26
2.2.6. Excel-based workbook of STRait Razor 3.0	26

2.2.7. A survey of tools for analysing DNA fingerprints	27
2.2.8. Short tandem repeat typing technologies used in human identity testing	27
2.2.9. STRAF (STR Analysis for Forensics).....	28
บทที่ 3	30
3.1. ภาพรวมการทำงานของระบบ	30
3.2. วิเคราะห์และออกแบบฐานข้อมูลของระบบ	31
3.2.1. ศึกษาโครงสร้างข้อมูลเอสทีอาร์.....	31
3.2.2. การออกแบบตารางข้อมูลเพื่อรองรับการตั้งค่าต่าง ๆ	35
3.2.3. แนวทางการออกแบบการกำหนดสิทธิของผู้ใช้ในระบบ	36
3.2.4. การวิเคราะห์เพื่อเปรียบเทียบลำดับเบสดีเอ็นเอ	37
3.3. สถาปัตยกรรมของระบบ	38
3.4. การประยุกต์ทฤษฎีทางคอมพิวเตอร์	41
3.4.1. โมเดล-วิว-คอนโทรลเลอร์ (Model-View-Controller).....	41
3.4.2. สถาปัตยกรรมเชิงคอมโพเนนต์ (Component-based architecture)	42
3.4.3. แบบรูปการออกแบบ (Design pattern).....	42
3.4.4. การแปลงอ็อบเจกต์เชิงสัมพันธ์ (Object-Relational Mapping: ORM)	46
3.5. การติดตั้งและการใช้งานระบบ	47
บทที่ 4	49
4.1. ภาพรวมของระบบ	49
4.2. การสืบค้นข้อมูลเอสทีอาร์	49
4.3. การแสดงผลข้อมูลภาพรวม.....	51
4.4. ส่วนต่อประสานโปรแกรมประยุกต์สำหรับบริการภายนอก	57
4.5. แนวทางการพัฒนาต่อยอดอัลกอริทึมวิเคราะห์ข้อมูลในระบบ	59

4.6. การจัดการข้อมูลของผู้ใช้งานห้องทดลอง	60
4.7. การจัดการระบบโดยผู้ดูแลระบบ	61
4.8. ที่มาของข้อมูลตัวอย่างภายในระบบ	65
4.9. การตรวจสอบความถูกต้องของการวิเคราะห์ข้อมูลภายในระบบ	65
4.10. เทคโนโลยีที่ใช้ในการพัฒนา	65
บทที่ 5	67
5.1. การอภิปรายผล.....	67
5.2. สรุปผลการวิจัย	68
5.3. แนวทางการวิจัยในอนาคต.....	69
ภาคผนวก.....	70
ภาคผนวก ก.....	71
ภาคผนวก ข.....	81
บรรณานุกรม.....	83
ประวัติผู้เขียน	90

สารบัญตาราง

	หน้า
ตารางที่ 1 ฐานข้อมูลและฟังก์ชันการทำงานของแต่ละระบบ	29
ตารางที่ 2 เลขเอร์ของระบบหลังบ้าน	39



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญรูป

	หน้า
รูปที่ 1 โครงสร้างโครโมโซม [16]	6
รูปที่ 2 ดีเอ็นเอ [19].....	7
รูปที่ 3 ยีนและโครโมโซม [21]	8
รูปที่ 4 แอลลีล จีโนไทป์ และฟีโนไทป์ของกรุปเลือด [25].....	8
รูปที่ 5 ตัวอย่างลำดับซ้ำเรียงต่อเนื่องแบบสั้นจากเทคนิคอิเล็กทรอนิกส์โทรฟอรีซิสหลอดเล็ก [30]	9
รูปที่ 6 แถบเจลที่แสดงแอลลีลเทียบกับแอลลีลมาตรฐาน (Allelic ladder) ของแต่ละโลคัส (ภาพที่ 1 ของ [32]).....	10
รูปที่ 7 อิเล็กโทรฟอรีซิสหลอดเล็ก [33]	11
รูปที่ 8 รูปแบบข้อมูลอิเล็กทรอนิกส์โทรฟอรีซิสหลอดเล็ก	11
รูปที่ 9 รูปแบบข้อมูลจาก ForenSeq (ก) ตารางจีโนไทป์ของเอสทีอาร์ (ข) ตารางแอลลีลและลำดับเบสดีเอ็นเอ	12
รูปที่ 10 การแทนที่ของลำดับเบส (รูปที่ 1 ของ [36])	13
รูปที่ 11 รูปแบบแฮพลไทป์บล็อกที่เกิดจากสลิป 3 ตำแหน่ง (ก) รูปแบบทั้งหมดที่เป็นไปได้ (ข) รูปแบบที่เกิดขึ้นจริง (รูปที่ 2 ของ [36]).....	14
รูปที่ 12 โปรแกรม OmniPop 200.1 (ก) หน้าสำหรับใส่ข้อมูลเพื่อค้นหา (ข) ข้อมูลความถี่แอลลีล	15
รูปที่ 13 ตัวอย่างการคำนวณความถี่จีโนไทป์	15
รูปที่ 14 ตัวอย่างผลลัพธ์ของการสืบค้นแบบโปรไฟล์ (ก) ผลลัพธ์ที่ได้จากการคูณค่าส่วนกลับของความถี่จีโนไทป์ของประเทศอเมริกา (ข) ค่าความสัมพันธ์ระหว่างโปรไฟล์ตัวอย่างกับข้อมูลที่มีอยู่ในฐานข้อมูล.....	16
รูปที่ 15 ตัวอย่างรูปเรขาคณิตที่สร้างจาก TopoJSON [48].....	17
รูปที่ 16 สถาปัตยกรรม 3-เลเยอร์.....	20
รูปที่ 17 สถาปัตยกรรมเชิงคอมพิวเตอร์ [50].....	21

รูปที่ 18 แบบรูปโมเดล-วิว-คอนโทรลเลอร์ [53].....	23
รูปที่ 19 ฐานข้อมูล STRidER [54].....	25
รูปที่ 20 ไฟล์ Excel-based workbook of STRait Razor 3.0	27
รูปที่ 21 แถบสีทั้งหมดของแอลลีมาตรฐานในชุดคิท AmpF/STR Identifier ที่ใช้สำหรับการเทียบ ขนาดของดีเอ็นเอ (ภาพที่ 1 ของ [13])	28
รูปที่ 22 ยูสเคสไดอะแกรมของระบบที่นำเสนอ	30
รูปที่ 23 ตัวอย่างข้อมูลและโครงสร้างของไฟล์เอสทีอาร์จากเทคนิคอิเล็กโทรโฟรีซิสหลอดเล็ก.....	32
รูปที่ 24 ตัวอย่างข้อมูลและโครงสร้างไฟล์ส่วนหัวของข้อมูลเอสทีอาร์ที่ได้จากการหาลำดับเบสยุคใหม่	32
รูปที่ 25 ตัวอย่างข้อมูลและโครงสร้างข้อมูลส่วนสรุปของข้อมูลเอสทีอาร์ในโลคัสต่าง ๆ	33
รูปที่ 26 ตัวอย่างข้อมูลและโครงสร้างข้อมูลส่วนแสดงลำดับเบสอย่างละเอียดของแต่ละแอลลีล และโลคัส	33
รูปที่ 27 อีอาร์ไดอะแกรมของข้อมูลเอสทีอาร์	34
รูปที่ 28 อีอาร์ไดอะแกรมของข้อมูลรายบุคคลและการสืบค้นแบบโปรไฟล์	35
รูปที่ 29 โครงสร้างไฟล์ที่ใช้สำหรับนำเข้าข้อมูลรายบุคคล	35
รูปที่ 30 อีอาร์ไดอะแกรมของข้อมูลโลคัสและชุดคิท	35
รูปที่ 31 อีอาร์ไดอะแกรมของข้อมูลฐานข้อมูลสาธารณะ โลคัสหลักและการค่าตัวแปรในระบบ	36
รูปที่ 32 อีอาร์ไดอะแกรมของข้อมูลของระบบจัดการผู้ใช้	37
รูปที่ 33 ไฟล์อ้างอิงเพื่อนำไปใช้วิเคราะห์เปรียบเทียบกับรูปแบบการซ้ำของข้อมูลลำดับเบสดีเอ็นเอ ของตัวอย่างที่ได้จากเทคโนโลยีเอ็นจีเอส	38
รูปที่ 34 โครงสร้างระบบ STRategy ภายใต้แนวคิดสถาปัตยกรรม 3-เลเยอร์.....	39
รูปที่ 35 การทำงานภาพรวมของระบบ	40
รูปที่ 36 แทปที่ใช้สำหรับการเลือกโลคัส.....	42
รูปที่ 37 ตัวอย่างการประยุกต์ใช้แบบรูปในระบบ	45
รูปที่ 38 ภาพรวมของระบบในรูปแบบ deployment diagram.....	48

รูปที่ 39 การสืบค้นข้อมูลเอสทีอาร์ (ก) สืบค้นข้อมูลเอสทีอาร์โดยใช้ไฟล์ตัวอย่าง (ข) สืบค้นข้อมูลเอสทีอาร์โดยใช้โลคัสและจีโนไทป์ (ค) สืบค้นข้อมูลเอสทีอาร์โดยใช้โลคัสและจีโนไทป์จากชุดคิท (ง) ผลลัพธ์จากการสืบค้นแบบรายตัวอย่าง (จ) ผลลัพธ์จากการสืบค้นแบบภาพรวม.....	50
รูปที่ 40 การสืบค้นแบบโปรไฟล์ (ก) ฟอรัมการสืบค้น (ข) ผลลัพธ์ภาพรวม	51
รูปที่ 41 ผลลัพธ์จากการสืบค้นแบบโปรไฟล์ที่ให้รายละเอียดในแต่ละประเทศ.....	51
รูปที่ 42 การแสดงผลความแปรปรวนทางพันธุกรรม (ก) รายการของโลคัส (ข) ค่าทางสถิติและกราฟ (ค) ความหลากหลายของแอลลีลของโลคัส DXS10074 และผลจากการเลือกแอลลีล 16 (ง) แทปของข้อมูลดีพลอยด์และแฮพลอยด์.....	52
รูปที่ 43 ตารางการแสดงผลข้อมูลแอลลีล.....	53
รูปที่ 44 การแสดงผลการกระจายของแอลลีลตามภูมิภาค	54
รูปที่ 45 การแสดงผลข้อมูลสรุปทางสถิติของข้อมูลสนิป	54
รูปที่ 46 การแสดงผลการเปรียบเทียบฐานข้อมูล STRidER และระบบที่ออกแบบและพัฒนา	55
รูปที่ 47 ตัวอย่างผลการวิเคราะห์เปรียบเทียบลำดับเบสดีเอ็นเอที่โลคัส FGA แอลลีล 24.2 ของชุดข้อมูลตัวอย่างทั้งหมดที่อยู่ในฐานข้อมูล.....	56
รูปที่ 48 หน้าเว็บเพื่อนำออกผลการวิเคราะห์เปรียบเทียบลำดับเบสดีเอ็นเอ	56
รูปที่ 49 ไฟล์ข้อมูลผลการวิเคราะห์เปรียบเทียบลำดับเบสดีเอ็นเอที่นำออกในรูปแบบไฟล์เอ็กเซล..	57
รูปที่ 50 หน้าเว็บอธิบายการใช้งานส่วนต่อประสานโปรแกรมประยุกต์สำหรับบริการภายนอก.....	59
รูปที่ 51 การประยุกต์ใช้แบบรูป Strategy เพื่อสนับสนุนการเพิ่มรูปแบบการสืบค้นข้อมูลแบบโปรไฟล์	60
รูปที่ 52 หน้าเว็บการนำเข้าข้อมูลที่ซ้ำซ้อนสู่ระบบ.....	61
รูปที่ 53 ไฟล์เอ็กเซลข้อมูลส่วนบุคคลของตัวอย่างในระบบ	61
รูปที่ 54 หน้าเว็บสำหรับการจัดการแผนที่.....	62
รูปที่ 55 หน้าเว็บการจัดการโลคัสและชุดคิท.....	62
รูปที่ 56 หน้าเว็บสำหรับการจัดการจังหวัด	63
รูปที่ 57 หน้าเว็บสำหรับการจัดการประเทศและภูมิภาค.....	64

รูปที่ 58 หน้าเว็บการจัดการโลโก้หลักของการสืบค้นแบบโปรไฟล์..... 64



บทที่ 1

บทนำ

1.1. ความเป็นมาและความสำคัญของปัญหา

นิติวิทยาศาสตร์คือการนำความรู้ทางวิทยาศาสตร์มาประยุกต์ใช้เพื่อประโยชน์ในการพิสูจน์ทางด้านกฎหมาย ตัวอย่างเช่น การพิสูจน์ความสัมพันธ์ทางเครือญาติ การตรวจพิสูจน์บุคคล กระบวนการทางนิติวิทยาศาสตร์ประกอบด้วยหลากหลายเทคนิคและวิธีการ เช่น การพิสูจน์ทางเครือญาติใช้การเปรียบเทียบลายพิมพ์ดีเอ็นเอ และในการเปรียบเทียบลายพิมพ์ดีเอ็นเอมีหลากหลายเทคนิคด้วยกัน หนึ่งในนั้นคือการพิจารณาลำดับซ้ำเรียงต่อเนื่องแบบสั้น (Short Tandem Repeats)

ลำดับซ้ำเรียงต่อเนื่องแบบสั้นหรือเอสทีอาร์ คือ ตำแหน่งบนสายดีเอ็นเอซึ่งมีรูปแบบการเรียงตัวของเบสซ้ำกันเป็นชุด (Unit) ตั้งแต่ชุดละ 2-6 เบส เช่น CACACACA เป็น Di-nucleotide ที่มีชุดเบส CA ซ้ำ 4 ชุด เป็นต้น จำนวนการซ้ำในแต่ละตำแหน่งแตกต่างกันไปแต่ละบุคคล และเอกลักษณ์นี้สามารถถ่ายทอดไปยังรุ่นลูกหลานได้ จึงสามารถใช้ประโยชน์จากลำดับซ้ำเรียงต่อเนื่องแบบสั้นในการระบุความสัมพันธ์ระหว่างเครือญาติได้ นอกจากนี้ยังมีงานวิจัย [1-3] ที่บ่งบอกว่าแอลลีลบางแอลลีลของลำดับซ้ำเรียงต่อเนื่องแบบสั้นมีความสัมพันธ์หรือมีความเกี่ยวข้องกับโรคบางชนิด นอกจากนี้ลำดับซ้ำเรียงต่อเนื่องแบบสั้นยังมีข้อมูลอีกหลายประเภทที่สามารถใช้ระบุหรือมีความเกี่ยวข้องกับโรค เช่น ภาวะพหุสัณฐานเดี่ยวของนิวคลีโอไทด์ (Single Nucleotide Polymorphism) หรือสนิป (SNP) โดยปกติแล้วดีเอ็นเอของมนุษย์มีลำดับเบสที่มีความคล้ายคลึงกันมาก โดยมีเพียงบางตำแหน่งที่แตกต่างกัน ถ้าในประชากรหนึ่งๆ พบตำแหน่งที่ถูกแทนที่ของเบสบริเวณเดียวกันบนสายดีเอ็นเอน้อยกว่า 1 เปอร์เซ็นต์ของจำนวนประชากรมักทำให้เกิดความผิดปกติจัดเป็นการกลาย (Mutation) แต่ถ้าการแทนที่นั้นมากกว่า 1 เปอร์เซ็นต์จะจัดว่าเป็นสนิป มีงานวิจัย [4, 5] ที่บ่งบอกว่าข้อมูลที่ได้จากสนิปนั้นมีความสัมพันธ์กับการก่อโรคบางชนิด รวมถึงใช้ในทางเภสัชพันธุศาสตร์ [6, 7]

ถึงแม้ว่าประโยชน์ของลำดับซ้ำเรียงต่อเนื่องแบบสั้นจะมีมากมาย แต่ก็ยังไม่มีแพลตฟอร์มซอฟต์แวร์ที่เป็นโอเพ่นซอร์สใดที่สามารถจัดการข้อมูล วิเคราะห์ข้อมูล รวมถึงการแสดงผลที่หลากหลายรวมไว้ในแพลตฟอร์มเดียว ในปัจจุบัน ผู้ใช้งานต้องรวบรวมผลลัพธ์ของข้อมูลเอสทีอาร์ลงในฐานข้อมูลที่แยกออกมาต่างหาก หรือไม่ก็เก็บในรูปแบบไฟล์และโฟลเดอร์ ซึ่งทำให้ยากต่อการ

จัดการหรือนำไปใช้งานต่อ นอกจากนี้สำหรับการวิเคราะห์และแสดงผลข้อมูลเอสทีอาร์ ผู้ใช้งานอาจต้องเรียนรู้เครื่องมือซอฟต์แวร์หลายตัว ซึ่งทำให้กระบวนการวิเคราะห์ที่ไม่มีประสิทธิภาพและยุ่งยาก เช่น ผู้ใช้งานต้องใช้ STRAF [8] สำหรับการคำนวณค่าทางสถิติของกลุ่มประชากร และต้องใช้ STRait Razor [9] และ Excel-based workbook of STRait Razor [10] ในการวิเคราะห์เพื่อเปรียบเทียบลำดับเบสดีเอ็นเอ รวมถึงต้องใช้ OmniPop200.1 [11] ในการสืบค้นแบบโปรไฟล์ เป็นต้น ทั้งนี้เครื่องมือบางชิ้นต้องมีการแปลงข้อมูลให้อยู่ในรูปแบบตามที่เครื่องมืออื่น ๆ กำหนดก่อนใช้งานอีกด้วย ซึ่งเพิ่มความซับซ้อนและเวลาสำหรับการใช้งานเครื่องมือเหล่านั้น

เพื่อแก้ไขปัญหาเหล่านี้ งานวิจัยนี้จึงออกแบบและพัฒนาระบบเพื่อสนับสนุนการจัดการเก็บและวิเคราะห์ข้อมูลเอสทีอาร์ ซึ่งสามารถช่วยให้ผู้ใช้งานสามารถจัดการและวิเคราะห์ข้อมูลเอสทีอาร์ได้อย่างมีประสิทธิภาพ ระบบดังกล่าวอนุญาตให้ผู้ใช้งานอัปโหลดข้อมูลเอสทีอาร์เข้าสู่ระบบ หลังจากนั้นระบบจะทำการเก็บข้อมูลเข้าสู่ฐานข้อมูลของระบบ รวมถึงมีการวิเคราะห์และแสดงผลข้อมูลในรูปแบบต่าง ๆ ให้อัตโนมัติ เช่น การคำนวณค่าสถิติของกลุ่มประชากร การแสดงผลความถี่แอลลิล และการแสดงผลของการวิเคราะห์เพื่อเปรียบเทียบลำดับเบสดีเอ็นเอ เป็นต้น นอกจากนี้ระบบอนุญาตให้ผู้ใช้งานสืบค้นและอัปโหลดข้อมูลเอสทีอาร์เพิ่มเติมในระบบ ซึ่งระบบที่นำเสนอถูกออกแบบมาให้ใช้ในองค์กรหรือห้องทดลอง จึงมีการใช้ระบบการกำหนดสิทธิ์ผู้ใช้ตามบทบาท (Role based access control) เพื่อกำหนดสิทธิ์ให้เหมาะสมกับผู้ใช้ในแต่ละบทบาท เช่น ผู้ใช้งานทั่วไปสามารถเข้าถึงข้อมูลภาพรวมได้ และแอดมินสามารถตั้งค่าระบบได้ เป็นต้น โดยผู้ใช้งานสามารถดาวน์โหลดและติดตั้งระบบที่นำเสนอได้โดยง่ายและได้หลากหลายแพลตฟอร์มผ่านดีออกเกอร์ (Docker) ซึ่งสามารถดาวน์โหลดระบบที่นำเสนอผ่าน github.com/cucpbioinfo/STRategy โดยสรุประบบที่นำเสนอทำหน้าที่เป็นฐานข้อมูลที่เก็บข้อมูลเอสทีอาร์พร้อมกับระบบจัดการข้อมูลที่ใช้งานในองค์กรหรือห้องทดลอง โดยมีเครื่องมือที่จำเป็นในการวิเคราะห์และการแสดงผลที่หลากหลาย รวมทั้งยังสามารถตั้งค่าต่าง ๆ โดยผู้ดูแลระบบ และพัฒนาต่อยอดโดยผู้พัฒนาได้โดยง่าย

1.2. วัตถุประสงค์ของงานวิจัย

วิจัยออกแบบและพัฒนาระบบสนับสนุนการจัดการเก็บและวิเคราะห์ข้อมูลลำดับซ้ำเรียงต่อเนื่องแบบสั้นสำหรับนิติวิทยาศาสตร์ โดยมีวัตถุประสงค์ดังต่อไปนี้

1. เพื่อสนับสนุนการบูรณาการข้อมูลที่ได้จากการกระบวนการทดลองทางนิติวิทยาศาสตร์

2. เพื่อสนับสนุนการวิเคราะห์และสืบค้นข้อมูลที่เป็นประโยชน์ทางนิติวิทยาศาสตร์
3. เพื่อสนับสนุนความยืดหยุ่นในการทำงานที่จำเพาะในแต่ละห้องปฏิบัติการ
4. เพื่อสนับสนุนการพัฒนาต่อยอดระบบได้โดยง่ายในอนาคต

1.3. ขอบเขตงานวิจัย

1. ระบบที่นำเสนอถูกออกแบบและพัฒนาในรูปแบบเว็บแอปพลิเคชันที่สามารถติดตั้งและใช้งานภายในหน่วยงานหรือองค์กร
 - ก. รองรับเบราว์เซอร์อย่างน้อย 2 ตัว คือ กูเกิลโครม (Google Chrome) และ ไมโครซอฟต์เอดจ์ (Microsoft Edge)
 - ข. ติดตั้งผ่านด็อกเกอร์ (Docker)
2. ระบบมีการควบคุมการเข้าถึงข้อมูลที่แตกต่างกันตามบทบาทโดยใช้หลักการ การกำหนดสิทธิผู้ใช้ระบบตามบทบาท (Role based access control)
 - ก. ผู้ใช้ทั่วไปไม่ต้องล็อกอินเข้าระบบ สามารถใช้งานฟังก์ชันค้นหาและเห็นข้อมูลระดับสถิติรวมเท่านั้น
 - ข. ผู้ใช้งานระดับห้องปฏิบัติการสามารถใช้งานฟังก์ชันได้เหมือนผู้ใช้ทั่วไปแต่สามารถเห็นข้อมูลระดับรายตัวอย่าง
 - ค. ผู้ดูแลระบบสามารถใช้งานได้ทุกฟังก์ชัน รวมถึงสามารถตั้งค่าระบบต่างๆ ได้
3. ระบบการจัดการฐานข้อมูลรองรับข้อมูลลำดับซ้ำเรียงต่อเนื่องแบบสั้นที่ได้จากเทคนิคการหาลำดับเบสยุคใหม่ (Next-generation sequencing) หรือเอ็นจีเอส (NGS) โดยเป็นผลลัพธ์ที่ได้จาก ForenSeq [12]
4. ผู้ใช้งานทั่วไปสามารถสืบค้นและดูข้อมูลภาพรวมของตัวอย่างในระบบ ดังนี้
 - ก. ดูข้อมูลสรุปเชิงสถิติตามภูมิภาค
 - ข. ดูข้อมูลสรุปเชิงสถิติตามโลคัส
 - ค. สืบค้นฐานข้อมูลโดยใช้ข้อมูลลำดับซ้ำเรียงซ้ำต่อเนื่องแบบสั้น
 - ง. การสืบค้นแบบโปรไฟล์ [11]
5. ผู้ใช้งานระดับห้องปฏิบัติการสามารถใช้งานระบบในลักษณะเดียวกับผู้ใช้ทั่วไปรวมทั้งสามารถใช้เครื่องมือสนับสนุนการวิเคราะห์ข้อมูลลำดับซ้ำเรียงต่อเนื่องแบบสั้น ดังนี้
 - ก. วิเคราะห์เปรียบเทียบลำดับเบสดีเอ็นเอกับแต่ละโลคัสในชุดของกลุ่มตัวอย่าง
 - ข. เพิ่ม-ลดข้อมูลตัวอย่างของข้อมูลลำดับซ้ำเรียงต่อเนื่องแบบสั้น

- ค. เพิ่ม-ลดข้อมูลรายละเอียดรายบุคคลของชุดข้อมูล เช่น เพศ ภูมิภาค เป็นต้น
6. ผู้ดูแลระบบสามารถใช้งานระบบในลักษณะเดียวกับผู้ใช้งานระดับห้องปฏิบัติการรวมทั้งสามารถปรับเปลี่ยนการตั้งค่าและตัวแปรในระบบสำหรับการใช้งานที่จำเพาะในแต่ละห้องปฏิบัติการดังนี้
- ก. เพิ่ม-ลดชุดคิของระบบหลัก [13, 14] และการสืบค้นแบบโปรไฟล์ [11]
 - ข. เพิ่ม-ลดข้อมูล ประเทศ ภูมิภาค และจังหวัด
 - ค. เพิ่ม-ลดรูปแบบทั่วไปที่ใช้ในการวิเคราะห์เปรียบเทียบลำดับเบสดีเอ็นเอกับแต่ละโลกัส
 - ง. เปลี่ยนแผนที่ในหน้าข้อมูลสรุปเชิงสถิติตามภูมิภาค ในรูปแบบไฟล์ GeoJSON [15]
7. ระบบสามารถเพิ่มส่วนการวิเคราะห์และแสดงผลโดยผู้พัฒนาอื่นได้โดยง่าย โดยแสดงตัวอย่างการพัฒนาส่วนการวิเคราะห์และแสดงผลเพิ่มเติม ดังนี้
- ก. เพิ่มอัลกอริทึมของการสืบค้นแบบโปรไฟล์
 - ข. เพิ่มหรือปรับเปลี่ยนส่วนของโลกัสหลักของการสืบค้นแบบโปรไฟล์
8. มีส่วนต่อประสานโปรแกรมประยุกต์ (Application Programming Interface) สำหรับบริการภายนอกสามารถเรียกใช้งานได้ ดังนี้
- ก. การการสืบค้นแบบโปรไฟล์
 - ข. ข้อมูลสรุปเชิงสถิติตามโลกัส
 - ค. ข้อมูลสรุปเชิงสถิติตามภูมิภาค

1.4. ขั้นตอนและแผนการดำเนินการวิจัย

1. ศึกษางานวิจัยที่เกี่ยวข้องและความเป็นไปได้ในการพัฒนาระบบ
2. ศึกษาลักษณะข้อมูลทางนิติวิทยาศาสตร์ ทฤษฎีทางนิติวิทยาศาสตร์และทฤษฎีทางวิศวกรรมซอฟต์แวร์สำหรับการออกแบบระบบ
3. ออกแบบ พัฒนา และทดสอบการทำงานของระบบ
4. ประเมินการใช้งานโดยผู้ใช้งานจากภาควิชานิติเวชศาสตร์ คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
5. ปรับปรุงระบบตามคำแนะนำจากภาควิชานิติเวชศาสตร์ คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
6. สอบโครงร่างวิทยานิพนธ์
7. ตีพิมพ์ผลงานวิจัย

8. สอบวิทยานิพนธ์

1.5. ประโยชน์ที่ได้รับ

1. ได้ระบบที่สนับสนุนการจัดการและบูรณาการข้อมูลประเภทต่างๆ ที่ได้จากการทดลองทางนิติวิทยาศาสตร์
2. ได้ระบบที่สนับสนุนการวิเคราะห์ข้อมูลและการแสดงผลข้อมูลที่มีความหลากหลาย และมีประโยชน์ทางนิติวิทยาศาสตร์
3. ได้เรียนรู้ขั้นตอนและกระบวนการในการตีพิมพ์งานวิจัย รวมถึงความรู้ด้านนิติวิทยาศาสตร์

1.6. การตีพิมพ์ในวารสารวิชาการนานาชาติ

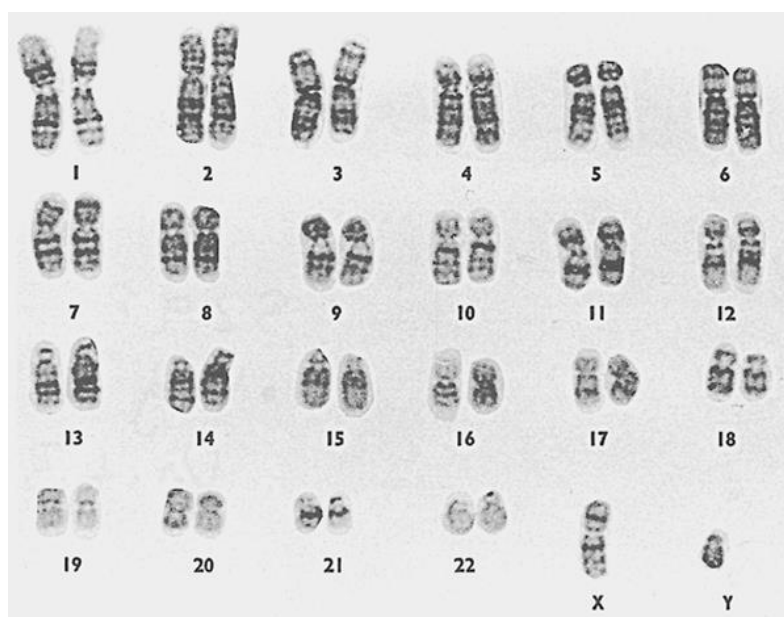
Nuttachai Kulthammanit, Tikumphorn Sathirapatya, Poonyapat Sukawutthiya, Hasnee Noh, Kornkiat Vongpaisarnsin, and Duangdao Wichadukul. (2023). STRategy: A support system for collecting and analyzing next-generation sequencing data of short tandem repeats for forensic science. PLOS One Journal on San Francisco, California, and Cambridge, United Kingdom.

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1. ทฤษฎีที่เกี่ยวข้อง

ทฤษฎีในหัวข้อที่ 2.2.1. จนถึง 2.2.12 เป็นทฤษฎีทางชีววิทยาที่ใช้สำหรับทำความเข้าใจการทำงานของระบบที่นำเสนอ ทฤษฎีในหัวข้อที่ 2.2.13. จนถึง 2.2.23. เป็นทฤษฎีทางคอมพิวเตอร์ที่ใช้ในการพัฒนาระบบที่นำเสนอให้มีประสิทธิภาพ



รูปที่ 1 โครงสร้างโครโมโซม [16]

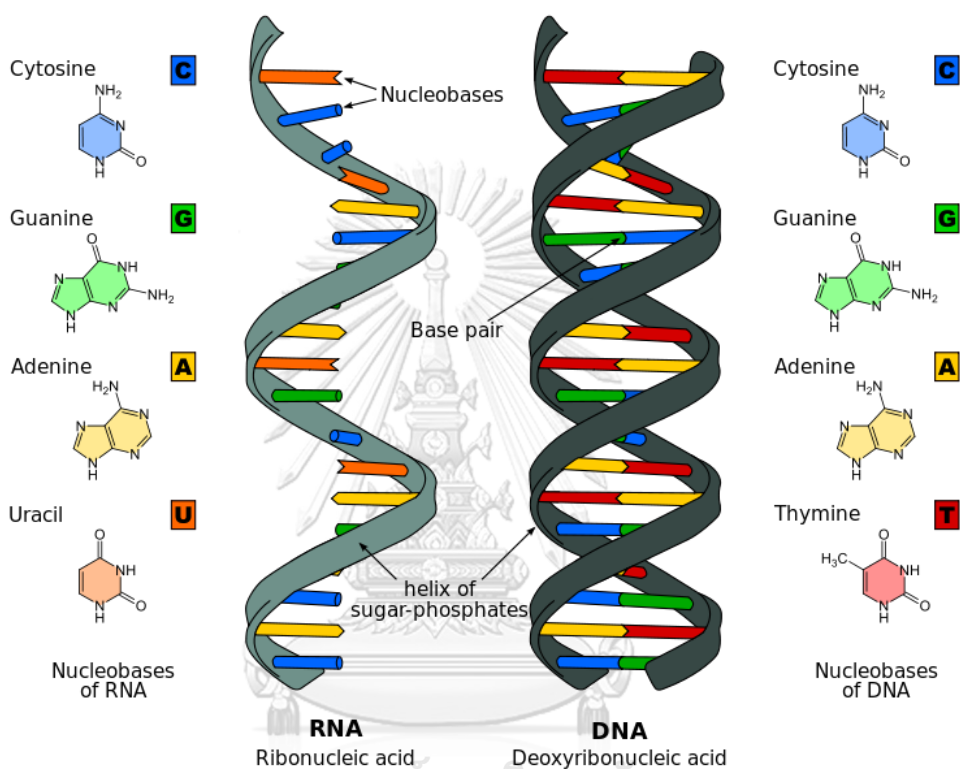
2.1.1. โครโมโซม (Chromosome)

นิวเคลียสของแต่ละเซลล์ประกอบด้วยหน่วยเก็บพันธุกรรม เรียกว่า โครโมโซม (Chromosome) [17] (รูปที่ 1) แต่ละโครโมโซมประกอบด้วยดีเอ็นเอ ที่ขดแน่นเป็นเกลียวรอบโปรตีนฮิสโตน (Histone) ในมนุษย์จะมีออโตโซม (Autosomes) 22 คู่ และ โครโมโซมเพศ (Sex chromosome) 1 คู่ รวมทั้งสิ้น 23 คู่ ออโตโซมหรือโครโมโซมร่างกาย เป็นโครโมโซมที่กำหนดลักษณะพันธุกรรมภายนอก ยกตัวอย่างเช่น สีผม สีตา เป็นต้น โครโมโซมเพศมี 2 ชนิด คือ โครโมโซมวาย (Y – chromosome) และ โครโมโซมเอ็กซ์ (X – chromosome) เพศชายจะมีโครโมโซมเอ็กซ์ และ โครโมโซมวายอย่างละหนึ่งแท่ง ส่วนเพศหญิงจะมีโครโมโซมเอ็กซ์สองแท่ง

2.1.2. ดีเอ็นเอ (DNA)

ดีเอ็นเอหรือชื่อเต็มคือ Deoxyribonucleic Acid [18] (รูปที่ 2) เป็นสารพันธุกรรมในมนุษย์และสิ่งมีชีวิตอื่นๆ ในกรณีของสิ่งมีชีวิตกลุ่มยูแคริโอต (Eukaryote) เช่น มนุษย์ แต่ละเซลล์มีดีเอ็นเอ

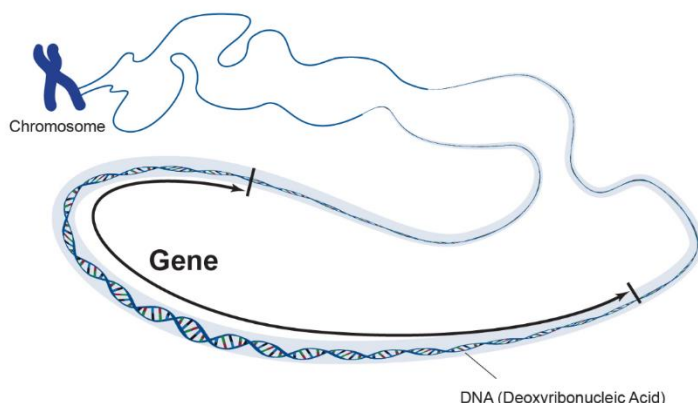
เดียวกันและดีเอ็นเออยู่ในนิวคลีอไซด์ของเซลล์ ดีเอ็นเอมีรูปร่างเป็นเกลียวคู่ ประกอบด้วยพอลินิวคลีโอไทด์ (Polynucleotide) 2 สายเรียงตัวในแนวที่ตรงข้ามกันพอดี โดยหันส่วนที่เป็นเบสเข้าหากันและส่วนที่เป็นน้ำตาลดีออกซีไรโบส (Deoxyribose sugar) อยู่ด้านนอก เบสในนิวคลีโอไทด์มี 4 ชนิด ได้แก่ อะดีนีน (Adenine; A) , ไทมีน (Thymine; T) , ไซโตซีน (Cytosine; C) และกวานีน (Guanine; G)



รูปที่ 2 ดีเอ็นเอ [19]

2.1.3. ยีน (Gene)

ยีน (Gene) [20] (รูปที่ 3) คือ ส่วนของดีเอ็นเอที่ถ่ายทอดทางพันธุกรรมได้ โดยแต่ละยีนอยู่บนตำแหน่งที่แน่นอนของโครโมโซมเรียกว่า โลคัส (Locus) ยีนมีหน้าที่ควบคุมและถ่ายทอดลักษณะทางพันธุกรรม โดยแต่ละโครโมโซมประกอบด้วยทั้งบริเวณที่เป็นยีนและบริเวณอื่น ๆ



รูปที่ 3 ยีนและโครโมโซม [21]

2.1.4. แอลลีล (Allele)

แอลลีล [22] คือรูปแบบของยีนที่ปรากฏบนโลคัสในลักษณะทางพันธุกรรมหนึ่งๆ ยีนอาจมีสองแอลลีล หรือมากกว่าสองแอลลีล เรียกว่า มัลติเปิลแอลลีล (Multiple alleles) [23] ตัวอย่างเช่น ยีนที่แสดงลักษณะกรุ๊ปเลือด มี 3 แอลลีล (รูปแบบ) คือ I^A , I^B , และ i แต่ละคนสามารถมีได้ 1 ใน 6 จีโนไทป์ (Genotype) [24] ได้แก่ ($I^A I^A$, $I^A i$, $I^B I^B$, $I^B i$, $I^A I^B$, ii) และสามารถแสดงออกมาได้ทั้งหมด 4 ฟิโนไทป์ (Phenotype) [24] ได้แก่ “กรุ๊ปเลือด A” ได้จากจีโนไทป์ ฮอมอไซกัส (Homozygous genotype) $I^A I^A$ และ จีโนไทป์เฮเทอโรไซกัส (Heterozygous genotype) $I^A i$ “กรุ๊ปเลือด B” ได้จากจีโนไทป์ฮอมอไซกัส $I^B I^B$ และ จีโนไทป์เฮเทอโรไซกัส $I^B i$ “กรุ๊ปเลือด AB” ได้จากจีโนไทป์เฮเทอโรไซกัส $I^A I^B$ และ “กรุ๊ปเลือด O” ได้จากจีโนไทป์ฮอมอไซกัส ii (รูปที่ 4)

จิว'	I^A	I^B	i
I^A	$I^A I^A$ A	$I^A I^B$ AB	$I^A i$ A
I^B	$I^B I^A$ AB	$I^B I^B$ B	$I^B i$ B
i	$i I^A$ A	$i I^B$ B	$i i$ O

Blood Type Alleles:

I^A (co-dominant)
I^B (co-dominant)
i (recessive)

phenotype (blood type)	genotype
Type A	$I^A I^A$ $I^A i$
Type B	$I^B I^B$ $I^B i$
Type AB	$I^A I^B$
Type O	$i i$

รูปที่ 4 แอลลีล จีโนไทป์ และฟิโนไทป์ของกรุ๊ปเลือด [25]

2.1.5. แฮพลอยด์ (Haploid) และ ดิพลอยด์ (Diploid)

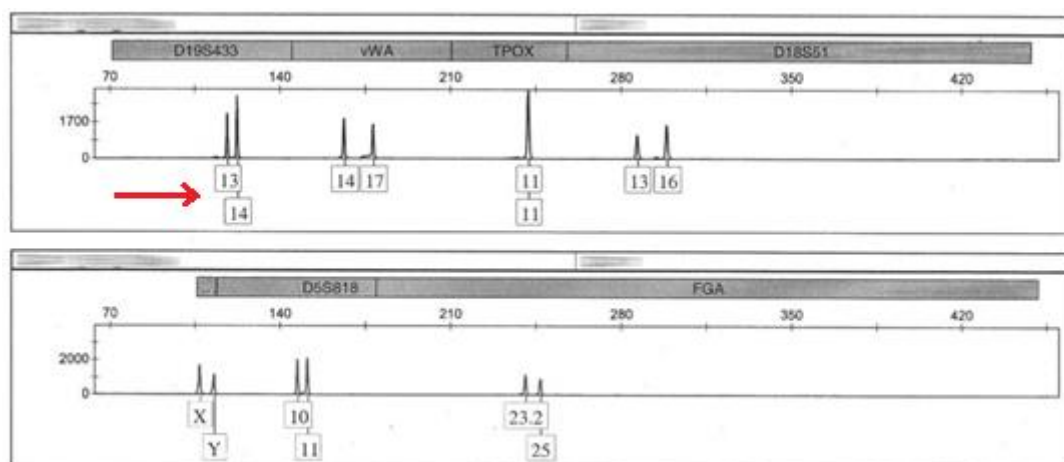
แฮพลอยด์ [26] คือเซลล์ที่ประกอบด้วยชุดของโครโมโซมเดี่ยวเท่านั้น ดิพลอยด์ [27] คือเซลล์ที่ประกอบด้วยชุดของโครโมโซมสองชุดได้จากพ่อและแม่อย่างละชุด สิ่งมีชีวิตที่สืบพันธุ์แบบอาศัยเพศ เช่น มนุษย์ ในเซลล์จะประกอบด้วยโครโมโซมทั้งหมด 2 ชุด เพราะฉะนั้นเซลล์ของมนุษย์จึงเป็นดิพลอยด์ อย่างไรก็ตามเซลล์ในไข่ของผู้หญิง และสเปิร์มของผู้ชายจะเป็นแฮพลอยด์ เพราะว่าในเซลล์มีแค่ชุดของโครโมโซมเพียงชุดเดียวเท่านั้น เมื่อเกิดการปฏิสนธิจึงกลายเป็นเซลล์แบบดิพลอยด์

2.1.6. การระบุบุคคลจากสารชีวโมเลกุล

ในปัจจุบันการระบุบุคคลจากสารชีวโมเลกุล (Identification from biomolecular evidence) [28] มีหลายวิธี หนึ่งในนั้นคือการตรวจดีเอ็นเอซึ่งเป็นวิธีที่มีความแม่นยำสูง มีความน่าเชื่อถือ และมีความจำเพาะต่อบุคคล ยกเว้นกรณีแฝดที่เกิดจากไข่ใบเดียวกัน โดยสามารถตรวจสอบได้จากสารชีววัตถุ เช่น เลือด ผม ผิวหนัง เป็นต้น โดยการเทียบจะต้องมีสารชีววัตถุที่มีการเก็บมาก่อนหน้า ปัจจุบันการตรวจพิสูจน์ดีเอ็นเอที่นิยมมีทั้งหมด 2 วิธีคือ ตรวจจากลำดับซ้ำเรียงต่อเนื่องแบบสั้น (Short tandem repeat) ของดีเอ็นเอในนิวเคลียสและตรวจดีเอ็นเอของไมโทคอนเดรีย (Mitochondrial DNA)

2.1.7. ลำดับซ้ำเรียงต่อเนื่องแบบสั้น (Short Tandem Repeats)

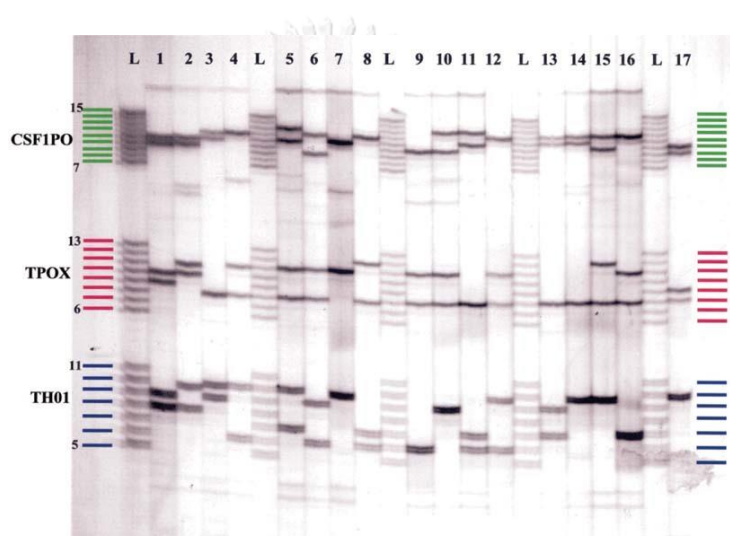
ลำดับซ้ำเรียงต่อเนื่องแบบสั้น (Short Tandem Repeats) หรือ เอสทีอาร์ (STR) [29] คือรูปแบบของเบสที่มีการเรียงซ้ำในหลายบริเวณ (โลคัส) ของโครโมโซม ซึ่งสามารถบอกอัตลักษณ์ของบุคคล เนื่องจากรูปแบบเอสทีอาร์ของแต่ละบุคคลมีความเป็นเอกลักษณ์ ดังแสดงในรูปที่ 5 ที่มีแถบแสดงโลคัสและแอลลีล เช่น โลคัสชื่อ D19S433 ของตัวอย่างนี้มีแอลลีลเป็น 13 และ 14 เป็นต้น



รูปที่ 5 ตัวอย่างลำดับซ้ำเรียงต่อเนื่องแบบสั้นจากเทคนิคอิเล็กโทรโฟรีซิสหลอดเล็ก [30]

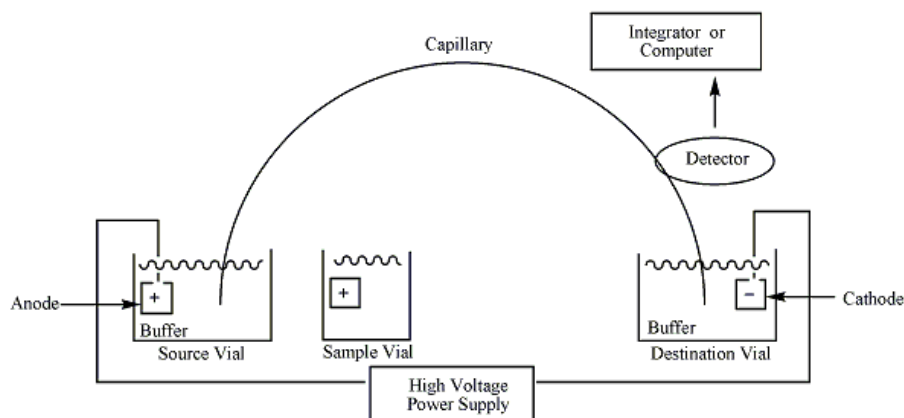
2.1.8. อิเล็กโทรโฟรีซิสหลอดเล็ก (Capillary Electrophoresis)

ห้องทดลองส่วนใหญ่ทั่วโลกในปัจจุบันใช้อิเล็กโทรโฟรีซิสหลอดเล็ก (Capillary Electrophoresis) [31] เป็นวิธีการสำหรับแยกเอสทีอาร์ การแยกเอสทีอาร์ที่มีความน่าเชื่อถือประกอบด้วย 3 เงื่อนไข คือ (1) มีความละเอียดเชิงพื้นที่ (Spatial resolution) สำหรับใช้ในการแยกแอลลีลของเอสทีอาร์ที่มีขนาดแตกต่างกันโดยนิวคลีโอไทด์ (2) มีความละเอียดเชิงสเปกตรัม (Spectral resolution) ที่แยกสีย้อมเรืองแสงออกจากกัน เพื่อให้พีซีอาร์ (PCR; Polymerase Chain Reaction) จากโลคัสที่ติดฉลากด้วยสีย้อมต่างกันสามารถแยกออกจากกันได้ (3) มีความแม่นยำของดีเอ็นเอที่ได้ สามารถนำมาเทียบกับแอลลีลมาตรฐาน (Allelic ladder) ที่ใช้ในการวัดขนาดได้ (รูปที่ 6)



รูปที่ 6 แถบเจลที่แสดงแอลลีลเทียบกับแอลลีลมาตรฐาน (Allelic ladder) ของแต่ละโลคัส (ภาพที่ 1 ของ [32])

เครื่องมือพื้นฐานที่ใช้ในการทำอิเล็กโทรโฟรีซิสหลอดเล็กประกอบด้วย หลอดแก้ว (Capillary tube) ขวดแก้วใส่สารตัวอย่าง (Buffer vials) สองขวด และอิเล็กโทรดสองขั้วที่เชื่อมต่อกับแหล่งจ่ายไฟแรงสูง และระบบอิเล็กโทรโฟรีซิสหลอดเล็ก เมื่อมีการเปิดเครื่องเพื่อให้ศักย์ไฟฟ้าสารจะไหลไปตามหลอดแก้วและผ่านเครื่องตรวจจับ (Detector) เพื่อตรวจจับตัวอย่าง (รูปที่ 7)



รูปที่ 7 อิเล็กโทรโฟรีซิสหลอดเล็ก [33]

Sample Name	Marker	Allele 1	Allele 2
58CUSP573-C1	D3S1358	15	17
58CUSP573-C1	TH01	9	
58CUSP573-C1	D21S11	28	29
58CUSP573-C1	D18S51	12	14
58CUSP573-C1	Penta E	5	11
58CUSP573-C1	D5S818	9	12
58CUSP573-C1	D13S317	8	13
58CUSP573-C1	D7S820	9	10
58CUSP573-C1	D16S539	10	12
58CUSP573-C1	CSF1PO	11	12
58CUSP573-C1	Penta D	7	14
58CUSP573-C1	AMEL	X	Y
58CUSP573-C1	vWA	14	
58CUSP573-C1	D8S1179	10	16
58CUSP573-C1	TPOX	8	
58CUSP573-C1	FGA	23	25
58CUSP573-C1	D19S433	13	14.2
58CUSP573-C1	D2S1338	20	24

รูปที่ 8 รูปแบบข้อมูลอิเล็กโทรโฟรีซิสหลอดเล็ก

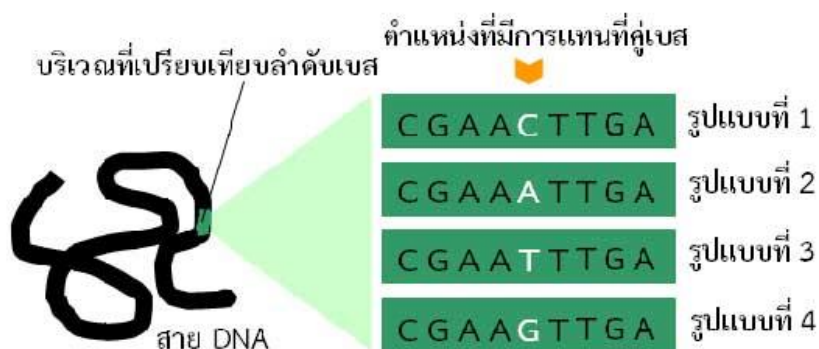
ตัวอย่างข้อมูลผลลัพธ์ (รูปที่ 8) ที่ได้มาจากเครื่องอิเล็กโทรโฟรีซิสหลอดเล็กของทางภาควิชานิติเวชศาสตร์ คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อยู่ในรูปแบบไฟล์ข้อความ โดยในไฟล์แต่ละคอลัมน์จะแบ่งด้วยคีย์แท็บ (tab) และประกอบด้วย 4 คอลัมน์ ดังต่อไปนี้

- Sample name – เป็นรหัสกำกับข้อมูล
- Marker – เป็นชื่อของโลคัส
- Allele 1 – แอลลีลแรก
- Allele 2 - แอลลีลที่สอง

2.1.9. การหาลำดับเบสยุคใหม่ (Next-Generation Sequencing)

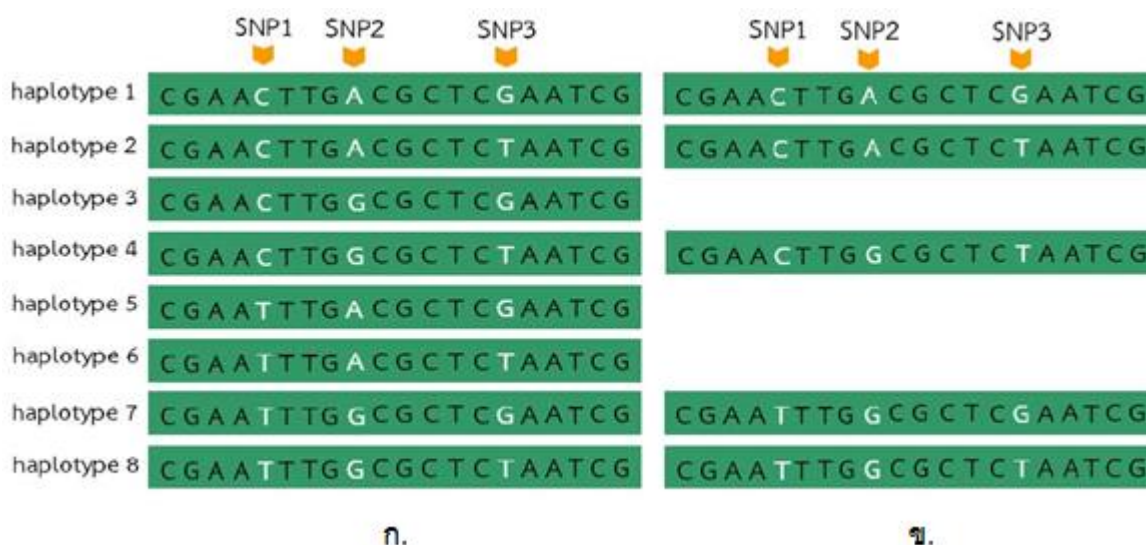
การหาลำดับนิวคลีโอไทด์ของดีเอ็นเอยุคแรกใช้การหาลำดับแบบแซงเกอร์ (Sanger sequencing) [34] แต่ข้อจำกัดของเทคนิคนี้ คือ ใช้เวลานานและไม่สามารถหาลำดับของนิวคลีโอ

การเรียงตัวของลำดับเบสบนดีเอ็นเอเหมือนกันแทบทั้งหมด แต่จะมีบางตำแหน่งที่มีลำดับเบสแตกต่างกัน ซึ่งเกิดจากการแทนที่ของเบสชนิดอื่น ดังแสดงในรูปที่ 10



รูปที่ 10 การแทนที่ของลำดับเบส (รูปที่ 1 ของ [36])

พหุสัณฐานเดี่ยวของนิวคลีโอไทด์ (Single Nucleotide Polymorphism) หรือสไนป (SNP) [37] ถือเป็นรูปแบบหนึ่งของภาวะพหุสัณฐานทางพันธุกรรม (Genetic polymorphism) โดยปกติแล้วสไนปที่เกิดขึ้นบนดีเอ็นเอจะพิจารณาจากหลายตำแหน่งร่วมกัน ทำให้สามารถเห็นภาวะพหุสัณฐานทางพันธุกรรมที่เกิดขึ้นในสิ่งมีชีวิตแต่ละชนิดได้ โดยตำแหน่งเหล่านี้เมื่อพิจารณาร่วมกันจะทำให้เกิดรูปแบบของลำดับนิวคลีโอไทด์บนสายดีเอ็นเอหลากหลายรูปแบบ เรียกว่าแฮพลอไทป์ [38] ตัวอย่าง เช่น มีสไนป 3 ตำแหน่งที่เกิดขึ้นใกล้ๆ กันบนสายดีเอ็นเอโดยที่ตำแหน่งแรกเกิดการแทนที่ระหว่าง C กับ T ตำแหน่งที่สองเกิดการแทนที่ระหว่าง A กับ G ตำแหน่งสุดท้ายเกิดการแทนที่ระหว่าง G กับ T ซึ่งทำให้สามารถเกิดแฮพลอไทป์ได้ทั้งหมด 8 แบบ (2^3) ดังแสดงในรูปที่ 11(ก) ลักษณะของแฮพลอไทป์จะกระจายอยู่ทั่วจีโนม [39] โดยแต่ละกลุ่มจะเรียกว่าแฮพลอไทป์บล็อก สไนปที่เกิดขึ้นในแฮพลอไทป์บล็อกเดียวกันมีแนวโน้มที่จะถูกถ่ายทอดไปด้วยกันเป็นชุดโดยไม่มีการเปลี่ยนแปลง ทั้งนี้แฮพลอไทป์ที่พบจากการสำรวจอาจมีเพียงไม่กี่รูปแบบ ตัวอย่าง เช่น สไนป 3 ตำแหน่งที่เกิดขึ้นในแฮพลอไทป์บล็อกเดียวกันในรูปที่ 11(ก) มีรูปแบบของแฮพลอไทป์ที่เป็นไปได้ทั้งหมด 8 แบบ แต่ในความเป็นจริงพบเพียง 5 แบบเท่านั้น ดังที่แสดงในรูปที่ 11(ข) แฮพลอไทป์อีก 3 รูปแบบ อาจจะไม่เกิดขึ้นในกลุ่มประชากรหรืออาจจะเกิดขึ้นน้อยมากทำให้ไม่สามารถถ่ายทอดมาสู่ลูกหลานได้

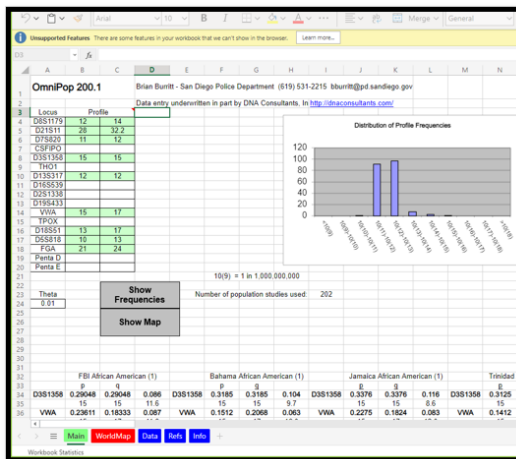


รูปที่ 11 รูปแบบแฮพลไทป์บล็อกที่เกิดจากสเนป 3 ตำแหน่ง (ก) รูปแบบทั้งหมดที่เป็นไปได้ (ข)
รูปแบบที่เกิดขึ้นจริง (รูปที่ 2 ของ [36])

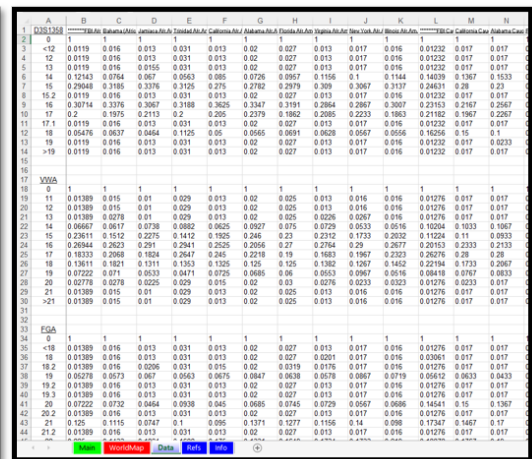
มีรายงานวิจัยว่าสเนปมีความสัมพันธ์กับการเกิดโรค [40] และการตอบสนองต่อยาที่ใช้รักษาโรค [41] จึงสามารถใช้ข้อมูลสเนปในการวินิจฉัยโรค การพัฒนายา [42] การเลือกใช้วิธีการรักษาและเลือกชนิดของยาให้เหมาะสมสำหรับแต่ละบุคคล นอกจากนี้สเนปยังเป็นอีกทางเลือกที่ถูกใช้ในนิติวิทยาศาสตร์ [43, 44]

2.1.11. การสืบค้นแบบโปรไฟล์

การสืบค้นแบบโปรไฟล์เป็นการหาค่าทางสถิติที่แสดงถึงความสัมพันธ์ระหว่างโปรไฟล์ (ชุดของโลคัสและแอลลีลของโลคัสเหล่านั้น) ที่ใช้ค้นหา และข้อมูลในฐานข้อมูลแต่ละประเทศ โปรแกรม OmniPop 200.1 เป็นโปรแกรมการสืบค้นแบบโปรไฟล์ที่สร้างขึ้นในรูปแบบไฟล์เอ็กเซล การค้นหาสามารถทำได้โดยการใส่ชุดของโลคัสและแอลลีลที่เกี่ยวข้องรูปที่ 12(ก) ให้กับโปรแกรม โดยโปรแกรมมีการกำหนดชุดแอลลีลและโลคัสที่ต้องใส่ขั้นต่ำเพื่อใช้ในการค้นหา เมื่อผู้ใช้ใส่ข้อมูลโปรไฟล์แล้ว โปรแกรมจะดึงค่าความถี่แอลลีล (Allele frequency) จากตารางข้อมูลในชีท Data มาใช้ในการคำนวณ ดังแสดงในรูปที่ 12(ข)

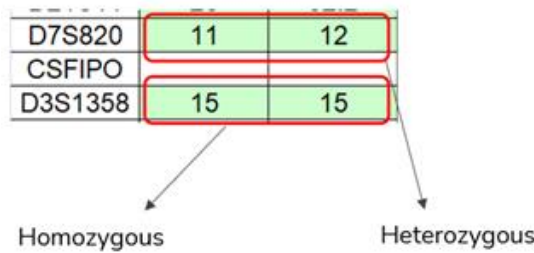


ก



ข

รูปที่ 12 โปรแกรม OmniPop 200.1 (ก) หน้าสำหรับใส่ข้อมูลเพื่อค้นหา (ข) ข้อมูลความถี่แอลลีล



	A	B	C	D	E	F	G
175	D3S1358						
176	>15	0.01397	0.015	0.01	0.03	0.013	0.02
177							
178	D7S820						
179	0	1	1	1	1	1	1
180	6	0.0119	0.015	0.01	0.03	0.013	0.02
181	6.3	0.0119	0.015	0.01	0.03	0.013	0.02
182	7	0.0119	0.015	0.01	0.03	0.013	0.02
183	8	0.17381	0.1512	0.1988	0.2083	0.2325	0.1774
184	8.1	0.0119	0.015	0.01	0.03	0.013	0.02
185	8.2	0.0119	0.015	0.01	0.03	0.013	0.02
186	9	0.15714	0.1296	0.1393	0.1131	0.1125	0.1331
187	9.1	0.0119	0.015	0.01	0.03	0.013	0.02
188	9.3	0.0119	0.015	0.01	0.03	0.013	0.02
189	10	0.32381	0.3364	0.3443	0.3333	0.34	0.2984
190	10.1	0.0119	0.015	0.01	0.03	0.013	0.02
191	10.3	0.0119	0.015	0.01	0.03	0.013	0.02
192	11	0.22381	0.2191	0.1844	0.2202	0.2175	0.2621
193	11.3	0.0119	0.015	0.01	0.03	0.013	0.02
194	12	0.09048	0.1235	0.1025	0.1012	0.0775	0.1089
195	13	0.01905	0.0247	0.0123	0.03	0.013	0.02
196	14	0.0119	0.015	0.01	0.03	0.013	0.02
197	>14	0.0119	0.015	0.01	0.03	0.013	0.02

รูปที่ 13 ตัวอย่างการคำนวณความถี่จีโนไทป์

โปรแกรม OmniPop นำค่าความถี่มาคำนวณโดยแบ่งออกเป็นสองกรณี (รูปที่ 13) กรณีแรกคือ ถ้าจีโนไทป์เป็นฮอโมไซกัส (Homozygous) จะได้ค่าความถี่จีโนไทป์ (Genotype frequency) ดังแสดงในสมการที่ (1)

$$\text{Genotype frequency} = x^2 + x \cdot (1 - x) \cdot \theta \quad (1)$$

ในกรณีที่จีโนไทป์เป็นเฮเทอโรไซกัส (Geterozygous) จะได้ค่าความถี่จีโนไทป์ (Genotype frequency) ดังแสดงในสมการที่ (2)

$$\text{Genotype frequency} = 2 \cdot x \cdot y \quad (2)$$

โดยที่ x, y คือความถี่แอลลีล และ θ คือค่าที่กำหนดโดยโปรแกรม ทั้งนี้ค่า x และ y ของโฮโมไซกัสและเฮเทอโรไซกัสจะเปลี่ยนไปตามแต่ละประเทศ หลังจากนั้น OmniPop จะนำค่าความถี่จีโนไทป์ดังกล่าวมาหาค่าส่วนกลับ และนำค่าส่วนกลับดังกล่าวในทุกๆ โลกส์มาคูณกัน และทำซ้ำแบบเดียวกันกับทุกประเทศ โดยโปรแกรมจะแสดงผลในรูปแบบ 1 ตารางต่อ 1 ประเทศดังในรูปที่ 14(ก) โดยผลลัพธ์ที่ได้จากการคูณค่าส่วนกลับของความถี่จีโนไทป์แสดงถึงความสัมพันธ์ระหว่างโปรไฟล์ตัวอย่างกับข้อมูลที่มีอยู่ในฐานข้อมูล โดยค่ายิ่งน้อยหมายถึงมีความสัมพันธ์กันมากดังแสดงในรูปที่ 14(ข)

FBI African American (1)				Bahama African American (1)			
	p	q		p	q		
D3S1358	0.29048	0.29048	0.086	D3S1358	0.3185	0.3185	0.104
	15	15	11.6		15	15	9.7
VWA	0.23611	0.18333	0.087	VWA	0.1512	0.2068	0.063
	15	17	11.6		15	17	16.0
FGA	0.125	0.18611	0.047	FGA	0.1115	0.2006	0.045
	21	24	21.5		21	24	22.4
D8S1179	0.10833	0.33333	0.072	D8S1179	0.1306	0.3376	0.088
	12	14	13.8		12	14	11.3
D21S11	0.21508	0.06983	0.030	D21S11	0.2197	0.0987	0.043
	28	32.2	33.3		28	32.2	23.1
D18S51	0.05556	0.16389	0.018	D18S51	0.051	0.1847	0.019
	13	17	54.9		13	17	53.1
D5S818	0.06389	0.24444	0.031	D5S818	0.0586	0.2253	0.026
	10	13	32.0		10	13	37.9
D13S317	0.48324	0.48324	0.236	D13S317	0.3951	0.3951	0.158
	12	12	4.2		12	12	6.3
D7S820	0.22381	0.09048	0.041	D7S820	0.2191	0.1235	0.054
	11	12	24.7		11	12	18.5
D16S539	1	1	1.000	D16S539	1	1	1.000
	0	0	1.0		0	0	1.0
TH01	1	1	1.000	TH01	1	1	1.000

Populations Sorted by Frequency (Most Common to Least Common)	
Chueta (Spain) (44)	3.20E+10
Bhumihar Brahmin (India) (72)	1.18E+11
Omani (28)	1.47E+11
Madeira Archipelago (53)	1.59E+11
Jewish (52)	1.97E+11
Mozabites (Algeria) (31)	2.00E+11
Bahama African American (1)	2.11E+11
African Spanish (29)	2.20E+11
CFS East Indian (Canada) (56)	2.22E+11
Alabama African American (2)	2.23E+11
VDFS Hispanic (4)	2.38E+11
Brahmin (India) (72)	2.38E+11
Virginia African American (2)	2.40E+11
FBI African American (1)	2.43E+11
Indiana African Americans (21)	2.45E+11

ก

ข

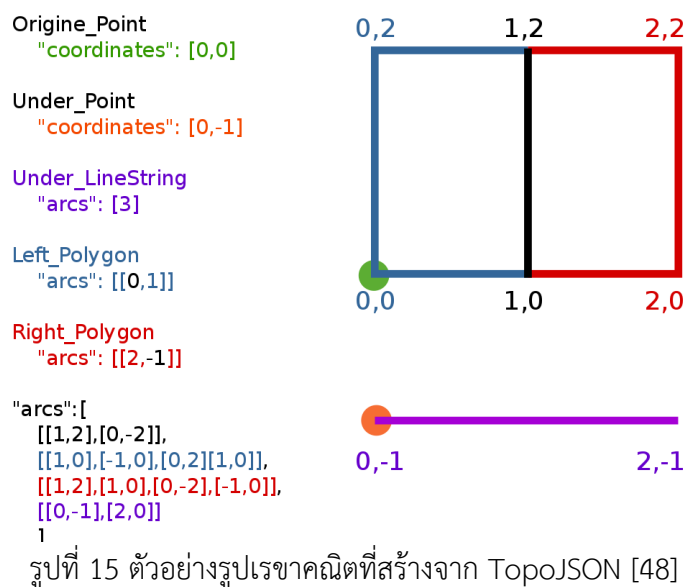
รูปที่ 14 ตัวอย่างผลลัพธ์ของการสืบค้นแบบโปรไฟล์ (ก) ผลลัพธ์ที่ได้จากการคูณค่าส่วนกลับของความถี่จีโนไทป์ของประเทศอเมริกา (ข) ค่าความสัมพันธ์ระหว่างโปรไฟล์ตัวอย่างกับข้อมูลที่มีอยู่ในฐานข้อมูล

2.1.12. กระบวนการรีเวิร์สคอมพลีเมนต์ (Reverse complement)

เป็นการแปลงลำดับนิวคลีโอไทด์ของสายดีเอ็นเอ โดยลำดับนิวคลีโอไทด์จะเรียงลำดับใหม่จากหลังไปหน้า เช่น ลำดับนิวคลีโอไทด์ GATC จะกลายเป็น CTAG จากนั้นแต่ละนิวคลีโอไทด์จะถูกเปลี่ยนเป็นนิวคลีโอไทด์ที่เป็นคู่จับ เช่น A เปลี่ยนเป็น T และ C เปลี่ยนเป็น G เป็นต้น ดังนั้น CTAG จะเปลี่ยนเป็น GATC ตัวอย่างของกระบวนการรีเวิร์สคอมพลีเมนต์ เช่น ลำดับเบส GGGGaaaaaaatttatat จะกลายเป็น atatataatttttttCCCC โดยที่ G ตัวแรกจะเปลี่ยนเป็น C และไปอยู่ท้ายสุด กระบวนการดังกล่าวจะสามารถแปลงสายดีเอ็นเอที่อ่านจากฝั่ง 3' กลายเป็นฝั่งที่ถูกอ่านจาก 5' หรือจากฝั่ง 5' กลายเป็นฝั่งที่อ่านจาก 3' ได้

2.1.13. GeoJSON

GeoJSON [45] เป็นรูปแบบมาตรฐานในการแสดงผลลักษณะทางภูมิศาสตร์ โดยเก็บเป็นโครงสร้างข้อมูลเชิงภูมิศาสตร์ ในรูปแบบ JSON [46] ซึ่งเป็นรูปแบบไฟล์พื้นฐาน เรขาคณิตที่ GeoJSON รองรับการใช้งานได้แก่ จุด (Point) เส้น (LineString) รูปหลายเหลี่ยม (Polygon) กลุ่มของจุด (Multipoint) กลุ่มของเส้น (MultiLineString) และ กลุ่มของรูปหลายเหลี่ยม (MultiPolygon) และมีอ็อบเจกต์ชื่อว่า Feature เป็นอ็อบเจกต์ทางเรขาคณิตที่มีคุณสมบัติบางอย่างเพิ่มเติมโดยกลุ่มของ Features จะถูกวางอยู่ในอ็อบเจกต์ชื่อ FeatureCollection นอกจากนี้ TopoJSON [47] (รูปที่ 15) ถูกพัฒนาจาก GeoJSON โดยมีคุณลักษณะเด่น คือ มีการเข้ารหัสตัว Topology เชิงพื้นที่ ซึ่งทำให้ขนาดไฟล์เล็กลง รูปที่ 15 แสดงตัวอย่างรูปเรขาคณิตที่สร้างจาก TopoJSON โดยแผนที่ที่แสดงอยู่ในหน้าเว็บ “การแสดงผลการกระจายของแอลลีลตามภูมิภาค” จะใช้ไฟล์ที่อยู่ในรูปแบบ TopoJSON และ GeoJSON



2.1.14. การเขียนโปรแกรมเชิงวัตถุ (Object Oriented Programming)

โปรแกรมเชิงวัตถุเป็นรูปแบบการเขียนโปรแกรมที่มองส่วนต่าง ๆ ของโปรแกรมเป็นวัตถุมาประกอบกัน โดยมีหลักการการเขียนโปรแกรมเชิงวัตถุประกอบด้วย

1. การห่อหุ้ม (Encapsulation)

การห่อหุ้มเป็นกระบวนการซ่อนรายละเอียดการทำงานและข้อมูลภายใน ไม่ให้สามารถมองเห็นจากภายนอกได้ กระบวนการดังกล่าวเรียกว่า การซ่อนข้อมูล (Information hiding) เนื่องจากภายนอกไม่สามารถเห็นสิ่งที่ถูกซ่อนไว้ได้ทำให้ไม่สามารถเปลี่ยนแปลง แก้ไข หรือสร้างความเสียหายให้กับสิ่งต่างๆ ที่อยู่ภายในได้

ตัวอย่างเช่น ถ้าโปรแกรมหนึ่งมีอ็อบเจกต์ (Object) หลายอ็อบเจกต์ แลกเปลี่ยนข้อมูลกัน การห่อหุ้มข้อมูลจะสมบูรณ์ก็ต่อเมื่อ อ็อบเจกต์แต่ละอ็อบเจกต์เก็บสถานะ (State) ไว้ในคลาส (Class) แบบส่วนตัว (Private) ไม่ให้ อ็อบเจกต์อื่นเข้าถึงได้โดยตรง แต่จะต้องเรียกผ่านฟังก์ชันหรือที่เรียกว่าเมธอด (Method) ที่คลาสต้นแบบของอ็อบเจกต์เตรียมไว้ให้ ในระบบที่นำเสนอใช้การห่อหุ้มโดยมีการประกาศเมธอดให้ใช้สำหรับการเข้าถึงข้อมูล ไม่ให้เข้าถึงข้อมูลโดยตรง

2. การกำหนดสาระสำคัญ (Abstraction)

การกำหนดสาระสำคัญ คือ การแสดงถึงคุณสมบัติและพฤติกรรมของอ็อบเจกต์เท่าที่จำเป็นต้องรับรู้และใช้งานเท่านั้น เรียกว่า การเปิดเผยกลไกระดับสูง (High-level mechanism) และซ่อนรายละเอียดการทำงานด้านในเอาไว้ ในการออกแบบโปรแกรมเชิงวัตถุ โปรแกรมส่วนใหญ่ประกอบด้วยอ็อบเจกต์จำนวนมากแลกเปลี่ยนข้อมูลกัน การบำรุงรักษาโค้ด (Code) ดังกล่าวจึงเป็นเรื่องยาก ประโยชน์ของหลักการกำหนดสาระสำคัญ คือช่วยในการลดการซ้ำซ้อนของโค้ด และเพิ่มการใช้ซ้ำ (Reusability) ของโค้ด ลดความซับซ้อนในโค้ด และช่วยเพิ่มความปลอดภัยให้กับระบบ เพราะผู้ใช้จะรู้แค่กลไกระดับสูงแต่ไม่รู้ถึงรายละเอียดการทำงาน ระบบที่นำเสนอมีการใช้หลักการกำหนดสาระสำคัญโดยใช้อินเทอร์เฟซในส่วนที่เป็นเลเยอร์ของวัตถุการเข้าถึงข้อมูล (Data Access Object) และเลเยอร์ของส่วนที่ใช้ประมวลผลข้อมูล

3. การสืบทอดคุณสมบัติ (Inheritance)

ในกรณีที่มีสิ่งหนึ่งมีลักษณะคล้ายกับอีกสิ่งหนึ่งมาก และมีส่วนที่แตกต่างกันเพียงเล็กน้อย เราไม่จำเป็นต้องสร้างสิ่งนั้นขึ้นมาใหม่ แต่สามารถนำหลักการการสืบทอดคุณสมบัติ (Inheritance) มาใช้งาน หลักการของการสืบทอดคุณสมบัติ คือการสืบทอดหรือรับเอา (Inherit) คุณสมบัติทั้งหมดมาจากสิ่งเดิมที่มีอยู่แล้ว และมีการเพิ่มเติมบางอย่างที่เป็นสิ่งใหม่เข้าไป ซึ่งจะช่วยให้ช่วยลดเวลาในการออกแบบและพัฒนาโปรแกรม ระบบที่นำเสนอมีการใช้หลักการการสืบทอดคุณสมบัติในส่วนของการจัดการกับข้อผิดพลาดในระบบ (Exception handling)

4. การพ้องรูป (Polymorphism)

หลักการของการพ้องรูป คือ ความสามารถของโปรแกรมในการประมวลผล อ็อบเจกต์ได้แตกต่างกันขึ้นอยู่กับประเภทของข้อมูลและคลาสที่ใช้ในการสร้างอ็อบเจกต์นั้น หรือคือการใช้โค้ดซ้ำ หลักการการพ้องรูปจะช่วยให้ผู้พัฒนาประหยัดเวลาในการพัฒนาโปรแกรม เช่น การสร้างคลาสแค่ครั้งเดียว แต่สามารถนำไปใช้ได้หลายๆ ที่ในโปรแกรม ช่วยทำให้ง่ายในการตรวจหาข้อบกพร่อง (Debug) ระบบที่นำเสนอมีการใช้

หลักการการพ้องรูปในส่วนการตั้งค่าความปลอดภัยของระบบที่ต้องมีการสืบทอดจากคลาสในเฟรมเวิร์กสปริงเพื่อตั้งค่าความปลอดภัยของระบบให้เหมาะสม

2.1.15. หลักการ SOLID (SOLID principle)

หลักการ SOLID [49] เป็นการออกแบบโปรแกรมเชิงวัตถุทำให้โปรแกรมมีความยืดหยุ่น เข้าใจง่าย และง่ายต่อการบำรุงรักษา ประกอบด้วยหลักการ 5 ข้อต่อไปนี้

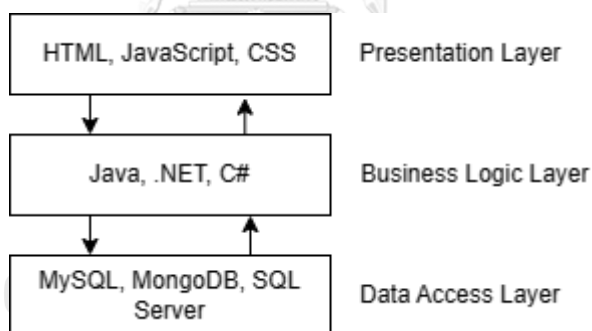
1. Single-responsibility principle หมายถึง คลาสหนึ่งๆ ควรมีหน้าที่เพียงอย่างเดียวเท่านั้น ระบบที่นำเสนอใช้หลักการ Single-responsibility principle โดยมีการแบ่งคลาสให้ทำหน้าที่เพียงอย่างเดียวยกตัวอย่าง เช่น คลาสที่เป็นเลเยอร์ของวัตถุการเข้าถึงข้อมูล (Data Access Object) จะทำหน้าที่ติดต่อกับฐานข้อมูลเท่านั้น
2. Open-closed principle หมายถึงคลาสหนึ่งๆ ควรง่ายต่อการเพิ่ม แต่ยากต่อการแก้ไข โดยผู้พัฒนาสามารถเพิ่มฟังก์ชันใหม่เข้าไปในโค้ดต้นฉบับ (Source code) ได้ง่ายโดยที่ไม่ต้องแก้ไขฟังก์ชันเดิมที่มีอยู่ ระบบที่นำเสนอใช้หลักการ Open-closed principle โดยมีการเรียกใช้งานผ่านอินเทอร์เฟซ เช่น คลาสในเลเยอร์ของการประมวลผลข้อมูลมีการใช้อินเทอร์เฟซของคลาสที่เป็นเลเยอร์ของวัตถุการเข้าถึงข้อมูล ซึ่งในอนาคตสามารถเพิ่มคลาสใหม่ได้ โดยที่คลาสในเลเยอร์ของการประมวลผลข้อมูลสามารถเรียกใช้งานคลาสใหม่นี้ได้โดยง่าย
3. Liskov substitution principle หมายถึงประเภท (Type) ของคลาสย่อยสามารถแทนที่ประเภทของคลาสแม่ได้โดยไม่ทำให้โปรแกรมเกิดความผิดพลาด ระบบที่นำเสนอมีการใช้หลักการนี้ เช่น คลาสที่เป็นคอนโทรลเลอร์เรียกใช้คลาสที่เป็นส่วนการประมวลผลข้อมูลผ่านอินเทอร์เฟซ ซึ่งสามารถถูกแทนที่ด้วยคลาสลูกได้
4. Interface segregation principle หมายถึงคลาสย่อย ต้องไม่ถูกบังคับให้จัดเตรียมหรือใช้ฟังก์ชันที่ไม่ต้องการใช้ หลักการนี้ช่วยระบบให้มีความแยกออกจากกัน (Decoupled) และง่ายต่อการปรับปรุงโครงสร้าง (Refactor) เปลี่ยนแปลง (Change) และ ปรับใช้ใหม่ (Redeploy) ระบบที่นำเสนอมีการใช้หลักการ Interface segregation principle เช่น อินเทอร์เฟซที่ใช้ในส่วนของการประมวลผลข้อมูลจะมีเฉพาะเมธอดที่จำเป็นในการใช้งานของส่วนนี้เท่านั้น
5. Dependency inversion principle หลักการมีทั้งหมดสองข้อ หนึ่ง คือ โมดูลระดับสูง (High-level module) ไม่ควรแปรตาม (Depend on) โมดูลระดับล่าง (Low-level module) แต่ควรแปรตามการกำหนดสาระสำคัญ (Abstractions) สอง คือ การกำหนดสาระสำคัญไม่ควรแปรตามข้อปลีกย่อย (Detail) แต่ข้อปลีกย่อยควรแปรตามการ

กำหนดสาระสำคัญ ระบบที่นำเสนอมีใช้หลักการ Dependency inversion principle เช่น คลาสที่เป็นเลเยอร์ของการประมวลผลข้อมูลเรียกใช้งานคลาสที่เป็นเลเยอร์ของวัตถุ การเข้าถึงข้อมูลผ่านอินเทอร์เฟซ

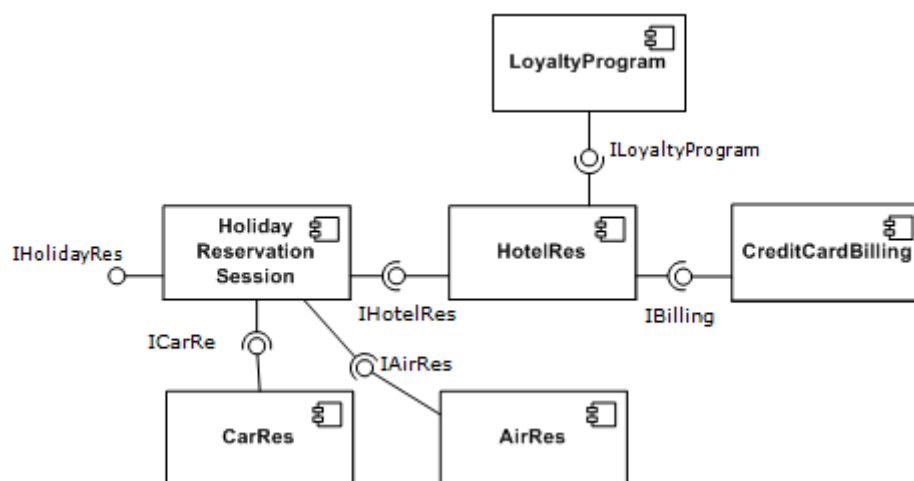
2.1.16. สถาปัตยกรรม 3-เลเยอร์ (3-layer architecture)

สถาปัตยกรรม 3-เลเยอร์ (รูปที่ 16) เป็นสถาปัตยกรรมซอฟต์แวร์ ประกอบด้วย

1. ชั้นการแสดงผล (Presentation layer) หรือส่วนต่อประสานกับผู้ใช้ (User Interface: UI) โดยทำหน้าที่แสดงสิ่งต่างๆ ให้ผู้ใช้เห็น ส่วนใหญ่สร้างด้วยเทคโนโลยี เช่น HTML5, JavaScript, CSS หรือผ่านเฟรมเวิร์ค (Framework)
2. ชั้นประมวลผลข้อมูล (Business logic layer) ประกอบด้วยตรรกะเชิงการค้า ที่เป็นตัวขับเคลื่อนหลักของแอปพลิเคชัน ส่วนใหญ่เขียนด้วย Java, .Net, C#, Python, JavaScript, Go และอื่น ๆ
3. ชั้นของการเข้าถึงข้อมูล (Data access layer) ตัวอย่างของฐานข้อมูลในชั้นนี้ เช่น MySQL, Oracle, PostgreSQL, Microsoft SQL Server, MongoDB และตัวอย่างของชั้นการเข้าถึงข้อมูล เช่น ORM และ JPA เป็นต้น



รูปที่ 16 สถาปัตยกรรม 3-เลเยอร์



รูปที่ 17 สถาปัตยกรรมเชิงคอมโพเนนท์ [50]

2.1.17. สถาปัตยกรรมเชิงคอมโพเนนท์ (Component-based architecture)

สถาปัตยกรรมเชิงคอมโพเนนท์ (รูปที่ 17) คือการแบ่งส่วนของระบบออกเป็นคอมโพเนนท์ เพื่อเน้นการนำคอมโพเนนท์กลับมาใช้ซ้ำ (Component reusability) สถาปัตยกรรมดังกล่าวมีประโยชน์หลายประการ เช่น ลดเวลา ลดค่าใช้จ่าย และเพิ่มความน่าเชื่อถือในการพัฒนาโปรแกรม

2.1.18. แบบรูปการออกแบบ (Design patterns)

ในทางวิศวกรรมซอฟต์แวร์ แบบรูปการออกแบบ [51] เป็นแบบรูปที่ใช้แก้ไขปัญหาที่เกิดขึ้นซ้ำๆ ในการออกแบบซอฟต์แวร์ แบบรูปการออกแบบไม่ได้เป็นโค้ดต้นฉบับ (Source code) สำเร็จรูป แต่เป็นเพียงแม่แบบ (Template) ที่ใช้ในการแก้ไขปัญหาในสถานการณ์ต่าง ๆ

แบบรูปของการออกแบบสามารถแบ่งออกเป็น 3 ประเภทใหญ่ ๆ ดังนี้

1. แบบรูปการออกแบบเชิงการสร้าง (Creational design patterns)

แบบรูปการออกแบบนี้เกี่ยวกับการสร้างอินสแตนซ์ (Instance) ของคลาส โดยแบบรูปการออกแบบนี้สามารถแบ่งย่อยได้เป็น แบบรูปการสร้างคลาส (Class-creation patterns) และแบบรูปการสร้างอ็อบเจกต์ (Object-creation patterns) โดยแบบรูปการสร้างคลาสใช้การสืบทอดในกระบวนการสร้างอินสแตนซ์ ส่วนแบบรูปการสร้างอ็อบเจกต์ใช้ในการทำให้งานหนึ่งๆ สำเร็จ ได้แก่

- Builder เป็นแบบรูปที่ช่วยสร้างอ็อบเจกต์ที่มีประเภทและการใช้งานต่างกันได้ โดยใช้โค้ดต้นฉบับเดียวกัน
- Factory method เป็นแบบรูปที่มีโค้ดใช้สำหรับสร้างอ็อบเจกต์ในคลาส แต่อ่อนุญาตให้ถ่ายทอดไปยังคลาสลูกและสามารถแก้ไขประเภทของอ็อบเจกต์ที่จะถูกสร้างได้

- Singleton เป็นแบบรูปที่มีการสร้างอินสแตนซ์หรืออ็อบเจกต์เพียงอินสแตนซ์หรืออ็อบเจกต์เดียวเท่านั้น

2. แบบรูปการออกแบบเชิงโครงสร้าง (Structural design patterns)

แบบรูปการออกแบบนี้เกี่ยวกับการรวมเข้ากันของคลาสและอ็อบเจกต์ในโครงสร้างขนาดใหญ่ แต่ยังคงคุณสมบัติที่ยืดหยุ่น (Flexible) และมีประสิทธิภาพ (Efficient)

- Private Class Data เป็นแบบรูปที่จำกัดการเข้าถึงและการเปลี่ยนโดยตรงของคุณลักษณะ (Attributes)
- Proxy เป็นแบบรูปที่ทำหน้าที่แทนอ็อบเจกต์ตัวจริง ผู้ใช้สามารถติดต่ออ็อบเจกต์จริงผ่านอ็อบเจกต์พร็อกซีที่ทำหน้าที่ควบคุมอ็อบเจกต์จริงอีกที

3. แบบรูปการออกแบบเชิงพฤติกรรม (Behavioral design patterns)

เป็นแบบรูปของการติดต่อสื่อสารระหว่างอ็อบเจกต์

- State เป็นแบบรูปที่พฤติกรรมของอ็อบเจกต์ขึ้นอยู่กับสถานะของอ็อบเจกต์
- Template method เป็นแบบรูปที่กำหนดรูปแบบของอัลกอริทึมสำหรับการทำงานบางอย่างไว้ และคลาสย่อย (Subclasses) สามารถกำหนดการทำงานบางขั้นตอนได้โดยไม่ต้องไม่เปลี่ยนโครงสร้างเดิม

2.1.19. หลักการอินเวอร์ชันออฟคอนโทรล (Inversion of control principle)

หลักการอินเวอร์ชันออฟคอนโทรล ซึ่งบางครั้งถูกเรียกว่าเป็นแบบรูป (Pattern) กลับทิศของการควบคุมในการเขียนโปรแกรมเชิงวัตถุ เพื่อลดความขึ้นต่อกันระหว่างคลาส (Loose coupling) ตัวอย่างที่มีการใช้หลักการอินเวอร์ชันออฟคอนโทรล ได้แก่ ดีเพนเดนซีอินเจกชัน (Dependency injection) แบบรูปเทมเพลตเมธอด (Template method pattern) แบบรูปแฟกทอรีอ็อบเจกต์ (Abstract factory pattern) และแบบรูปแฟกทอรี (Factory pattern)

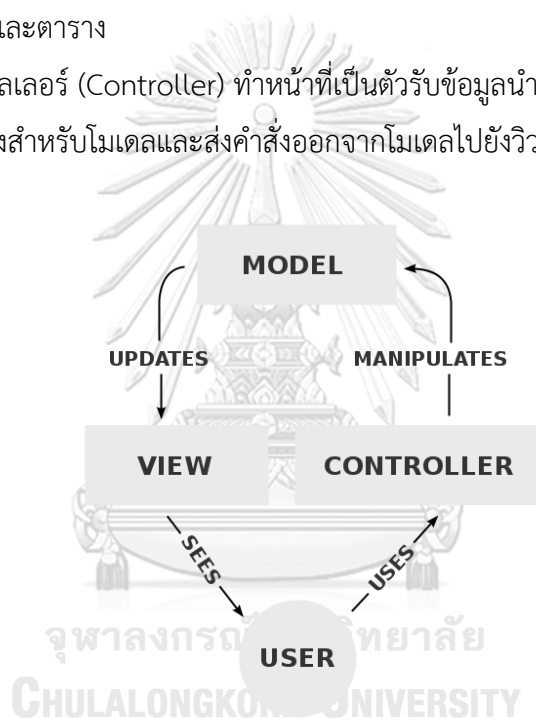
2.1.20. หลักการดีเพนเดนซีอินเจกชัน (Dependency Injection principle)

หลักการดีเพนเดนซีอินเจกชัน [49] เป็นเทคนิคที่ใช้หลักการอินเวอร์ชันออฟคอนโทรลเพื่อลดความขึ้นต่อกันระหว่างคลาส โดยเทคนิคนี้จะแยกการสร้างอ็อบเจกต์ออกมาจากคลาสที่มีความขึ้นต่อกัน ยกตัวอย่างเช่น คลาส A ต้องการใช้คลาส B คลาส A สามารถสร้างอ็อบเจกต์จากคลาส B ภายในคลาส A ได้ แต่รูปแบบดังกล่าวจะเพิ่มการขึ้นต่อกันระหว่างคลาส A และคลาส B เมื่อนำหลักการดีเพนเดนซีอินเจกชันมาใช้คลาส A ไม่จำเป็นต้องสร้างอ็อบเจกต์ของคลาส B ภายในคลาส แต่ให้คลาส Injector เป็นคนจัดการเรื่องการสร้างอ็อบเจกต์คลาส B และส่งอ็อบเจกต์ดังกล่าวมาให้คลาส A ใช้ เพื่อลดความขึ้นต่อกันระหว่างคลาส A และคลาส B

2.1.21. แบบรูปโมเดล-วิว-คอนโทรลเลอร์ (Model-view-controller design pattern)

แบบรูปโมเดล-วิว-คอนโทรลเลอร์ [52] (รูปที่ 18) เป็นแบบรูปที่ถูกใช้มากในการสร้างส่วนต่อประสานกับผู้ใช้ (User Interface) เพื่อแยกข้อมูลออกจากที่มาของข้อมูล และสิ่งที่ผู้ใช้เห็น โดยแบ่งตรรกะออกเป็นสามส่วนประกอบด้วย

1. โมเดล (Model) ทำหน้าที่จัดการข้อมูล ตรรกะ และกฎของแอปพลิเคชันโดยตรง โดยรับข้อมูลนำเข้า (Input) จากผู้ใช้ (User) ผ่านทางคอนโทรลเลอร์
2. วิว (View) ทำหน้าที่เป็นส่วนแสดงผลข้อมูล เช่น แผนภูมิ ตาราง หรือไดอะแกรม และในข้อมูลตัวเดียวกันอาจมีได้หลายวิว เช่น อาจมีข้อมูลเกี่ยวกับบัญชีที่แสดงเป็นทั้งแผนภูมิและตาราง
3. คอนโทรลเลอร์ (Controller) ทำหน้าที่เป็นตัวรับข้อมูลนำเข้า (Input) จากผู้ใช้ แปลงให้เป็นคำสั่งสำหรับโมเดลและส่งคำสั่งออกจากโมเดลไปยังวิวเพื่อทำการแสดงผล



รูปที่ 18 แบบรูปโมเดล-วิว-คอนโทรลเลอร์ [53]

2.1.22. Object Relational Mapping (ORM)

Object Relational Mapping (ORM) เป็นเทคนิคในการเขียนโปรแกรมที่จะช่วยแปลงข้อมูลระหว่างข้อมูลในฐานข้อมูลและอ็อบเจกต์ในภาษาโปรแกรมที่เป็นเชิงวัตถุ ระบบที่นำเสนอใช้เฟรมเวิร์ก Hibernate ที่มีการนำเทคนิคนี้มาใช้ ประโยชน์ของ ORM มีดังนี้

1. การเขียนโค้ดใช้อ็อบเจกต์ในการเข้าถึงข้อมูลแทนที่การใช้คำสั่งเพื่อเข้าถึงตารางในฐานข้อมูลโดยตรง
2. โค้ดของโปรแกรม จะไม่สามารถเห็น SQL queries ได้โดยตรง

3. เอนทิตี (Entities) ที่เป็นตัวแทนชุดข้อมูลในฐานข้อมูล ถูกพิจารณาจากมุมมองของฝั่งธุรกิจ (Business) มากกว่าโครงสร้างของฐานข้อมูล
4. ในการจัดการทรานแซคชัน (Transaction) จะมีการสร้างคีย์ขึ้นโดยอัตโนมัติ
5. เพิ่มความเร็วในการพัฒนาโปรแกรมเนื่องจากการเขียนโค้ดซับซ้อนน้อยลง สามารถจดจ่อกับฟีเจอร์ (Feature) ได้มากขึ้น

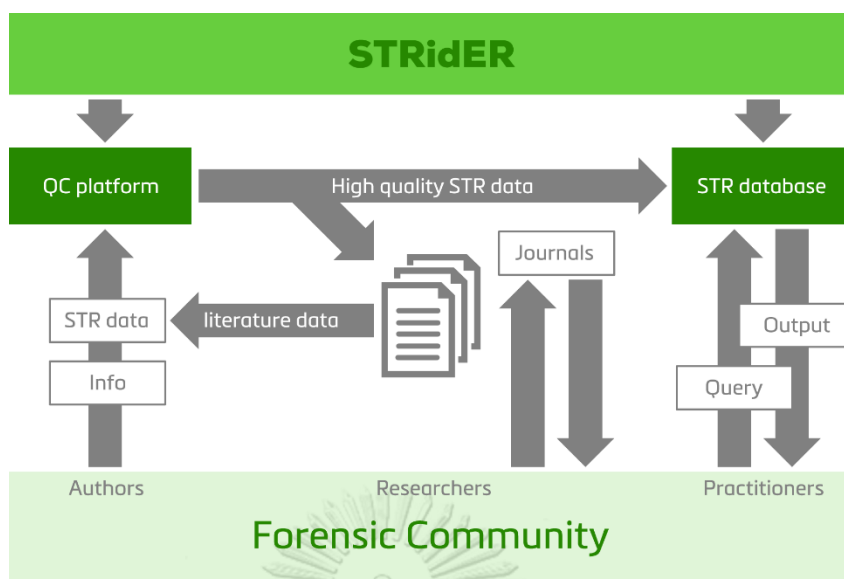
2.1.23. Java Persistence API (JPA)

Java Persistence API (JPA) เป็นมาตรฐานของภาษาจาวาที่ใช้ในการแปลงระหว่างอ็อบเจกต์ของภาษาจาวาและข้อมูลในฐานข้อมูล เพื่อลดภาระการเขียนโค้ดที่ใช้ในการจัดการความสัมพันธ์ระหว่างอ็อบเจกต์ โปรแกรมเมอร์ที่ใช้ JPA provider framework (เป็น implementation ของ JPA) สามารถใช้งานอินสแตนซ์ของฐานข้อมูลได้โดยง่าย ซึ่ง JPA เป็นมาตรฐานที่ใช้ในการพัฒนา ORM โดยที่ผู้พัฒนาสามารถแมพ (Map) เก็บ (Store) หรืออัปเดต (Update) ข้อมูล ไปยังฐานข้อมูลได้ และสามารถดึงข้อมูลจากฐานข้อมูลมาเป็นอ็อบเจกต์ของจาวา โดยผ่าน JPA ได้ ในระบบที่แนะนำให้ใช้ภาษาจาวาและใช้เฟรมเวิร์คชื่อ Hibernate ที่เป็น Object Relational Mapping ที่ถูกสร้างโดยใช้มาตรฐาน JPA

2.2. งานวิจัยที่เกี่ยวข้อง

2.2.1. ฐานข้อมูล STRidER (STRs for Identity ENFSI Reference Database)

ฐานข้อมูล STRidER [54] (รูปที่ 19) เป็นระบบฐานข้อมูล ที่ถูกพัฒนาต่อจาก ENFSI STRbase (2004-2016) ฐานข้อมูล STRidER ถูกพัฒนาโดย Institute of Legal Medicine, Medical University of Innsbruck เพื่อทำการคัดกรองข้อมูลเอสทีอาร์เชิงประชากร และเพิ่มคุณภาพของชุดข้อมูลเพื่อให้มีความน่าเชื่อถือสำหรับการประมาณความถี่ของแอลลีล (Allele) STRidER [55] ทำหน้าที่เป็นฐานข้อมูลเชิงความถี่และซอฟต์แวร์แพลตฟอร์ม สำหรับการพัฒนาเครื่องมือใหม่ๆ ควบคุมคุณภาพของข้อมูลเอสทีอาร์และเพิ่มเติมการวิเคราะห์ทางนิติวิทยาศาสตร์แบบอื่นๆ โดยเครื่องมือที่ทาง STRidER จัดเตรียมไว้ เช่น คำนวณค่าความถี่แอลลีล (Allele frequency) ชนิดของชุดคิทที่ใช้ในการทดลองได้แก่ SGMplus, ESSplex, Identifier, ESSplex SE, Powerplex 16, Powerplex 18, Powerplex 21, NGM, Fusion, NGM-SE, ESI-16, ESI-17 และ Globalfiler เป็นต้น



รูปที่ 19 ฐานข้อมูล STRidER [54]

2.2.2. NIST Standard Reference Database SRD 130

NIST Standard Reference Database SRD 130 [56] เป็นระบบฐานข้อมูลทางนิติวิทยาศาสตร์เกี่ยวกับดีเอ็นเอที่ใช้ยืนยันตัวบุคคลมากกว่า 20 ปี โดยมีข้อมูลเป็นประโยชน์ในการวิจัยและทดสอบดีเอ็นเอเพื่อระบุตัวบุคคล เช่น ข้อมูลพื้นฐานเกี่ยวกับเอสทีอาร์ การวิเคราะห์ความสัมพันธ์ทางสายโลหิตในรูปแบบต่างๆ (Kinship analysis) และรายงานความแตกต่างของแอลลีล (Variant allele) เป็นต้น ฐานข้อมูลนี้จัดทำโดย จอห์น เอ็ม บัตเลอร์ (John M. Butler) และ เดนนิส เจ ริดเตอร์ (Dennis J. Reeder) โดยมีความช่วยเหลือจาก เจน เรดแมน (Jan Redman) คริสเตียน ลุยเบิร์ต (Christian Ruitberg) และ ไมเคิล ถัง (Michael Tung)

2.2.3. US Y-STR Database (YHRD)

US Y-STR Database [57] เป็นระบบฐานข้อมูลที่สามารถสืบค้นโดยการระบุโพลีคัสและจำนวนการซ้ำของดีเอ็นเอแต่ละแอลลีล ผู้ใช้งานสามารถใช้ฐานข้อมูล US Y-STR ในการสืบค้นแฮพลไทป์ ของ Y-STR ในรูปแบบต่างๆ ภายในฐานข้อมูลระดับชาติและข้อมูลกลุ่มประชากรที่ระบุ สามารถเรียกดูข้อมูล เช่น ลักษณะพิเศษของเอสทีอาร์ การกลายพันธุ์ การวิเคราะห์ทางพันธุกรรมของประชากรรายโพลีคัส YSTR สร้างการประมาณความถี่แฮพลไทป์ที่เชื่อถือได้เพื่อใช้ในการประเมินเชิงปริมาณของความสัมพันธ์ในลำดับวงศ์ตระกูล ประเมินการแบ่งประชากรชายในกลุ่มประชากรทั่วโลก โดยการใช้การแจกแจงความถี่ในรูปแบบ Y-STR เป็นต้น

2.2.4. The ALlele FREquency Database (ALFRED)

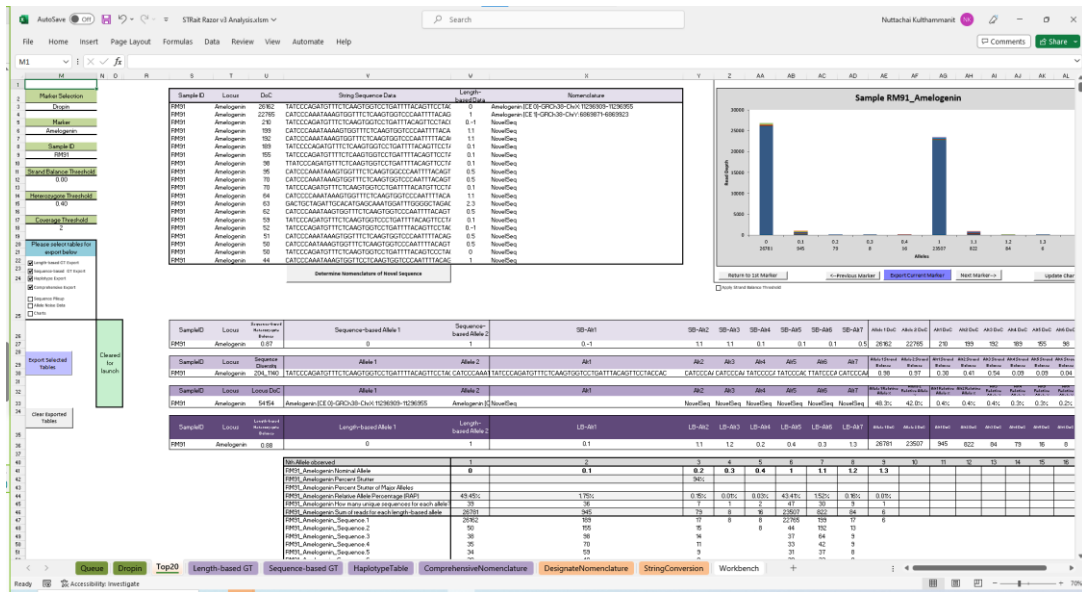
ALFRED [58] เป็นเว็บสาธารณะสำหรับการเปรียบเทียบข้อมูลของความถี่แอลลีลที่เกี่ยวข้องกับความหลากหลายของลำดับดีเอ็นเอในประชากรมนุษย์ ALFRED เน้นเก็บความถี่แอลลีลในประชากรที่กำหนดทางมานุษยวิทยาที่หลากหลาย โดยเน้นเก็บข้อมูลที่มีการศึกษาในประชากรหลายกลุ่ม ข้อมูลทั้งหมดใน ALFRED ถือเป็นสาธารณสมบัติสำหรับใช้ในการวิจัยและการสอน ตัวอย่างฟังก์ชันใน ALFRED เช่น สืบค้นข้อมูลโดยใช้ Unique Identifiers (UID) สืบค้นงานตีพิมพ์โดยใช้นามสกุลของผู้เขียน สืบค้นความถี่ของแอลลีล (Allele frequency) เป็นต้น

2.2.5. STRait Razor 3.0

STRait Razor 3.0 เป็นเครื่องมือที่ใช้ลักษณะเฉพาะของแฮพลไทป์เอสทีอาร์ในข้อมูล เอ็นจีเอสในการระบุเอสทีอาร์ STRait Razor 3.0 ถูกออกแบบและพัฒนาเพิ่มเติมให้มีความเร็วในการระบุเอสทีอาร์มากกว่าเวอร์ชันก่อนหน้าถึง 660 เท่า โดยความเร็วที่เพิ่มขึ้นมากเกิดจากการใช้วิธีการจัดดัชนีแบบใหม่ โปรแกรม STRait Razor 3.0 ถูกพัฒนาด้วยภาษา C++ สามารถใช้งานบนระบบปฏิบัติการหลักทั้งหมดรวมถึงไมโครซอฟต์วินโดวส์ ด้านความแม่นยำและความถูกต้องของ STRait Razor v3.0 รับรองความแม่นยำและความถูกต้อง 100% ซึ่งผลลัพธ์สอดคล้องกับ STRait Razor v2.0 แต่ STRait Razor v3.0 เพิ่มคุณสมบัติหลักหลายประการที่ช่วยลดความซับซ้อนของกระบวนการรายงานแฮพลไทป์ที่มีความถี่รวมต่ำ รวมไปถึงการรวบรวมแฮพลไทป์ทั้งหมดภายในโลคัสที่ถูกเข้ารหัสบนเกลียวตรงข้ามของโมเลกุลดีเอ็นเอ

2.2.6. Excel-based workbook of STRait Razor 3.0

Excel-based workbook of STRait Razor 3.0 (รูปที่ 20) เป็นเครื่องมือเสริมของ STRait Razor 3.0 ที่พัฒนาโดย โจนาธาน คิง (Jonathan King) ซึ่งเครื่องมือนี้จะอยู่ในรูปแบบของไฟล์เอ็กเซล โดยจะใช้ในการแสดงผลลัพธ์ข้อมูลที่ได้จาก STRait Razor 3.0 ออกมาในรูปแบบต่าง ๆ เช่น แอลลีลในแต่โลคัส ลำดับซ้ำเรียงต่อเนื่องแบบสั้นของแต่ละแอลลีล เป็นต้น



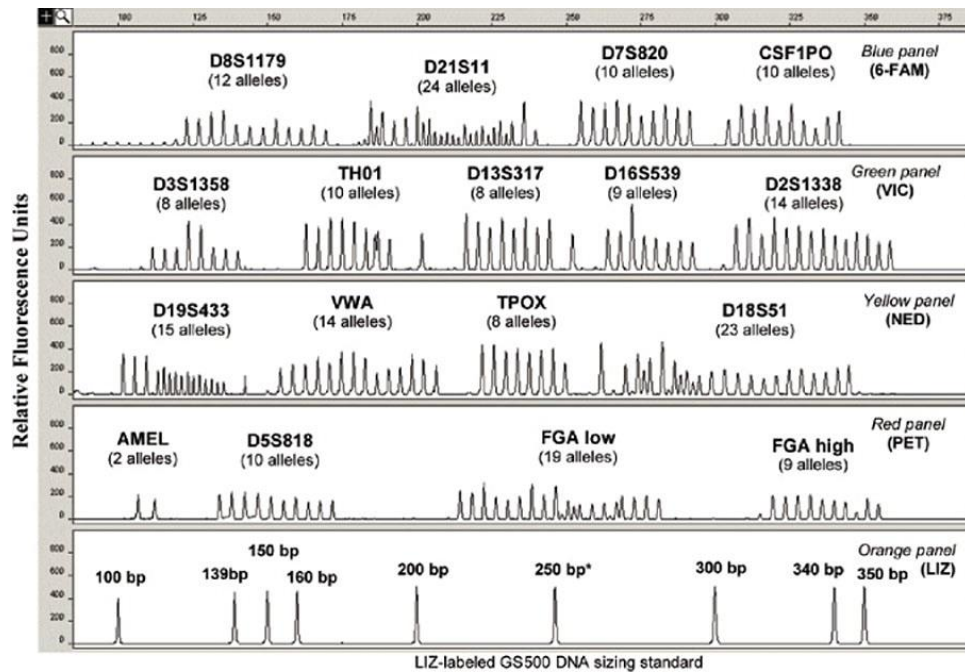
รูปที่ 20 ไฟล์ Excel-based workbook of STRait Razor 3.0

2.2.7. A survey of tools for analysing DNA fingerprints

งานวิจัยนี้ [59] นำเสนอผลการสำรวจและเปรียบเทียบเครื่องมือสำหรับวิเคราะห์ลายพิมพ์ดีเอ็นเอทั้งหมด 33 รายการ โดยใช้เกณฑ์ที่กำหนดไว้ล่วงหน้า การเปรียบเทียบเป็นการทดลองที่เกิดจากการใช้งานจริง รวมทั้งตรวจสอบเอกสารของเครื่องมือต่างๆ ทั้งนี้ไม่มีเครื่องมือที่ดีที่สุดในทุกสถานการณ์ แต่มีการสรุปเครื่องมือที่ดีที่สุดสำหรับสถานการณ์ต่างๆ เช่น เครื่องมือสำหรับการวิเคราะห์ภาพดีเอ็นเอที่มีความละเอียดต่ำ ได้มีการแนะนำให้ใช้ Molecular Imaging Software และ GelComparII เป็นต้น

2.2.8. Short tandem repeat typing technologies used in human identity testing

งานวิจัยนี้ [13] อธิบายเกี่ยวกับวิธีที่ใช้ในการหาเอสทีอาร์เพื่อใช้ในการระบุตัวตน (Human identity) รวมถึงการวิเคราะห์ลำดับเบสดีเอ็นเอ ซึ่งมีขั้นตอนประกอบด้วย การเพิ่มจำนวนดีเอ็นเอแบบมัลติเพล็กซ์ (Multiplex PCR amplification) – เป็นการเพิ่มจำนวนดีเอ็นเอโดยใช้ไพรเมอร์ (Primer) พร้อมกันหลายคู่ การแยกดีเอ็นเอที่ได้ที่ประกอบไปด้วยเอสทีอาร์ที่มีความยาวแตกต่างกันจากตอนแรกด้วยอิเล็กโทรโฟรีซิสหลอดเล็ก จากนั้นทำการเทียบกับแอลลีลมาตรฐาน (Allelic ladder) ที่ได้จากชุดคิท (Kit) เพื่อหาจีโนไทป์ โดยชุดคิทประกอบด้วยไพรเมอร์ แอลลีลมาตรฐาน (รูปที่ 21) รวมถึงสารประกอบมาตรฐานที่ใช้ในการทดลอง แอลลีลมาตรฐานจะถูกใช้เทียบกับขนาดของดีเอ็นเอที่ได้จากกระบวนการเพิ่มจำนวน



รูปที่ 21 แถบสีทั้งหมดของแอลลีลมาตรฐานในชุดคิท AmpF/STR Identifier ที่ใช้สำหรับการเทียบขนาดของดีเอ็นเอ (ภาพที่ 1 ของ [13])

2.2.9. STRAF (STR Analysis for Forensics)

งานวิจัยนี้เป็นเว็บแอปพลิเคชันที่เขียนโดยภาษาอาร์ เว็บแอปพลิเคชันนี้จะใช้ในการคำนวณค่าทางสถิติของกลุ่มประชากรจากข้อมูลเอสทีอาร์ สามารถคำนวณได้ทั้งข้อมูลตัวอย่างที่เป็น แฮพลอยด์ (Haploid) และ ดิพลอยด์ (Diploid) ข้อมูลนำเข้าจะใช้เป็นไฟล์ข้อความที่คั่นด้วยแท็บ (Text: Tab-delimited) หลังจากนำเข้าข้อมูลระบบ สามารถดาวน์โหลดผลลัพธ์ออกมาได้ โดยผลลัพธ์จะประกอบไปด้วยค่าทางสถิติ เช่น ความหนาแน่นของยีน (Gene Diversity: GD) และ ค่าความสามารถในการจำแนก (Polymorphism Information Content: PIC)

ปัจจุบันยังไม่มีระบบที่เหมือนกับระบบที่นำเสนอแต่ระบบที่คล้ายคลึงกันแต่มีฟังก์ชันบางส่วนที่เหมือนกับระบบที่นำเสนอ คือ STRider, YHRD และ ALFRED ซึ่งมีการเปรียบเทียบฟังก์ชันการทำงานในตารางที่ 1

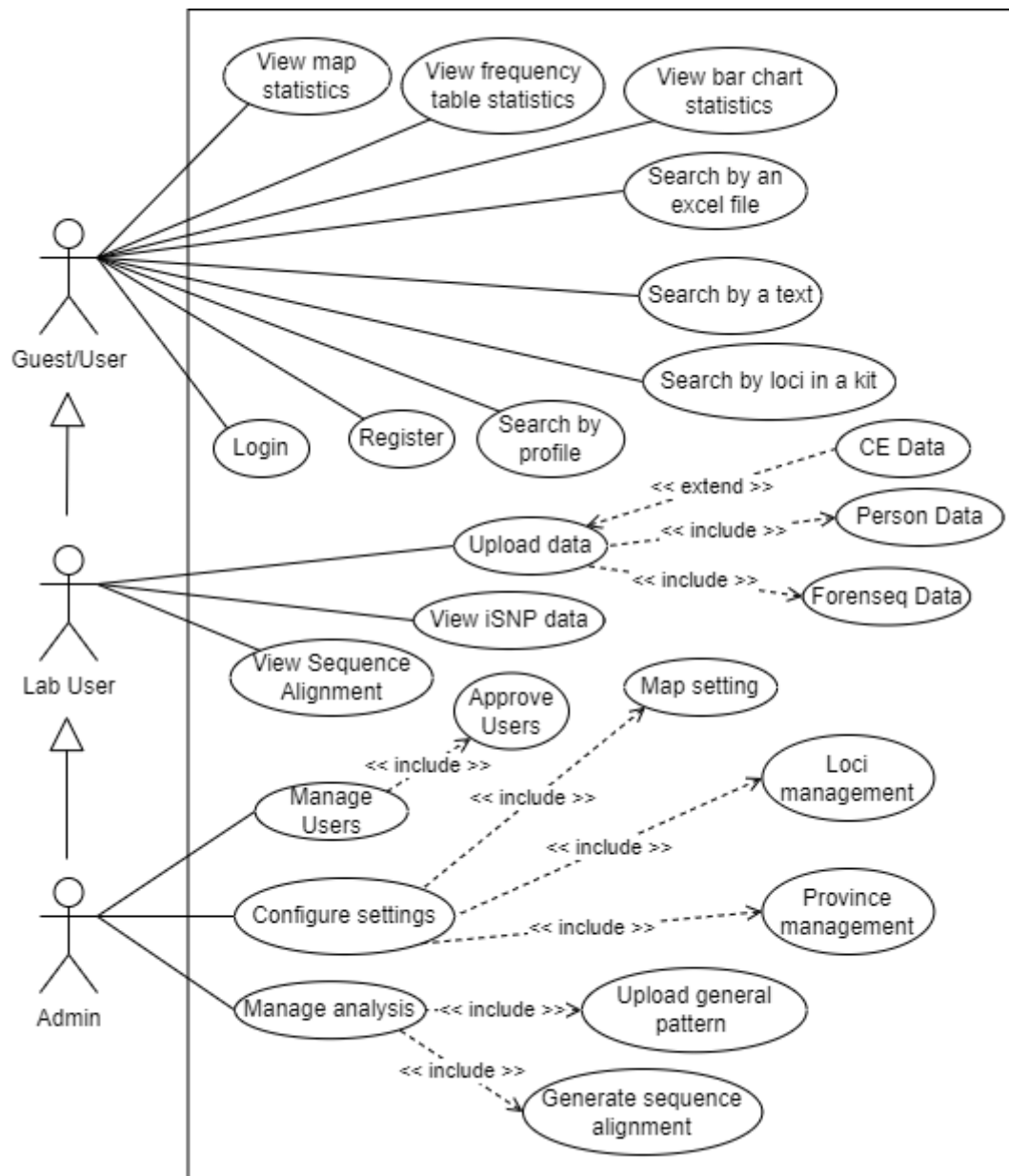
ตารางที่ 1 ฐานข้อมูลและฟังก์ชันการทำงานของแต่ละระบบ

ฟังก์ชันการทำงาน	STRIDER	STRait Razor	OmniPop	STRAF	YHRD	STRategy
สามารถติดตั้งภายในหน่วยงานหรือองค์กร						x
มีระบบการควบคุมการเข้าถึงข้อมูล						x
ระบบการจัดการข้อมูลเอสทีอาร์						x
ดูสืบค้นและ ข้อมูลภาพรวม	ข้อมูลภาพรวมตามภูมิภาค				x	x
	ข้อมูลภาพรวมตามโลเคชัน	x		x		x
	สืบค้นข้อมูลเอสทีอาร์	x			x	x
	สืบค้นข้อมูลโดยโปรไฟล์					x
ใช้งานระบบและ เครื่องมือสนับสนุน การวิเคราะห์	การเปรียบเทียบรูปแบบของลำดับซ้ำเรียงต่อเนื่องแบบสั้น			x		x
	เพิ่ม/ลดข้อมูลเอสทีอาร์					x
	เพิ่ม/ลดข้อมูลรายละเอียดของบุคคลในชุดข้อมูล					x
สามารถตั้งค่าสำหรับใช้งานที่แตกต่างในแต่ละห้องปฏิบัติการ						x
เพิ่มส่วนการวิเคราะห์และแสดงผลโดยผู้พัฒนาอื่นได้โดยง่าย						x
มีส่วนต่อประสานโปรแกรมประยุกต์						x

บทที่ 3

วิธีการดำเนินการวิจัย

3.1. ภาพรวมการทำงานของระบบ



รูปที่ 22 ยูสเคสไดอะแกรมของระบบที่นำเสนอ

ระบบที่นำเสนอถูกออกแบบให้อยู่ในรูปแบบเว็บแอปพลิเคชัน ฟังก์ชันหลักของระบบถูกแสดงผ่านยูสเคสไดอะแกรม (รูปที่ 22) โดยมีแอกเตอร์ (Actor) ผู้ใช้งานทั่วไป (User) เป็นผู้ใช้ที่ไม่ต้องเข้าสู่ระบบในการใช้งานและสามารถดูข้อมูลเชิงสถิติได้ทั้งหมด 3 แบบ ได้แก่ แผนที่ ตาราง และ

แผนภูมิแท่ง และสามารถใช้ฟังก์ชันการค้นหาได้ทั้งหมด 4 แบบ ได้แก่ ค้นหาโดยไฟล์เอ็กเซล ค้นหาโดยข้อความ ค้นหาโดยโลคัสในชุดคิทที่กำหนด และค้นหาโดยโปรไฟล์ สามารถลงทะเบียนและเข้าสู่ระบบได้ แอคเตอร์ผู้ใช้งานระดับห้องปฏิบัติการ (Lab user) ถ่ายทอดมาจากแอคเตอร์ผู้ใช้งานทั่วไป แต่สามารถเข้าถึงข้อมูลสเนป (SNP) ผลการวิเคราะห์ของการเปรียบเทียบลำดับเบสดีเอ็นเอ (Pattern alignment) และสามารถอัปโหลดข้อมูลได้ 2 ประเภทคือ ข้อมูลเอสทีอาร์ที่ได้จากเทคนิคการหาลำดับเบสยุคใหม่ (เอ็นจีเอส) ที่เป็นผลลัพธ์จาก ForenSeq และ ข้อมูลเพิ่มเติมที่เกี่ยวข้องกับข้อมูลเอสทีอาร์อื่นๆ แอคเตอร์ผู้ดูแลระบบ (Admin) ถ่ายทอดมาจากผู้ใช้งานระดับห้องปฏิบัติการ แต่สามารถจัดการผู้ใช้งานภายในระบบ ตั้งค่าตัวแปรต่างๆ ภายในระบบ เช่น แผนที่ โลคัส และจังหวัด สามารถควบคุมระบบจัดการการวิเคราะห์ ได้แก่ อัปโหลดรูปแบบทั่วไปของการเปรียบเทียบลำดับเบสดีเอ็นเอและสร้างรูปแบบการเปรียบเทียบลำดับเบสดีเอ็นเอ

3.2. วิเคราะห์และออกแบบฐานข้อมูลของระบบ

3.2.1. ศึกษาโครงสร้างข้อมูลเอสทีอาร์

ศึกษาโครงสร้างข้อมูลเอสทีอาร์จากห้องปฏิบัติการภาควิชานิติเวชศาสตร์ คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย (Department of Forensic Medicine, Faculty of Medicine, Chulalongkorn University) โดยไฟล์ข้อมูลเอสทีอาร์มี 2 รูปแบบ ได้แก่ รูปแบบไฟล์ที่ได้จากเทคนิคอิเล็กโทรโฟรีซิสหลอดเล็ก (รูปที่ 23) และรูปแบบไฟล์ที่ได้จากเทคนิคการลำดับเบสยุคใหม่ (รูปที่ 24 ถึงรูปที่ 26) โครงสร้างไฟล์ดังกล่าวสามารถจัดเก็บอยู่ในฐานข้อมูลเชิงสัมพันธ์ ไฟล์ที่ได้จากเทคนิคอิเล็กโทรโฟรีซิสหลอดเล็กจะอยู่ในรูปแบบของไฟล์ข้อความ ประกอบด้วยข้อมูลสรุปจีโนไทป์ โดยที่คอลัมน์แรกเป็น รหัสตัวอย่าง คอลัมน์ที่สองเป็นชื่อโลคัส คอลัมน์ที่สามและสี่ เป็นแอลลีลลำดับที่ 1 และ 2 ตามลำดับ โดยค่าแอลลีล แสดงถึงจำนวนซ้ำของรูปแบบลำดับนิวคลีโอไทด์สายสั้นที่โลคัสนี้ๆ

1	Sample Name	Marker	Allele 1	Allele 2
2	58CUSP573-C1	D3S1358	15	17
3	58CUSP573-C1	TH01	9	
4	58CUSP573-C1	D21S11	28	29
5	58CUSP573-C1	D18S51	12	14
6	58CUSP573-C1	Penta E	5	11
7	58CUSP573-C1	D5S818	9	12
8	58CUSP573-C1	D13S317	8	13
9	58CUSP573-C1	D7S820	9	10
10	58CUSP573-C1	D16S539	10	12
11	58CUSP573-C1	CSF1PO	11	12
12	58CUSP573-C1	Penta D	7	14
13	58CUSP573-C1	AMEL	X	Y
14	58CUSP573-C1	vWA	14	
15	58CUSP573-C1	D8S1179	10	16
16	58CUSP573-C1	TPOX	8	
17	58CUSP573-C1	FGA	23	25
18	58CUSP573-C1	D19S433	13	14.2
19	58CUSP573-C1	D2S1338	20	24

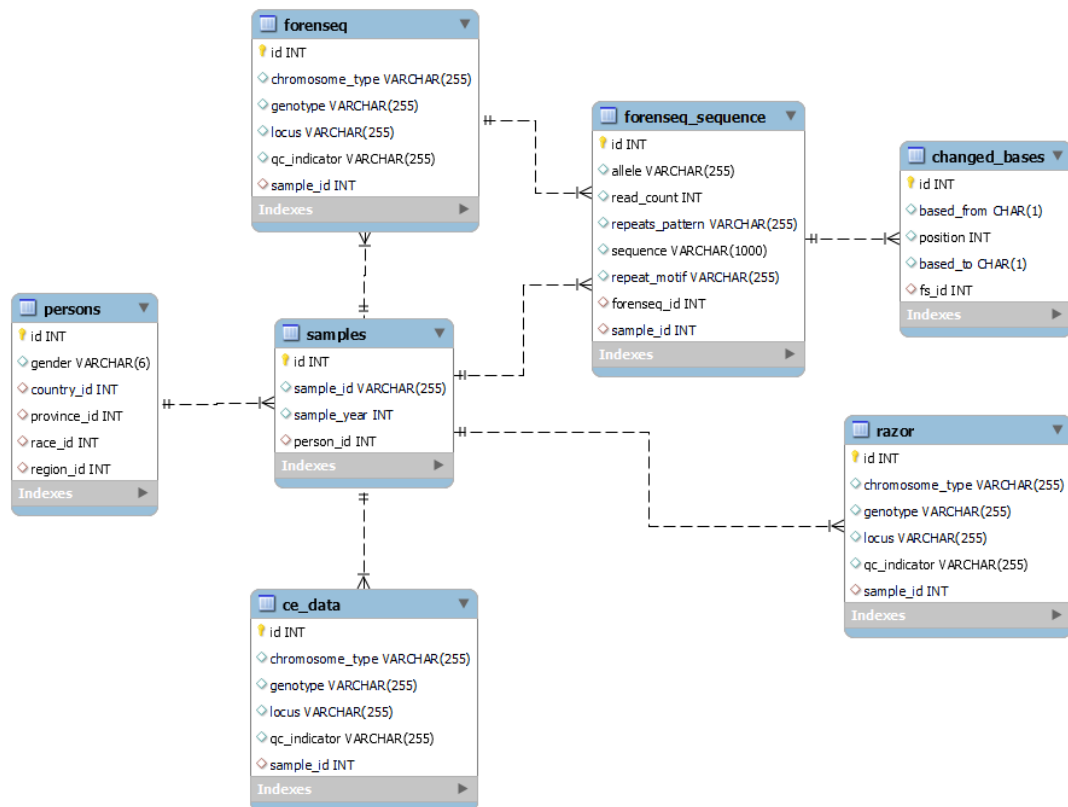
รูปที่ 23 ตัวอย่างข้อมูลและโครงสร้างของไฟล์เอสทีอาร์จากเทคนิคอิเล็กโทรพอเรซิสหลอดเล็ก

ไฟล์ที่ได้จากเทคนิคการหาลำดับเบสยุคใหม่อยู่ในรูปแบบไฟล์เอ็กเซล โดยข้อมูลในไฟล์ประกอบด้วยข้อมูลสรุปของ ออโตโซม วายโครโมโซม เอ็กซ์โครโมโซม และสลิป แยกออกเป็น 4 ซีท แต่ละซีทประกอบด้วย ส่วนหัวของไฟล์ที่เป็นข้อมูลของตัวอย่าง (รูปที่ 24) ข้อมูลสรุปจีโนไทป์ในแต่ละโลคัส (รูปที่ 25) ซึ่งสรุปจำนวนซ้ำของลำดับซ้ำเรียงต่อเนื่องแบบสั้นของลำดับเบสดีเอ็นเอของแต่ละแอลลีล (รูปที่ 26) ที่ผ่านเงื่อนไขการวิเคราะห์จากเครื่องหาลำดับเบสเอ็นจีเอส (Typed Allele เป็น Yes ในรูปที่ 29)

Sample Autosomal STR Report

Sample	032FA		
Project	training		
Analysis	1.0 - Training260416		
Run	Training260416		
Gender	XY		
Created	11 Nov 2016 at 04:08PM by admin@forenseq.uas		

รูปที่ 24 ตัวอย่างข้อมูลและโครงสร้างไฟล์ส่วนหัวของข้อมูลเอสทีอาร์ที่ได้จากการหาลำดับเบสยุคใหม่



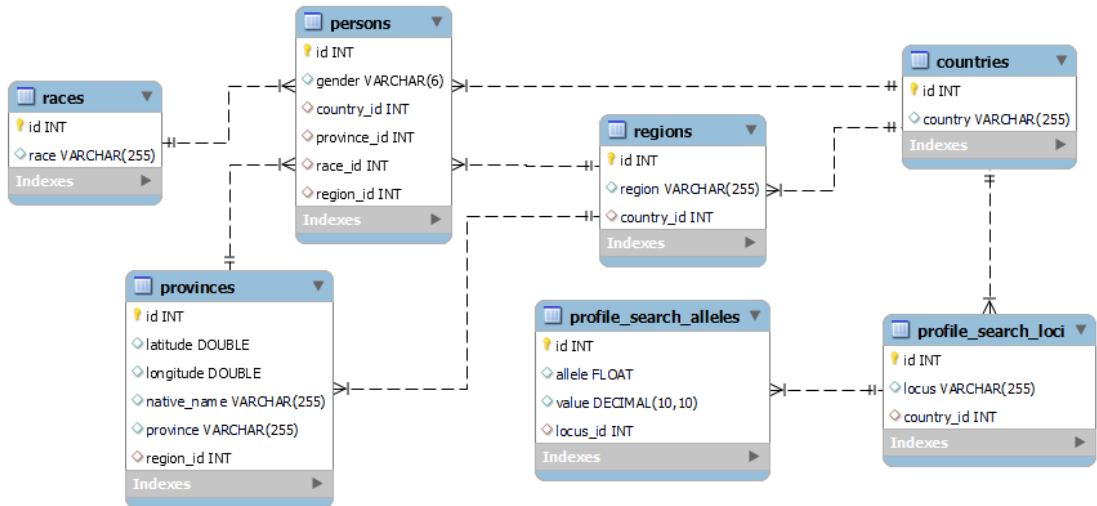
รูปที่ 27 อีอาร์ไดอะแกรมของข้อมูลเอสทีอาร์

ตาราง FORENSEQ_SEQUENCE (รูปที่ 27) เก็บข้อมูลผลการวิเคราะห์ของการเปรียบเทียบลำดับลำดับเบสดีเอ็นเอ Pattern Alignment ของแต่ละโลคัสที่ได้จากการวิเคราะห์ โดยตาราง CHANGED_BASES เก็บรายละเอียดนิวคลีโอไทด์ที่เปลี่ยนไปจากลำดับเบสอ้างอิงที่อัลกอริทึมของระบบค้นพบ เช่น ลำดับเบสอ้างอิงเดิมเป็น [AGTG]10 แต่ลำดับของตัวอย่างเป็น [AGTG]3 AGTA [AGTG]6 ระบบจะเก็บข้อมูลว่านิวคลีโอไทด์ G เปลี่ยนเป็น A เทียบจากลำดับเบสอ้างอิง

รูปที่ 28 แสดงอีอาร์ไดอะแกรม (ER Diagram) ของระบบในส่วนที่เกี่ยวข้องข้อมูลรายบุคคลและข้อมูลที่เกี่ยวข้องกับการสืบค้นแบบโปรไฟล์ ตารางข้อมูลรายบุคคล PERSONS สัมพันธ์กับตารางข้อมูลเชื้อชาติ RACES ตารางข้อมูลจังหวัด PROVINCES ตารางข้อมูลภูมิภาค REGIONS และตารางข้อมูลประเทศ CONTRIES ข้อมูลรายบุคคลสามารถอัปเดตได้ตามรูปที่ 29 ระบบได้กำหนดชุดคอลัมน์ข้อมูลพื้นฐานของรายบุคคล โดยคอลัมน์สุดท้ายคือคอลัมน์ที่ใช้อ้างอิงกับตาราง SAMPLES สำหรับเชื่อมความสัมพันธ์ระหว่าง SAMPLES และ PERSONS

ตาราง PROFILE_SEARCH_ALLELES ใช้เก็บค่าความถี่ของแต่ละแอลลีล และตาราง PROFILE_SEARCH_LOCI เก็บค่าโลคัส ตาราง PROFILE_SEARCH_ALLELES สัมพันธ์กับตาราง

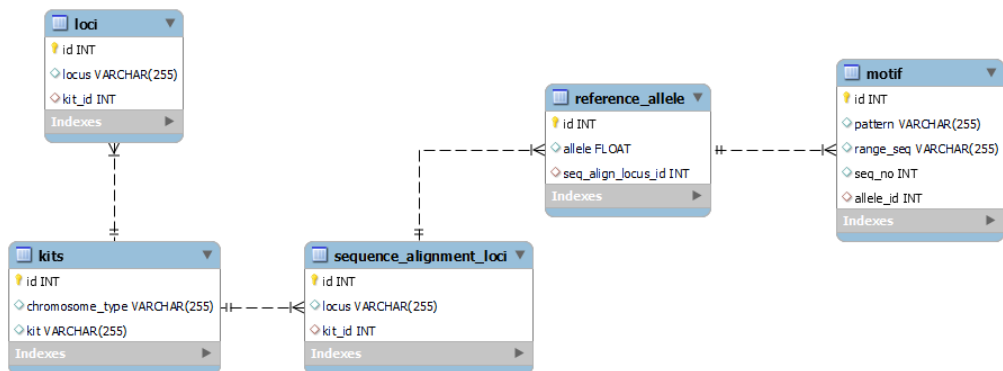
PROFILE_SEARCH_LOCI ผ่านค่าโลคัส ซึ่งสัมพันธ์ต่อไปยังตาราง COUNTRIES โดยการเก็บข้อมูลในลักษณะดังกล่าวทำให้สามารถเก็บค่าความถี่แอลลีล ตามประเทศ โลคัส และแอลลีลได้



รูปที่ 28 อีอาร์ไดอะแกรมของข้อมูลรายบุคคลและการสืบค้นแบบโปรไฟล์

	A	B	C	D	E	F	G	H	I
1	Firstname	Lastname	Age	Gender	Province	Region	Country	Race	SampleId
2	Somchai	Trilong	19	MALE	Bangkok	Central	Thailand	Thai	038M
3	Somsri	Thamnut	25	FEMALE	Bangkok	Central	Thailand	Thai	046M

รูปที่ 29 โครงสร้างไฟล์ที่ใช้สำหรับนำเข้าข้อมูลรายบุคคล



รูปที่ 30 อีอาร์ไดอะแกรมของข้อมูลโลคัสและชุดคิท

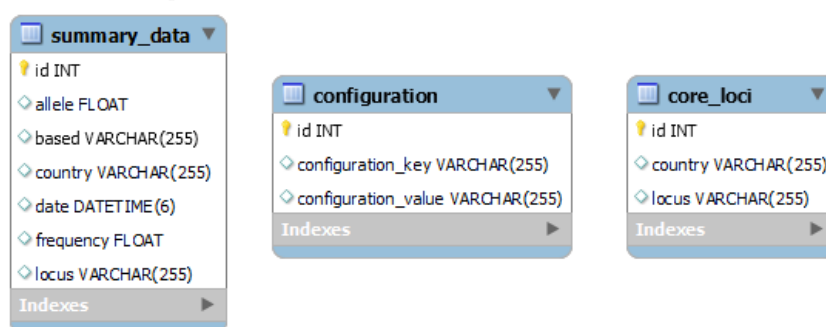
3.2.2. การออกแบบตารางข้อมูลเพื่อรองรับการตั้งค่าต่าง ๆ

รูปที่ 30 แสดงอีอาร์ไดอะแกรม (ER Diagram) ของระบบในส่วนความสัมพันธ์ของโลคัสในตาราง LOCI และชุดคิทในตาราง KITS โดยชุดคิทแต่ละชุดประกอบไปด้วยหลายโลคัสที่แตกต่างกัน

ไป การใช้ตารางในรูปแบบดังกล่าวจะสามารถทำให้ผู้ดูแลระบบสามารถเพิ่มชุดคิพและชุดโลคัสจากส่วนต่อประสานกับผู้ใช้ได้โดยง่าย

ตาราง MOTIF (รูปที่ 30) เก็บโมติฟ (Motif) หรือรูปแบบของเอสทีอาร์ ซึ่งมีความแตกต่างกันในแต่ละแอลลีลของแต่ละโลคัส ตาราง REFERENCE_ALLELE สัมพันธ์กับตาราง MOTIF โดยแต่ละแอลลีลสามารถมีรูปแบบการซ้ำที่แตกต่างกันได้มากกว่า 1 แบบและตาราง REFERENCE_ALLELE นี้สัมพันธ์กับตาราง SEQUENCE_ALIGNMENT_LOCI เนื่องจาก 1 โลคัสมีได้หลายแอลลีล และสุดท้ายสามารถจัดการกลุ่มโลคัสตามชุดคิพโดยสัมพันธ์กับตาราง KITS ทั้งนี้ตารางเหล่านี้ถูกนำไปใช้เทียบและอ้างอิงกับข้อมูลเอสทีอาร์ในตาราง FORENSEQ_SEQUENCE (รูปที่ 27) ที่อยู่ในระบบและสรุปผลออกมาเป็นรูปแบบการจัดเรียงเอสทีอาร์

ตาราง CONFIGURATION (รูปที่ 31) จะเก็บตัวแปรที่ใช้ในการตั้งค่าของระบบ เช่น ที่อยู่ของไฟล์แผนที่ ขนาดของแผนที่ ตำแหน่งเริ่มต้นของแผนที่ การสร้างตารางในลักษณะดังกล่าวทำให้สามารถเก็บค่าตัวแปรและเรียกใช้ทั้งระบบได้ง่าย ในส่วนของตาราง SUMMARY_DATA เป็นตารางที่ใช้สำหรับเก็บข้อมูลที่รวบรวมจากฐานข้อมูลสาธารณะ เช่น STRidER โดยในอนาคตสามารถนำเข้าฐานข้อมูลจากแหล่งอื่นเพื่อบูรณาการหรือเปรียบเทียบกับข้อมูลที่อยู่ภายในระบบได้โดยง่าย คอลัมน์ based ในตาราง SUMMARY_DATA จะระบุถึงแหล่งที่มาของข้อมูล โดยนำเข้าข้อมูลใส่ในตารางดังกล่าวและระบบจะนำข้อมูลนี้ไปประมวลผลรวมกับข้อมูลที่มีอยู่ในระบบและแสดงผลออกมาในรูปแบบต่างๆ ตาราง CORE_LOCI เก็บโลคัสหลักที่ใช้ในการสืบค้นแบบโปรไฟล์ (Profile search) ของแต่ละประเทศ ซึ่งผู้ดูแลระบบสามารถจัดการเพิ่ม ลบ หรือแก้ไขโลคัสหลักได้

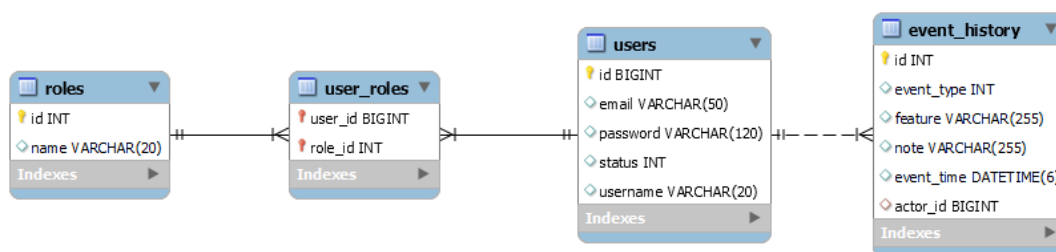


รูปที่ 31 อีอาร์ไออะแกรมของข้อมูลฐานข้อมูลสาธารณะ โลคัสหลักและการค่าตัวแปรในระบบ

3.2.3. แนวทางการออกแบบการกำหนดสิทธิของผู้ใช้ในระบบ

ระบบถูกออกแบบให้รองรับการกำหนดสิทธิผู้ใช้ระบบตามบทบาท (Role) โดยผู้ใช้นั้นคนสามารถมีได้มากกว่า 1 บทบาท และแต่ละบทบาทถูกกำหนดสิทธิในการเข้าถึงฟีเจอร์ต่างๆ ของ

ระบบได้แตกต่างกัน ในระบบที่นำเสนอได้กำหนดบทบาทเบื้องต้นไว้ 3 บทบาทได้แก่ ผู้ใช้ (User) ผู้ใช้งานระดับห้องปฏิบัติการ (Lab user) และผู้ดูแลระบบ (Administrator) โดยระบบรองรับการเพิ่มบทบาทในอนาคต ดังแสดงในรูปที่ 32 ผ่านตาราง ROLES, USER_ROLES และ USERS สำหรับตาราง EVENT_HISTORY ใช้สำหรับเก็บข้อมูลประวัติย้อนหลังของการกระทำที่สำคัญของแต่ละผู้ใช้ เพื่อตรวจสอบย้อนหลังในกรณีที่ระบบเกิดข้อผิดพลาด



รูปที่ 32 อีอาร์ไดอะแกรมของข้อมูลของระบบจัดการผู้ใช้

3.2.4. การวิเคราะห์เพื่อเปรียบเทียบลำดับเบสดีเอ็นเอ

การวิเคราะห์เพื่อเปรียบเทียบลำดับเบสดีเอ็นเอต้องใช้ข้อมูลอ้างอิง โดยข้อมูลอ้างอิงอยู่ในรูปแบบไฟล์เอ็กซ์เซล (รูปที่ 33) ซึ่งประกอบด้วย 4 คอลัมน์ คอลัมน์แรกแสดงโลคัส คอลัมน์ที่สองแสดงรูปแบบการซ้ำของเอสทีอาร์หรือลำดับเบสดีเอ็นเออ้างอิง โดยรูปแบบการเขียนจะอ้างอิงจากงานวิจัย [60] คอลัมน์ที่สามแสดงแอลลีลของโลคัสถ้าเป็น Default คือ ใช้สำหรับทุกๆ แอลลีลของโลคัสนั้น คอลัมน์สุดท้ายใช้กำหนดว่ารูปแบบในคอลัมน์ที่สองเป็นรีเวิร์ส (Reverse) หรือไม่

การวิเคราะห์เพื่อเปรียบเทียบลำดับเบสดีเอ็นเอในแต่ละโลคัสและแอลลีล ระบบวิเคราะห์รูปแบบการซ้ำเปรียบเทียบกับลำดับเบสดีเอ็นเออ้างอิง โดยค้นหารูปแบบอ้างอิงของแอลลีลดังกล่าวจากตาราง REFERENCE_ALLELE (รูปที่ 33) ถ้าไม่มี ระบบจะใช้ลำดับการซ้ำของแอลลีลเริ่มต้นในการวิเคราะห์

อัลกอริทึมที่ใช้ในการเปรียบเทียบลำดับเบสดีเอ็นเอของตัวอย่างกับลำดับเบสดีเอ็นเออ้างอิงจะทำงานโดยใช้อินพุต 3 ค่า ประกอบด้วย

1. ลำดับเบสดีเอ็นเอของตัวอย่างที่ได้จากเทคโนโลยีเอ็นจีเอส
2. แอลลีลของลำดับเบสดีเอ็นเอในข้อ 1.
3. ลำดับซ้ำเรียงต่อเนื่องแบบสั้นอ้างอิงของแอลลีลที่กำหนดในข้อ 2.

และแสดงขั้นตอนการหารูปแบบการซ้ำที่ปรากฏในลำดับเบสดีเอ็นเอของตัวอย่างที่ได้จากเทคโนโลยีเอ็นจีเอส ในภาคผนวก ข.

	A	B	C	D
1	Locus	Reference STR repeat motifs (Always Forward)	Allele	Orientation
2	D12S391	[AGAT] _n [AGAC] _n AGAT	Default	Forward
3	CSF1PO	[ATCT] _n	Default	Reverse
4	D10S1248	[GGAA] _n	Default	Forward
5	D13S317	[TATC] _n	Default	Forward
6	D16S539	[GATA] _n	Default	Forward
7	D17S1301	[AGAT] _n	Default	Forward
8	D18S51	[AGAA] _n	Default	Forward
9	D19S433	N18 [CCTT] _n ccta [CCTT] _n cttt [CCTT] _n	Default	Reverse
10	D1S1656	N10 CCTA [TCTA] _n	Default	Reverse
11	D20S482	[AGAT] _n	Default	Forward
12	D21S11	[TCTA] _n [TCTG] _n [TCTA] _n ta [TCTA] _n tca [TCTA] _n tccata [TCTA] _n	Default	Forward
13	D22S1045	[ATT] _n ACT [ATT] _n	Default	Forward
14	D2S1338	[GGAA] _n GGAC [GGAA] _n [GGCA] _n	Default	Reverse
15	D2S441	[TCTA] _n	Default	Forward
16	D2S441	[TCTA] _n TCA [TCTA] _n	10.3	Forward
17	D3S1358	[TCTA] _n [TCTG] _n [TCTA] _n	Default	Forward
18	D4S2408	[ATCT] _n	Default	Forward
19	D5S818	[ATCT] _n	Default	Reverse
20	D6S1043	[ATCT] _n	Default	Reverse
21	D7S820	N16 [TATC] _n	Default	Reverse
22	D8S1179	[TCTA] _n [TCTG] _n [TCTA] _n	Default	Forward
23	D9S1122	[TAGA] _n	Default	Forward
24	PentaD	AAAAG [AAAGA] _n	Default	Forward
25	PentaE	[TCTTT] _n	Default	Reverse
26	TH01	[AATG] _n	Default	Forward

รูปที่ 33 ไฟล์อ้างอิงเพื่อนำไปใช้วิเคราะห์เปรียบเทียบกับรูปแบบการซ้ำของข้อมูลลำดับเบสดีเอ็นเอของตัวอย่างที่ได้จากเทคโนโลยีเอ็นจีเอส

3.3. สถาปัตยกรรมของระบบ

สถาปัตยกรรมของระบบที่นำเสนอ (รูปที่ 34) เป็นระบบจัดการฐานข้อมูลเอสทีอาร์สำหรับงานนิติวิทยาศาสตร์ที่ใช้สถาปัตยกรรม 3 เลเยอร์ประกอบด้วย

1. ระบบหน้าบ้าน (Frontend)

เป็นส่วนของการแสดงผล (Presentation layer) ใช้สำหรับการตอบสนองกับผู้ใช้งาน และการแสดงผลต่าง ๆ เช่น ผลของการค้นหา ผลการเรียกดูข้อมูลเชิงสถิติ และผู้ดูแลระบบสามารถระบบหน้าบ้านพัฒนาโดยใช้ภาษาจาวาสคริปต์ (JavaScript) และไลบรารีที่ชื่อว่า รีแอค (React)

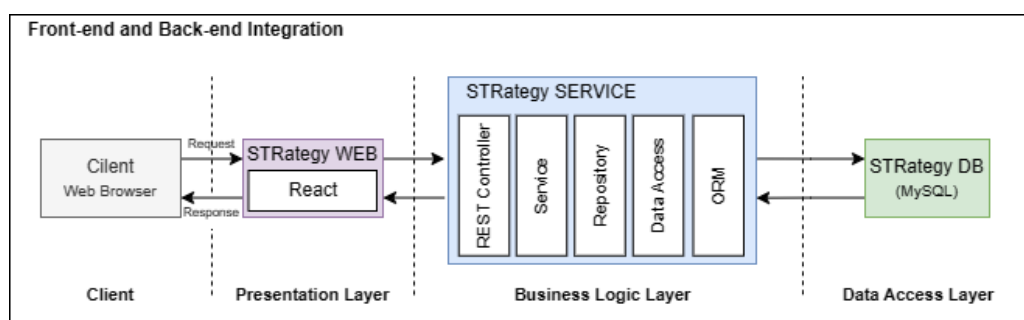
2. ระบบหลังบ้าน (Backend)

ส่วนของการประมวลผลข้อมูล (Business logic layer) โดยส่วนนี้จะประกอบไปด้วย 5 ส่วนย่อยดังที่แสดงในตารางที่ 2 ส่วนนี้จะมีหน้าที่สำหรับประมวลผลข้อมูลเอสทีอาร์ เช่น การ

อัปโหลดไฟล์เอ็กเซล การอัปโหลดไฟล์ข้อความ การวิเคราะห์ข้อมูล เป็นต้น ระบบหลังพัฒนาโดยใช้เฟรมเวิร์กสปริง (Spring framework) ที่ถูกสร้างบนเฟรมเวิร์กสปริง (Spring framework) ซึ่งเขียนโดยภาษาจาวา

3. ฐานข้อมูล (Database)

ส่วนของฐานข้อมูล เลือกใช้ MySQL ซึ่งเป็นระบบฐานข้อมูลเชิงสัมพันธ์สำหรับเก็บข้อมูลทั้งหมดในระบบ เช่น ข้อมูลของผู้ใช้งานในระบบ ข้อมูลเอสทีอาร์ ผลการวิเคราะห์ข้อมูล ข้อมูลตัวอย่าง จังหวัด และตัวแปรต่างๆ ภายในระบบ เป็นต้น



รูปที่ 34 โครงสร้างระบบ STRategy ภายใต้แนวคิดสถาปัตยกรรม 3-เลเยอร์

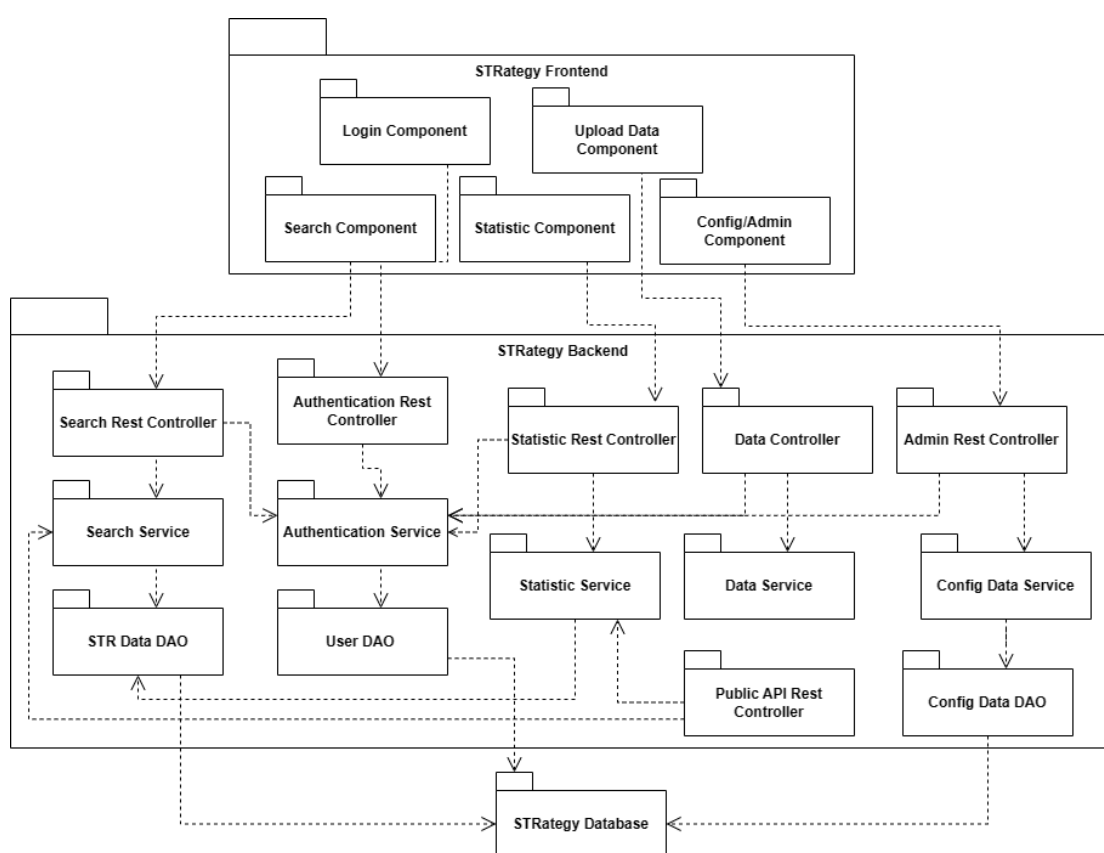
ตารางที่ 2 เลเยอร์ของระบบหลังบ้าน

คอลัมน์	เลเยอร์	หน้าที่
1	Rest Controller	ส่วนควบคุมการเข้าถึงฟังก์ชันต่างๆ ภายในระบบทำหน้าที่รับส่งข้อมูลในรูปแบบเจสัน (JSON) กับหน้าเว็บ
2	Service	ส่วนตรรกะที่ใช้ในการประมวลผลข้อมูลโดยรับส่งข้อมูลกับเว็บผ่าน Rest Controller และส่วนจัดการข้อมูลในฐานข้อมูลโดยรับส่งข้อมูลกับ Repository
3	Repository	ส่วนดึงข้อมูลจากเลเยอร์ของวัตถุการเข้าถึงข้อมูลเพื่อซ่อนวิธีการค้นหา เก็บ ลบ และแก้ไขข้อมูล
4	Data Access	คลาสที่ใช้สำหรับเก็บชุดคำสั่งหรือเมธอดที่ใช้ติดต่อกับฐานข้อมูล
5	ORM	ตัวกลางที่เชื่อมความสัมพันธ์ระหว่างอ็อบเจกต์และตารางข้อมูลในฐานข้อมูล

จากสถาปัตยกรรม 3 เลเยอร์ข้างต้น แต่ละเลเยอร์มีการแบ่งออกเป็นโมดูลย่อย (รูปที่ 35) โดยเลเยอร์แรก ระบบของหน้าบ้านมีชื่อ STRategy Frontend แบ่งออกเป็น 5 โมดูลย่อย ได้แก่

1. Login Component – เป็นส่วนของหน้าเว็บที่ทำหน้าที่เกี่ยวกับการเข้าสู่ระบบ

2. Upload Data Component – เป็นส่วนของหน้าเว็บที่ทำหน้าที่เกี่ยวกับการอัป-โหลดข้อมูลทั้งหมดในระบบ
3. Search Component – เป็นส่วนของหน้าเว็บที่ทำหน้าที่เกี่ยวกับการสืบค้นทั้งหมด
4. Statistic Component – เป็นส่วนของหน้าเว็บที่ทำหน้าที่เกี่ยวกับแสดงผลค่าสถิติและแสดงผลการวิเคราะห์ทั้งหมดในหน้าเว็บ
5. Config/Admin Component - เป็นส่วนของหน้าเว็บที่ทำหน้าที่เกี่ยวกับการตั้งค่าและการดำเนินการทุกอย่างที่เกี่ยวกับผู้ดูแลระบบ



รูปที่ 35 การทำงานภาพรวมของระบบ

เลเยอร์ถัดมาเป็นเลเยอร์ของส่วนองระบบหลังบ้านมีชื่อว่า STRategy Backend โดยแบ่งออกเป็น 3 กลุ่มหลัก ได้แก่

1. Controller – เป็นส่วนการติดต่อกับระบบหน้าบ้าน โดยมี 6 ส่วนหลัก ได้แก่
 - a. Search Rest Controller – เป็นส่วนติดต่อรองรับการสืบค้น
 - b. Authentication Rest Controller – เป็นส่วนติดต่อรองรับการตรวจสอบสิทธิการเข้าถึงข้อมูล

- c. Statistic Rest Controller – เป็นส่วนติดต่อรองรับการเข้าถึงการคำนวณค่าสถิติและการวิเคราะห์ข้อมูล
 - d. Data Rest Controller – เป็นส่วนติดต่อรองรับการอัปเดตข้อมูล
 - e. Admin Rest Controller – เป็นส่วนติดต่อรองรับการดำเนินการต่างๆ ของผู้ดูแล
 - f. Public API Rest Controller - เป็นส่วนต่อประสานโปรแกรมประยุกต์สำหรับบริการภายนอก
2. Service – เป็นส่วนที่ใช้ในการประมวลผลข้อมูล โดยมี 5 ส่วนหลัก ได้แก่
- a. Search Service – เป็นส่วนสืบค้นข้อมูลในรูปแบบต่างๆ
 - b. Authentication Service – เป็นส่วนตรวจสอบสิทธิการเข้าถึงข้อมูล
 - c. Statistic Service – เป็นส่วนคำนวณค่าสถิติและการวิเคราะห์ข้อมูลต่างๆ
 - d. Data Service – เป็นส่วนจัดการการอัปเดตข้อมูล
 - e. Config Data Service – เป็นส่วนจัดการการตั้งค่าข้อมูล
3. Data Access Object (DAO) – เป็นส่วนจัดการการเข้าถึงฐานข้อมูล โดยมี 3 ส่วนหลัก ได้แก่
- a. STR Data DAO – เป็นส่วนเข้าถึงข้อมูลเอสทีอาร์
 - b. User DAO – เป็นส่วนเข้าถึงข้อมูลผู้ใช้งานในระบบ
 - c. Config Data DAO – เป็นส่วนเข้าถึงข้อมูลการตั้งค่า

3.4. การประยุกต์ทฤษฎีทางคอมพิวเตอร์

งานวิจัยนี้นำเสนอระบบที่สนับสนุนการวิเคราะห์และการจัดการข้อมูลเอสทีอาร์ที่มีความหลากหลายทั้งรูปแบบข้อมูลและการแสดงผล เพื่อให้ระบบมีความยืดหยุ่นสูงและง่ายต่อการพัฒนาต่อยอดในอนาคต ระบบถูกออกแบบภายใต้แนวคิดและทฤษฎีทางคอมพิวเตอร์หลักๆ ได้แก่

3.4.1. โมเดล-วิว-คอนโทรลเลอร์ (Model-View-Controller)

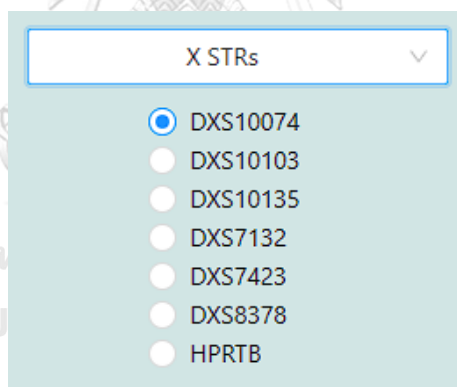
การออกแบบระบบใช้แบบรูปทางสถาปัตยกรรมที่ชื่อว่าโมเดล-วิว-คอนโทรลเลอร์เพื่อแยกระบบออกเป็นสามส่วน เพื่อทำให้ง่ายต่อการวางแผนการออกแบบและพัฒนา รวมทั้งการบำรุงรักษา เนื่องจากมีการแบ่งส่วนที่ชัดเจนและลดการพึ่งพากัน (Depend on) รองรับการผลิตผลหลากหลายรูปแบบเนื่องจากวิถูกแยกออกจากคอนโทรลเลอร์

ในระบบที่นำเสนอได้แบ่งระบบออกเป็นสามส่วนตามแบบรูปนี้ ได้แก่ ส่วนของการแสดงผล (View) เป็นส่วนของหน้าบ้านที่ใช้ไลบรารีที่ชื่อว่า React ส่วนต่อมาก็คือ โมเดล (Model) ที่ใช้ในการติดต่อกับฐานข้อมูลและส่วนของคอนโทรลเลอร์ (Controller) ที่ใช้สำหรับการประมวลผลข้อมูลจะอยู่ในส่วนของระบบหลังบ้านที่ใช้เฟรมเวิร์คสปริง

3.4.2. สถาปัตยกรรมเชิงคอมโพเนนท์ (Component-based architecture)

สถาปัตยกรรมเชิงคอมโพเนนท์เป็นหลักการที่แบ่งคอมโพเนนท์ออกเป็นส่วนย่อยและเล็กที่สุด จนคอมโพเนนท์นั้นมีหน้าที่แค่อย่างเดียวตามหลักการ Single responsibility principle สถาปัตยกรรมดังกล่าวช่วยลดความซ้ำซ้อนของโค้ด เนื่องจากสามารถนำคอมโพเนนท์กลับมาใช้ซ้ำ (Reusability) ได้

ในระบบที่นำเสนอหลักการถูกใช้ในส่วนที่เป็นระบบหน้าบ้านโดยมีการใช้ไลบรารีที่ชื่อว่า React โดยมีการแบ่งคอมโพเนนท์ที่ใช้แสดงผลออกเป็นส่วนย่อยและนำกลับมาใช้ซ้ำ ยกตัวอย่างเช่น ในส่วนของแท็บที่ใช้สำหรับเลือกโลคัส (รูปที่ 36) แท็บดังกล่าวมีการใช้ซ้ำในหน้าอื่นที่มีการทำงานที่คล้ายกัน



รูปที่ 36 แท็บที่ใช้สำหรับการเลือกโลคัส

3.4.3. แบบรูปการออกแบบ (Design pattern)

แบบรูปในระบบนี้จะถูกแบ่งออกเป็น 2 ประเภท ได้แก่ แบบรูปที่ใช้ในเฟรมเวิร์ก และแบบรูปที่ผู้พัฒนาเขียนเพิ่มเติม หัวข้อดังต่อไปนี้จะเป็นแบบรูปและหลักการที่เกี่ยวข้องที่ใช้ในเฟรมเวิร์ก

1. แบบรูปสถานะ (State)

แบบรูปนี้เป็นแบบรูปที่พฤติกรรมของอ็อบเจกต์จะเปลี่ยนแปลงไปตามสถานะของอ็อบเจกต์นั้น ซึ่งจะช่วยจัดการอ็อบเจกต์ที่มีสถานะที่ซับซ้อนให้ง่ายขึ้น แบบรูปนี้ถูกใช้ในไลบรารีที่ชื่อว่า React ที่ใช้คอมโพเนนท์ในการแสดงผล และคอมโพเนนท์จะเปลี่ยนแปลงไปตามสถานะของคอมโพเนนท์

นั้นๆ ซึ่งระบบที่นำเสนอ นำไลบรารีดังกล่าวมาใช้ในระบบหน้าบ้านที่แสดงส่วนต่อประสานกราฟิกกับผู้ใช้

2. หลักการอินเวอร์ชันออฟคอนโทรล (Inversion of Control principle)

หลักการอินเวอร์ชันออฟคอนโทรลเป็นหลักการที่มอบหมายให้อ็อบเจกต์อื่นทำหน้าที่ในการสร้างและจัดการอ็อบเจกต์ที่อ็อบเจกต์ดังกล่าวต้องการใช้งาน ซึ่งหลักการดังกล่าวมีประโยชน์หลายอย่าง เช่น ช่วยเรื่องการบำรุงรักษาโค้ด หรือช่วยให้ทดสอบระบบได้ง่าย เป็นต้น หลักการดังกล่าวถูกใช้ในหลายส่วนของเฟรมเวิร์กของระบบที่นำเสนอ เช่น แบบรูปดีเพนเดนซีอินเจกชันที่แยกการสร้างอ็อบเจกต์ออกจากคลาส หรือแบบรูปแพกทอรีที่ Spring IoC container ทำหน้าที่ในการสร้างและจัดการอ็อบเจกต์ทั้งหมดภายในระบบ

3. แบบรูปดีเพนเดนซีอินเจกชัน (Dependency Injection: DI pattern)

แบบรูปดีเพนเดนซีอินเจกชันใช้สำหรับแยกการสร้างอ็อบเจกต์ออกจากคลาสที่เกี่ยวข้อง เพื่อลดการขึ้นต่อกันระหว่างคลาส แบบรูปดังกล่าวทำให้การบำรุงรักษาโค้ดทำได้ง่าย เนื่องจากการแก้ไขในคลาสใดๆ จะไม่ส่งผลกระทบต่อคลาสอื่น โดยในเฟรมเวิร์กที่เชื่อว่าสปริงนำหลักการนี้มาใช้โดยมีอ็อบเจกต์ที่ชื่อว่า Spring IoC container เป็นตัวกลางในการทำหน้าที่การสร้างและส่งอ็อบเจกต์ที่จำเป็นไปยังอ็อบเจกต์ต่าง ๆ ซึ่งระบบที่นำเสนอพัฒนาภายใต้แบบรูปดังกล่าว

4. แบบรูปแพกทอรีเมธอด (Factory method design pattern)

แบบรูปแพกทอรีเมธอดเป็นแบบรูปที่แยกหน้าที่การสร้างอ็อบเจกต์ที่ต้องการให้กับคลาส แพกทอรี แบบรูปนี้ใช้แบบรูปดีเพนเดนซีอินเจกชันเป็นพื้นฐาน เฟรมเวิร์กสปริงมีการใช้แบบรูปดังกล่าว โดยมีคลาสที่ชื่อว่าแอปพลิเคชันคอนเท็กซ์ (Application context) ซึ่งเป็นคลาสที่ใช้แบบรูปนี้ในการจัดการอ็อบเจกต์และค่าต่างๆ ในส่วนของระบบหลังบ้านของระบบที่นำเสนอ

5. วัตถุการเข้าถึงข้อมูล (Data Access Object: DAO)

วัตถุการเข้าถึงข้อมูลเป็นรูปแบบที่แยกการเข้าถึงข้อมูลในฐานข้อมูลให้ออกจากส่วนการประมวลผลข้อมูล แบบรูปนี้สนับสนุนหลักการ single responsibility principle เนื่องจากการแยกให้แต่ละส่วนทำหน้าที่แค่อย่างเดียว และมีประโยชน์อีกหลายประการ เช่น มีความปลอดภัยมากขึ้น เนื่องจากรายละเอียดในส่วนของกรเก็บข้อมูลจะถูกแยกจากส่วนประมวลผลข้อมูล เกิดความไม่ขึ้นต่อกันระหว่างส่วนของการประมวลผลข้อมูลและส่วนของการเก็บข้อมูล ทำให้การแก้ไขในส่วนของ

การประมวลผลข้อมูลไม่ส่งผลกระทบต่อส่วนของการเก็บข้อมูล ระบบที่นำเสนอใช้เฟรมเวิร์กที่ชื่อว่า Hibernate ในการติดต่อกับฐานข้อมูลซึ่งมีการนำแบบรูปนี้มาใช้งานเช่นกัน

6. แบบรูปเทมเพลตเมธอด (Template method pattern)

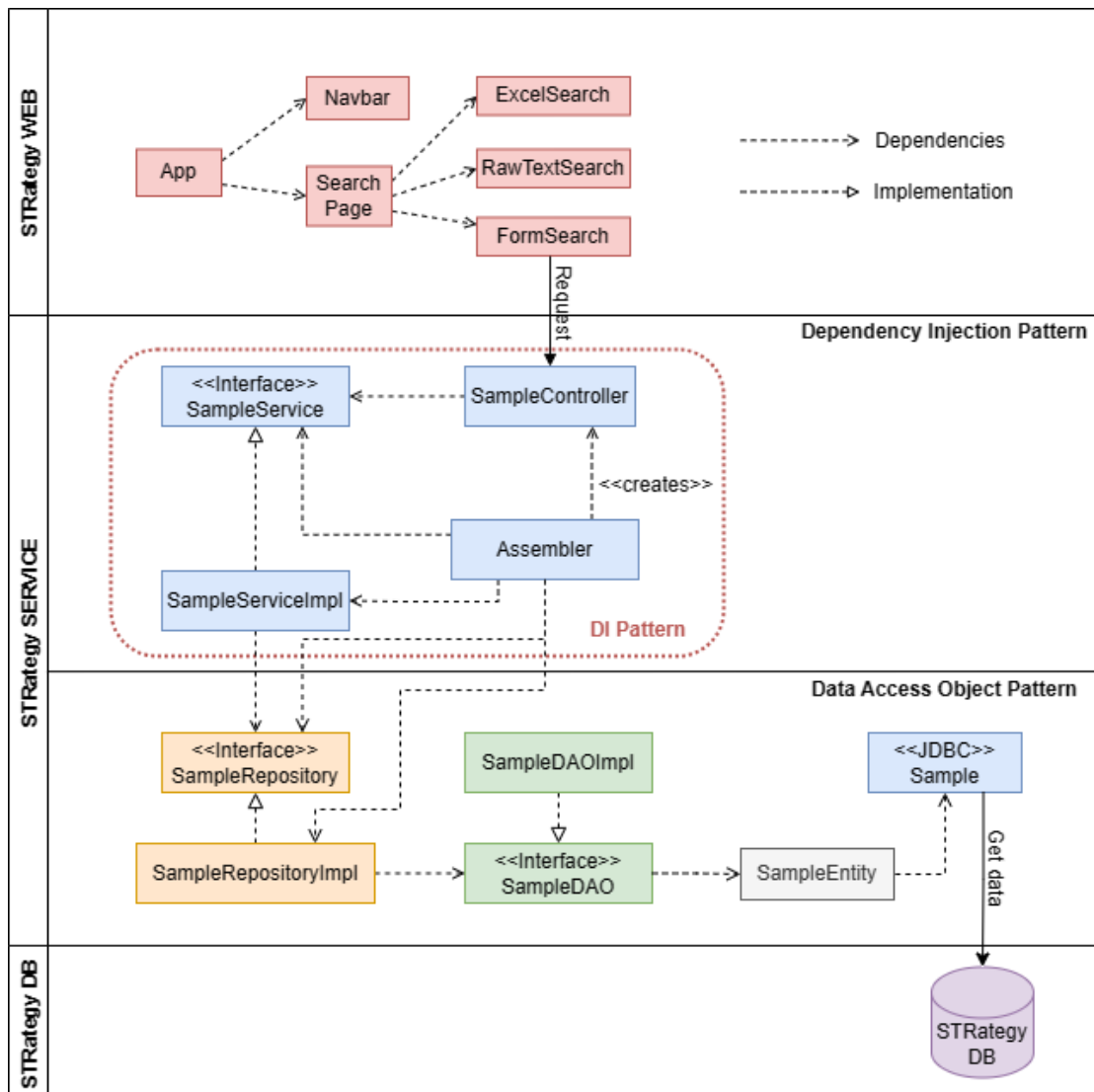
แบบรูปเทมเพลตเมธอด เป็นแบบรูปที่ใช้ในการกำหนดวิธีการทำขั้นตอนบางอย่างที่ถูกใช้งานบ่อย โดยผู้ใช้สามารถปรับเปลี่ยนวิธีการบางขั้นตอนได้ ในระบบที่นำเสนอมีการใช้แบบรูปเทมเพลตเมธอดของเฟรมเวิร์ก Hibernate ในการดึงข้อมูล (Query) จากฐานข้อมูลของระบบหลังบ้าน

7. แบบรูปซิงเกิลตัน (Singleton)

แบบรูปซิงเกิลตัน เป็นแบบรูปที่อ็อบเจกต์ของคลาสใดๆ มีเพียงอ็อบเจกต์เดียวเท่านั้น ในระบบที่นำเสนอ นำแบบรูป Singleton มาใช้ในคลาสที่ทำหน้าที่จัดการข้อมูลในฐานข้อมูลเพื่อลดความผิดพลาดในการจัดการข้อมูลภายในฐานข้อมูลเนื่องจากมีอ็อบเจกต์เพียงตัวเดียวเท่านั้นที่ทำหน้าที่จัดการข้อมูล เฟรมเวิร์กสปริงที่ใช้ในระบบหลังบ้านของระบบที่นำเสนอมีการสร้างวิธีการสร้างอ็อบเจกต์ได้หลายหลายรูปแบบ หนึ่งในนั้นคือการใช้แบบรูปซิงเกิลตัน ซึ่งระบบที่นำเสนอได้ตั้งค่าให้เฟรมเวิร์กใช้แบบรูปดังกล่าวในการสร้างอ็อบเจกต์

8. แบบรูปพร็อกซี (Proxy design pattern)

แบบรูปพร็อกซี เป็นแบบรูปที่ผู้ใช้ติดต่อกับพร็อกซีอ็อบเจกต์ที่ทำหน้าที่ควบคุมอ็อบเจกต์จริงอีกที โดยระบบที่นำเสนอมีการนำแบบรูปพร็อกซีมาใช้ในการสร้างทรานแซคชันเพื่อติดต่อกับฐานข้อมูลของระบบหลังบ้าน เพื่อเพิ่มประสิทธิภาพในการติดต่อระหว่างแอปพลิเคชันและฐานข้อมูล เฟรมเวิร์ก Hibernate ที่ระบบที่นำเสนอใช้ในการติดต่อกับฐานข้อมูลมีการใช้แบบรูปพร็อกซี โดยในการติดต่อกับฐานข้อมูล Hibernate จะมีพร็อกซีอ็อบเจกต์ที่ทำหน้าที่จัดเตรียมข้อมูลที่ต้องการแก้ไขทั้งหมดเพื่อติดต่อฐานข้อมูลเพียงครั้งเดียว



รูปที่ 37 ตัวอย่างการประยุกต์ใช้แบบรูปในระบบ

รูปที่ 37 แสดงตัวอย่างแบบรูปบางส่วนที่ถูกใช้งานในระบบที่นำเสนอ โดยระบบที่นำเสนอมีการเขียนโค้ดให้สอดคล้องกับแบบรูปดังกล่าว ตัวอย่างการค้นหาด้วยชุดของแอลลีและโลคัส โดยที่ในส่วนของระบบหน้าบ้านมีการใช้ไลบรารีที่ชื่อว่า React โดยไลบรารีนี้ใช้แบบรูปสถานะ เมื่อผู้ใช้เรียกใช้ฟังก์ชันสืบค้นโดยใช้ฟอร์มที่ถูกเขียนในฟังก์ชันคอมโพเนนต์ `FormSearch` หน้าเว็บจะติดต่อถึงระบบหลังบ้านที่พัฒนาโดยเฟรมเวิร์กที่ชื่อว่าสปริง เฟรมเวิร์กสปริงกำหนดให้หน้าเว็บติดต่อผ่านคลาส `SampleController` โดยที่คลาสดังกล่าวจะเรียกใช้งานคลาส `SampleService` ที่ทำหน้าที่ในการประมวลผลข้อมูล และคลาส `SampleService` จะขอข้อมูลจากคลาส `SampleRepository` โดยคลาสนี้ทำหน้าที่ในการดึงข้อมูลจาก `SampleDAO` ที่มีการทำงานแบบรูปวัตถุการเข้าถึงข้อมูล โดยอี

อบเจกต์ของทุกคลาสที่กล่าวมา มีการประยุกต์แบบรูปดีเพนเดนซีอินเจกชันผ่าน Spring IoC Container ซึ่งทำหน้าที่จัดการและส่งอ็อบเจกต์ที่สำคัญออกไปให้อ็อบเจกต์ต่าง ๆ ในระบบใช้งาน

แบบรูปต่อไปนี้เป็นแบบรูปที่ไม่ได้เป็นส่วนของเฟรมเวิร์กสปริง แต่ผู้พัฒนาประยุกต์ใช้เพิ่มเติมเข้ามาเพื่อให้โค้ดมีความยืดหยุ่น บำรุงรักษา และต่อยอดได้ง่ายในอนาคต

9. แบบรูป Private Class Data

แบบรูป Private Class Data เป็นแบบรูปที่ไม่อนุญาตให้ผู้ใช้แก้ไขคุณลักษณะของอ็อบเจกต์ได้โดยตรง แบบรูปนี้สนับสนุนหลักการห่อหุ้มซึ่งป้องกันไม่ทำให้ผู้ใช้สามารถเปลี่ยนแปลง แก้ไข หรือสร้างความเสียหายให้กับตัวแปรต่างๆ ที่อยู่ภายในอ็อบเจกต์ได้ คลาสที่ทำหน้าที่ในการส่งข้อมูลออก และรับข้อมูลเข้ามาในระบบหลังบ้านของระบบที่นำเสนอมีการใช้แบบรูปดังกล่าว

10. แบบรูป Builder

แบบรูป Builder เป็นแบบรูปที่ช่วยให้สามารถสร้างอ็อบเจกต์ที่มีความซับซ้อนทีละขั้นตอนได้ แบบรูปนี้ช่วยสร้างอ็อบเจกต์ที่มีประเภทและการทำงานต่างกันได้ โดยใช้โค้ดต้นฉบับเดียวกัน ซึ่งในระบบที่นำเสนอใช้แบบรูปนี้ในการสร้างอ็อบเจกต์ที่มีคุณลักษณะแตกต่างกัน แบบรูปดังกล่าวถูกใช้ในระบบที่นำเสนอในส่วนของการสร้างอ็อบเจกต์ผู้ใช้งาน (User) ที่ใช้สำหรับการล็อกอินเข้าสู่ระบบ

11. แบบรูป Strategy

แบบรูป Strategy เป็นแบบรูปที่ทำให้เราสามารถกำหนดชุดของพฤติกรรมที่หลากหลายของอ็อบเจกต์และสามารถเลือกและเปลี่ยนพฤติกรรมนั้นได้ขณะรันไทม์ (Run time) แบบรูปดังกล่าวถูกใช้ในระบบที่นำเสนอในส่วนของการสืบค้นแบบโปรไฟล์ เพื่อให้ผู้ใช้สามารถเลือกใช้อัลกอริทึมในการสืบค้นแบบโปรไฟล์ได้ในขณะรันไทม์ โดยมีการอธิบายแบบละเอียดในหัวข้อ 4.5. แนวทางการพัฒนาต่อยอดอัลกอริทึมวิเคราะห์ข้อมูลในระบบ ในบทที่ 4

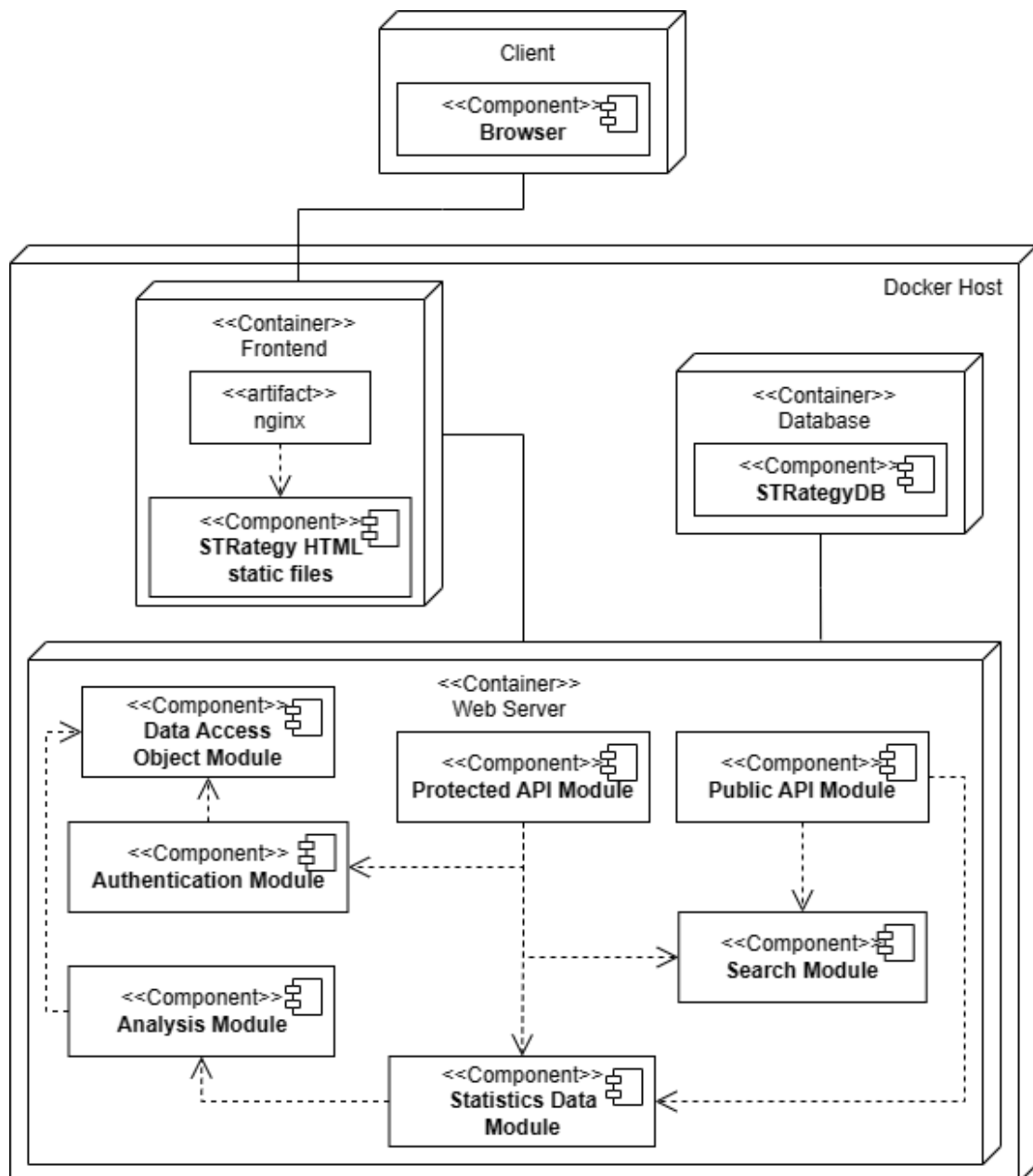
3.4.4. การแปลงอ็อบเจกต์เชิงสัมพันธ์ (Object-Relational Mapping: ORM)

การแปลงอ็อบเจกต์เชิงสัมพันธ์เป็นเทคนิคที่มีการใช้อ็อบเจกต์ในการจัดการข้อมูลในฐานข้อมูลแทนการเข้าถึงตารางภายในฐานข้อมูลโดยตรง ทำให้โค้ดต้นฉบับอ่านง่ายเนื่องจากการจัดการข้อมูลใช้ภาษาเดียวกับที่ผู้พัฒนาใช้ในการพัฒนาระบบ ช่วยให้ส่วนของการประมวลผลข้อมูลไม่ขึ้นกับฐานข้อมูล และช่วยลดเวลาในการพัฒนาเนื่องจากผู้พัฒนาไม่จำเป็นต้องรู้ภาษาสอบถามเชิงโครงสร้าง (Structured Query Language: SQL) เฟรมเวิร์ก Hibernate เป็นเฟรมเวิร์กที่ใช้เทคนิค

นี้ในการติดต่อฐานข้อมูล ซึ่งระบบที่นำเสนอมีการใช้เฟรมเวิร์กที่ชื่อว่า Hibernate ในการติดต่อกับฐานข้อมูล

3.5. การติดตั้งและการใช้งานระบบ

ระบบที่นำเสนอรองรับเบราว์เซอร์ กูเกิลโครม (Google Chrome) และไมโครซอฟต์เอดจ์ (Microsoft Edge) โดยมีการนำส่วนของหน้าบ้าน หลังบ้าน และฐานข้อมูลบรรจุไว้ในคอนเทนเนอร์ (Container) ของด็อกเกอร์ซึ่งทำให้ระบบที่นำเสนอมีความยืดหยุ่นและติดตั้งง่าย โดยสามารถติดตั้งและใช้งานบนระบบปฏิบัติการที่หลากหลาย เช่น วินโดวส์ (Windows) ลินุกซ์ (Linux) และแมคโอเอส (Mac OS) คอนเทนเนอร์ภายในระบบ (รูปที่ 38) ที่นำเสนอจะประกอบด้วย 3 คอนเทนเนอร์อยู่ภายในคอนเทนเนอร์ด็อกเกอร์ เบราวเซอร์ติดต่อสื่อสารกับคอนเทนเนอร์ที่เป็นระบบหน้าบ้าน หลังจากนั้นระบบหลังบ้านจะทำหน้าที่ติดต่อสื่อสารกับระบบหลังบ้าน และมีคอนเทนเนอร์ของฐานข้อมูลที่ทำหน้าที่เก็บข้อมูลและส่งข้อมูลให้กับคอนเทนเนอร์ของระบบหลังบ้าน



รูปที่ 38 ภาพรวมของระบบในรูปแบบ deployment diagram

บทที่ 4

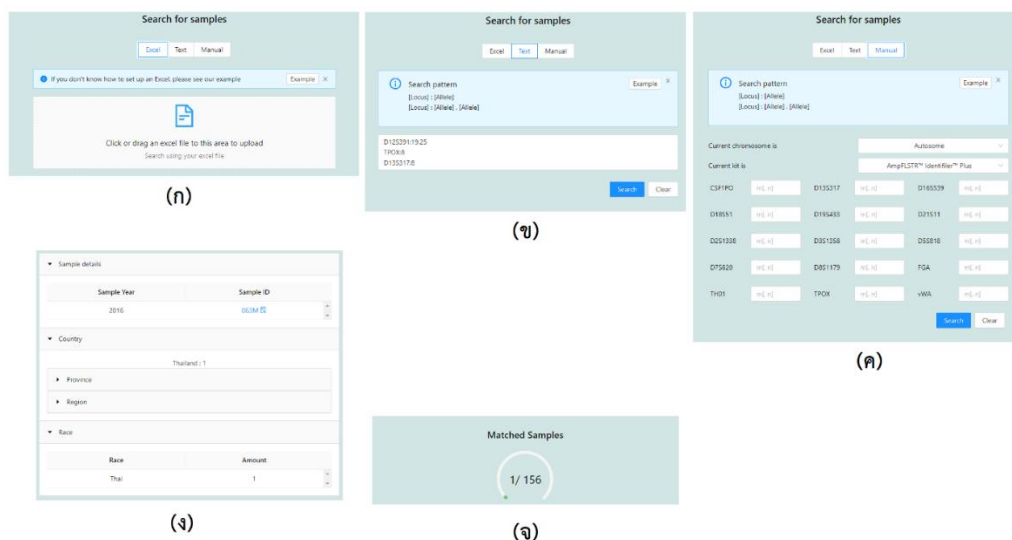
ผลวิจัยและอภิปรายผล

4.1. ภาพรวมของระบบ

ระบบที่ได้ออกแบบและพัฒนาขึ้น เป็นระบบที่อยู่ในรูปแบบเว็บแอปพลิเคชันที่ช่วยในการจัดเก็บและวิเคราะห์ข้อมูลเอสทีอาร์ ระบบมีการควบคุมการเข้าถึงข้อมูลที่แตกต่างกันตามบทบาท โดยใช้หลักการการกำหนดสิทธิผู้ใช้ระบบตามบทบาท ผู้ใช้ทั่วไปสามารถค้นหาข้อมูลเอสทีอาร์ภายในฐานข้อมูลได้โดยใช้โลคัสและแอลลีล แต่ผลลัพธ์ในการค้นหาจะเป็นสถิติภาพรวม เช่น จำนวนตัวอย่างที่ตรงกัน ผู้ใช้ทั่วไปสามารถเข้าถึงข้อมูลที่เป็นภาพรวมต่าง ๆ เช่น การแสดงความถี่แอลลีลตามภูมิภาค การเปรียบเทียบความถี่แอลลีลระหว่างข้อมูลในฐานข้อมูลของระบบที่พัฒนาขึ้นกับฐานข้อมูลสาธารณะ STRidER การแสดงความถี่แอลลีลตามโลคัส ผู้ใช้งานห้องทดลองสามารถเข้าถึงคุณลักษณะที่ผู้ใช้งานทั่วไปสามารถเข้าถึงได้รวมถึงสามารถเข้าถึงข้อมูลระดับรายบุคคล เช่น การค้นหาข้อมูลเอสทีอาร์จะแสดงรหัสตัวอย่าง การวิเคราะห์เพื่อเปรียบเทียบลำดับเบสดีเอ็นเอ การอัปโหลดข้อมูลเอสทีอาร์เข้าสู่ระบบ ผู้ดูแลระบบสามารถเข้าถึงคุณลักษณะที่ผู้ใช้งานห้องทดลองสามารถเข้าถึงได้และสามารถตั้งค่าต่าง ๆ ในระบบ เช่น การตั้งค่าแผนที่ในการแสดงผล การจัดการผู้ใช้ในระบบได้

4.2. การสืบค้นข้อมูลเอสทีอาร์

การสืบค้นข้อมูลเอสทีอาร์ในฐานข้อมูลที่ออกแบบและพัฒนาขึ้น มีตัวเลือกในการค้นหาทั้งหมด 3 แบบ แบบที่หนึ่งคือการสืบค้นโดยใช้ไฟล์ตัวอย่าง (รูปที่ 39ก) ที่เป็นผลลัพธ์จากการอ่านโดยใช้เทคนิคการลำดับเบสยุคใหม่ แบบที่สองคือการค้นหาโดยใช้รูปแบบโลคัสและแอลลีล (รูปที่ 39ข) แบบที่สามคือการค้นหาโดยใช้โลคัสและแอลลีลตามชุดคิทที่เลือก (รูปที่ 39ค) หลังจากกดค้นหาจะได้ผลลัพธ์ออกมา 2 รูปแบบ รูปแบบแรก รูปแบบแรกจะแสดงในส่วนของรายละเอียดในแต่ละตัวอย่าง (รูปที่ 39ง) ด้วย ซึ่งการแสดงผลแบบนี้จะเป็นการแสดงผลให้กับผู้ใช้งานในห้องทดลองหรือผู้ดูแลระบบเท่านั้น สำหรับรูปแบบที่สองจะแสดงเฉพาะจำนวนตัวอย่างที่ตรงกับข้อมูลนำเข้าที่ผู้ใช้ใส่เข้ามา (รูปที่ 39จ) ซึ่งการแสดงผลแบบนี้เป็นการแสดงผลให้กับผู้ใช้ทั่วไป



รูปที่ 39 การสืบค้นข้อมูลเอสทีอาร์ (ก) สืบค้นข้อมูลเอสทีอาร์โดยใช้ไฟล์ตัวอย่าง (ข) สืบค้นข้อมูลเอสทีอาร์โดยใช้โลคัสและจีโนไทป์ (ค) สืบค้นข้อมูลเอสทีอาร์โดยใช้โลคัสและจีโนไทป์จากชุดคิท (ง) ผลลัพธ์จากการสืบค้นแบบรายตัวอย่าง (จ) ผลลัพธ์จากการสืบค้นแบบภาพรวม

การสืบค้นแบบโปรไฟล์ (รูปที่ 40ก) ผู้ใช้งานต้องเลือกชุดโลคัสหลัก (Core loci) ของประเทศที่ต้องการ โดยโลคัสหลักจะแตกต่างกันไปในแต่ละประเทศ หลังจากเลือกได้แล้วผู้ใช้งานต้องใส่จีโนไทป์ลงในฟอร์มดังกล่าว วิธีการใส่จีโนไทป์สามารถทำได้สองวิธี วิธีแรกคือกรอกด้วยมือที่ละช่อง วิธีที่สองคือการอัปโหลดไฟล์ตัวอย่างที่เป็นผลลัพธ์จากเทคนิคลำดับเบสยุคใหม่ ระบบจะนำค่าจีโนไทป์ในแต่ละโลคัสไปใส่ในแต่ละช่องให้อัตโนมัติ โดยค่าของจีโนไทป์ในโลคัสหลักห้ามเป็นค่าว่าง หลังจากนั้นระบบจะคำนวณค่าจีโนไทป์ที่ได้และแสดงผลออกมาเป็นตาราง (รูปที่ 40ข) ที่เรียงค่าจากประเทศที่มีความคล้ายคลึงกับข้อมูลโปรไฟล์มากที่สุดไปประเทศที่มีความคล้ายกับข้อมูลโปรไฟล์น้อยสุด (Most common to least common) และจะแสดงรายละเอียดที่ใช้ในการคำนวณออกมาเป็นตารางรายประเทศ (รูปที่ 41)

▼ Select Core Loci Country

UK ▼

▼ Core Loci List

Fields File

* Amelogenin

D1S1656

TPOX

D2S441

* D2S1338

* D3S1358

D4S2408

▼ Summary result

Populations Sorted by Frequency (Most Common to Least Common)

Country	Frequency
Thailand	4.15e+35
Asia	4.40e+35
GREECE	8.11e+36
Entire Database	8.50e+36
BOSNIA AND HERZEGOWINA	2.27e+37
SAUDI ARABIA	2.27e+37
AUSTRIA	4.06e+37
SWEDEN	4.46e+37
CZECH REPUBLIC	6.94e+37
FINLAND	9.29e+37

< 1 2 3 >

(ก)

(ข)

รูปที่ 40 การสืบค้นแบบโปรไฟล์ (ก) ฟอรัมการสืบค้น (ข) ผลลัพธ์ภาพรวม

▼ Each country result

Thailand (4.15e+35)

Locus	p	q	value	1/value
Amelogenin	0	0	0.000	1
D1S1656	0.069672131	0.100409836	0.014	71
TPOX	0.110655738	0.026639344	0.006	170
D2S441	0.295081967	0.129098361	0.076	13
D2S1338	0.235655738	0.114754098	0.054	18
D3S1358	0.381147541	0.209016393	0.159	6
D4S2408	0.317622951	0.317622951	0.103	10
FGA	0.204918033	0.204918033	0.044	23
D5S818	0.225409836	0.295081967	0.133	8
CSF1PO	0.024590164	0.420081967	0.021	48
D6S1043	0.163934426	0.129098361	0.042	24
D7S820	0.19057377	0.176229508	0.067	15
D8S1179	0.129098361	0.159836066	0.041	24
D9S1122	0.391393443	0.342213115	0.268	4
D10S1248	0.225409836	0.225409836	0.053	19

GREECE (8.11e+36)

Locus	p	q	value	1/value
Amelogenin	0	0	0.000	1
D1S1656	0.115385	0.0649038	0.015	67
TPOX	0.110655738	0.026639344	0.006	170
D2S441	0.382212	0.0504808	0.039	26
D2S1338	0.09375	0.115385	0.022	46
D3S1358	0.245192	0.257212	0.126	8
D4S2408	0.317622951	0.317622951	0.103	10
FGA	0.161058	0.161058	0.027	37
D5S818	0.225409836	0.295081967	0.133	8
CSF1PO	0.024590164	0.420081967	0.021	48
D6S1043	0.163934426	0.129098361	0.042	24
D7S820	0.19057377	0.176229508	0.067	15
D8S1179	0.0480769	0.134615	0.013	77
D9S1122	0.391393443	0.342213115	0.268	4
D10S1248	0.362981	0.362981	0.134	7
TH01	0.216346	0.216346	0.049	21
vWA	0.09375	0.242788	0.046	22

รูปที่ 41 ผลลัพธ์จากการสืบค้นแบบโปรไฟล์ที่ให้รายละเอียดในแต่ละประเทศ

4.3. การแสดงผลข้อมูลภาพรวม

ระบบวิเคราะห์ข้อมูลจากฐานข้อมูลและแสดงออกมาเป็นข้อมูลเชิงสถิติในหลากหลายรูปแบบ ซึ่งการแสดงผลแต่ละรูปแบบก็มีความแตกต่างกัน เช่น แสดงในรูปแบบภาพรวมของข้อมูล หรือแสดงในรูปแบบรายบุคคล ผู้ใช้ทั่วไปสามารถเข้าถึงการแสดงผลที่เป็นภาพรวมเท่านั้น การแสดงผลรายบุคคลสามารถเข้าถึงได้โดยผู้ใช้งานห้องทดลองหรือผู้ดูแลระบบเท่านั้น



รูปที่ 42 การแสดงผลความแปรปรวนทางพันธุกรรม (ก) รายการของโลคัส (ข) ค่าทางสถิติและกราฟ (ค) ความหลากหลายของแอลลีลของโลคัส DXS10074 และผลจากการเลือกแอลลีล 16 (ง) แท็บของข้อมูลดิพลอยด์และแฮพลอยด์

การแสดงผลความแปรปรวนทางพันธุกรรม (Genetic variation) (รูปที่ 42) จะแสดงความถี่แอลลีลของแต่ละโลคัส ผู้ใช้งานสามารถเลือกโครโมโซมและโลคัสที่ต้องการจากทางด้านซ้าย (รูปที่ 42ก) หลังจากผู้ใช้เลือกโลคัสที่ต้องการดูข้อมูลแล้วระบบจะแสดงค่าความถี่แอลลีลในโลคัสนั้นในรูปแบบแผนภูมิแท่งและมีค่าทางสถิติแสดงที่ด้านขวาของหน้าเว็บ (รูปที่ 42ข) ผู้ใช้งานสามารถคลิกที่แท่งข้อมูลที่อยู่ในแผนภูมิแท่งเพื่อให้ระบบแสดงตารางความหลากหลายทางแอลลีลของแอลลีลนั้น ๆ ได้ (รูปที่ 42ค) และเนื่องจาก จีโนไทป์ของแต่ละโครโมโซมมีความแตกต่างกัน โดยโครโมโซมกลุ่มที่เป็นออโตโซม (Autosome) หรือโครโมโซม 1-22 ควบคุมลักษณะต่างๆ ที่ปรากฏของร่างกายยกเว้นเพศ ประกอบไปด้วยจีโนไทป์ที่เป็นดิพลอยด์ (Diploid) เท่านั้น ส่วนโครโมโซมเอ็กซ์ประกอบไปด้วยจีโนไทป์ที่เป็นได้ทั้งดิพลอยด์และแฮพลอยด์ โดยจีโนไทป์ที่มาจากเพศหญิงจะเป็นดิพลอยด์และจีโนไทป์ที่มาจากเพศชายจะเป็นแฮพลอยด์ ทั้งนี้ในทางเทคนิคโครโมโซมเอ็กซ์จะมีจีโนไทป์ที่เป็นแฮพลอยด์เท่านั้นถ้ามาจากเพศชาย แต่ว่ามีข้อยกเว้นสำหรับบางโลคัสที่มีแอลลีลมากกว่า 1 ตำแหน่งจะเป็นดิพลอยด์ ยกตัวอย่างเช่น โลคัส DYF387S1 และ DYS385a-b ซึ่งโลคัสที่มีทั้งจีโนไทป์แบบดิพลอยด์และแฮพลอยด์ตัวเว็บจะอนุญาตให้ผู้ใช้เลือกได้ทั้งสองแท็บ (รูปที่ 42ง) แต่ถ้าโลคัสใดมีดิพลอยด์หรือแฮพลอยด์แบบเดียว ผู้ใช้จะสามารถใช้ได้เพียงแท็บที่เกี่ยวข้องเท่านั้น

ในส่วนของการแสดงค่าทางสถิติผลความแปรปรวนทางพันธุกรรม [61] มีสองลักษณะ คือ ในกรณีที่โลคัสดังกล่าวเป็นดิพลอยด์จะมีการแสดงค่า Polymorphic Information Content (PIC) [62], Match Probability (PM), Discrimination (PD), Power of Exclusion (PE) , Observed

Heterozygosity (Hobs) และ Expected Heterozygosity (Hexp) [63] ส่วนโลคัสที่เป็นแฮพลอยด์ จะแสดงค่า Polymorphic Information Content (PIC), Match Probability (PM) และ Power of Discrimination (PD) เท่านั้น

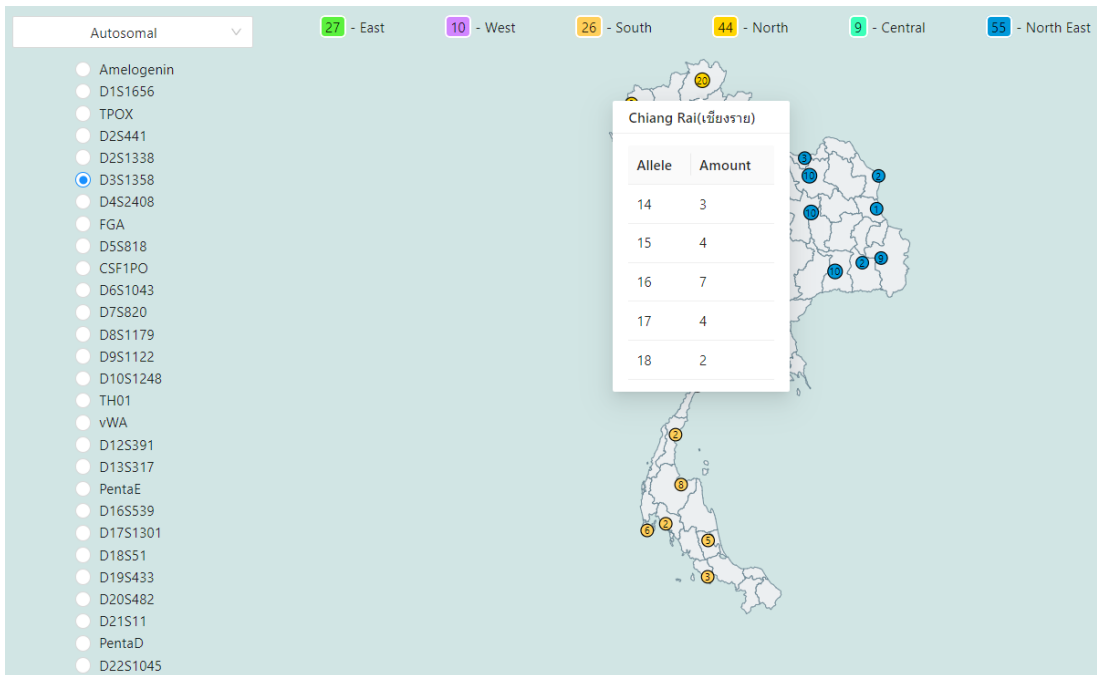
การแสดงผลข้อมูลแอลลีล (Allele detail) (รูปที่ 43) จะแสดงผลความถี่แอลลีล เช่นเดียวกับการแสดงผลความแปรปรวนทางพันธุกรรม แต่จะอยู่ในรูปแบบตารางที่แสดงทุกค่าสถิติ ของทุกแอลลีลในโลคัสนั้น และมีการแยกการแสดงผลระหว่างดิพลอยด์กับแฮพลอยด์ด้วย เช่นเดียวกัน

Allele	Count	Allele Frequency	Allele detail (Export to excel click here)		
			Variant count	Variant allele frequency	Repeat Structure
15	1	0.013158	1	0.013158	[TAGA]2 ctga CAGA [TAGA]7 [CAGA]4 TAGA
16	21	0.276316	21	0.276316	[TAGA]2 ctga CAGA [TAGA]8 [CAGA]4 TAGA
17	7	0.092105	7	0.092105	[TAGA]2 ctga CAGA [TAGA]9 [CAGA]4 TAGA
18	16	0.210526	16	0.210526	[TAGA]2 ctga CAGA [TAGA]10 [CAGA]4 TAGA
19	27	0.355263	1	0.013158	TAGA TTGA CAGA [TAGA]12 [CAGA]4 TAGA
			26	0.342105	[TAGA]2 ctga CAGA [TAGA]11 [CAGA]4 TAGA
20	4	0.052632	4	0.052632	[TAGA]2 ctga CAGA [TAGA]12 [CAGA]4 TAGA

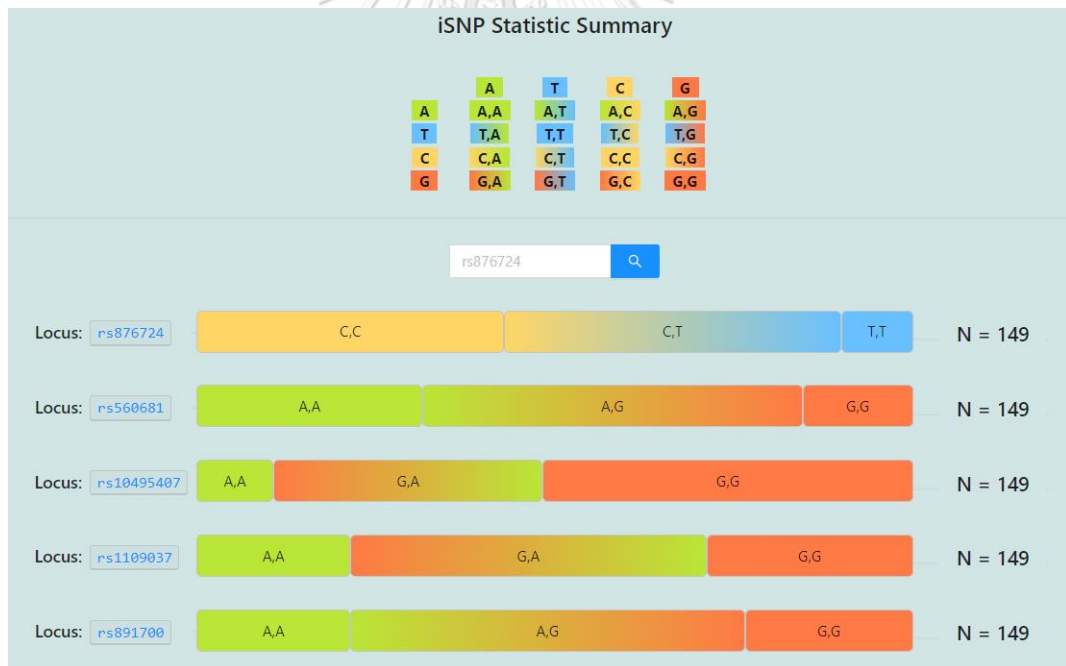
รูปที่ 43 ตารางการแสดงผลข้อมูลแอลลีล

การแสดงผลการกระจายของแอลลีลตามภูมิศาสตร์ (Allele distribution by geographic) (รูปที่ 44) จะแสดงความถี่แอลลีลในรูปแบบแผนที่ โดยมีวงกลมเป็นตัวแทนของแต่ละจังหวัด เมื่อนำเมาส์ไปชี้ที่วงกลมดังกล่าว ระบบจะแสดงตารางแอลลีลและจำนวนตัวอย่างที่พบแอลลีลเหล่านั้น ในจังหวัด ตามข้อมูลที่อยู่ของตัวอย่าง นอกจากนี้ผู้ใช้อังสามารถนำเมาส์ไปชี้ที่สี่เหลี่ยมที่รวมจำนวนของแต่ละภูมิภาค ซึ่งระบบก็จะแสดงตารางเช่นเดียวกันแต่จะรวมแอลลีลของทั้งภูมิภาค

การแสดงผลข้อมูลสรุปทางสถิติของข้อมูลสเนป (iSNP statistic summary) (รูปที่ 45) แสดงข้อมูลสรุปของแต่ละโลคัสของสเนป (SNP locus) ที่ผู้ใช้งานสามารถค้นหาข้อมูลได้โดยใช้ชื่อโลคัสของสเนป ตัวกราฟแท่งจะแสดงข้อมูลสรุปจำนวนของคู่นิวคลีโอไทด์ที่เปลี่ยนไปของข้อมูลตัวอย่างทั้งหมดในฐานข้อมูลจำแนกตามโลคัสของสเนป เมื่อนำเมาส์ไปชี้ในแต่ละแท่งระบบจะแสดงสัดส่วนเปอร์เซ็นต์ของคู่นิวคลีโอไทด์ที่เปลี่ยนของสเนปนั้นๆ



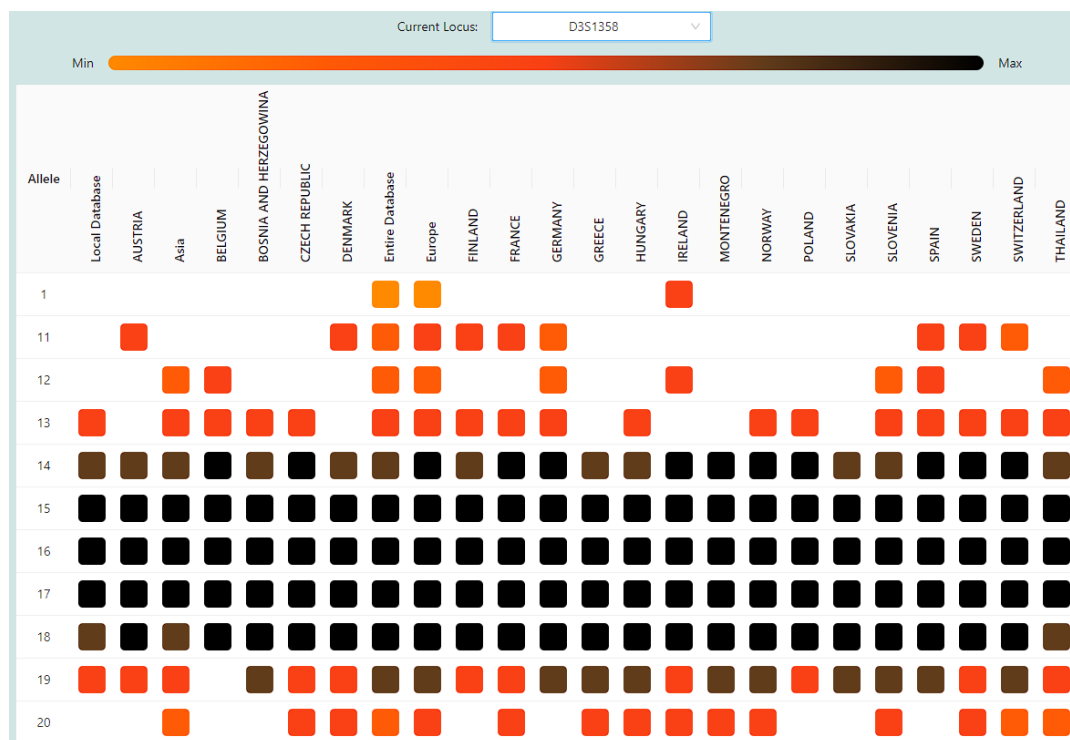
รูปที่ 44 การแสดงผลการกระจายของแอลลีลตามภูมิภาค



รูปที่ 45 การแสดงผลข้อมูลสรุปทางสถิติของข้อมูลสลิป

รูปที่ 46 แสดงการเปรียบเทียบความถี่แอลลีลที่มาจากฐานข้อมูล STRidER กับข้อมูลความถี่แอลลีลจากตัวอย่างทั้งหมดภายในฐานข้อมูล ผู้ใช้สามารถเลือกโลคัสที่ต้องการเปรียบเทียบข้อมูลจากรายการที่แสดงอยู่ด้านบน หลังจากเลือกโลคัส ระบบจะแสดงผลข้อมูลในรูปแบบตารางสองมิติที่แกนตั้งเป็นแอลลีลและแกนนอนเป็นประเทศ แต่ละช่องจะแสดงถึงความถี่แอลลีลโดยใช้เฉดสีและ

แสดงค่าเป็นตัวเลขเมื่อนำเมาส์ไปวางบนช่องนั้น คอลัมน์ Local Database แสดงถึงค่าเฉลี่ยความถี่แอลลีลในฐานข้อมูลของระบบ และคอลัมน์อื่นๆ แสดงความถี่ที่รวบรวมจากฐานข้อมูล STRidER [54]



รูปที่ 46 การแสดงผลการเปรียบเทียบฐานข้อมูล STRidER และระบบที่ออกแบบและพัฒนา

สำหรับการแสดงผลการวิเคราะห์เปรียบเทียบลำดับเบสดีเอ็นเอของข้อมูลเอสทีอาร์ (STR Pattern alignment) (รูปที่ 47) ผู้ใช้สามารถเลือกโครโมโซม โลคัส และแอลลีลที่ต้องการเปรียบเทียบข้อมูลระหว่างกลุ่มตัวอย่าง โดยระบบจะแสดงผลลัพธ์ในรูปแบบตารางที่ประกอบด้วย 5 คอลัมน์ ได้แก่ Sample Year, Sample ID, Sequence, Read count และ STR Repeat motifs โดยที่คอลัมน์ Sequence จะเป็นลำดับนิวคลีโอไทด์ของชุดตัวอย่าง และระบบจะไฮไลต์แต่ละโมทิฟ (รูปแบบลำดับนิวคลีโอไทด์) เพื่อให้ง่ายต่อการดู และคอลัมน์ STR Repeat Motifs แสดงลำดับเบสของตัวอย่างในรูปแบบสรุป เนื่องจากข้อมูลการแสดงผลนี้เป็นข้อมูลรายบุคคล ผู้ใช้ทั่วไปจะไม่สามารถเข้าถึงการแสดงผลนี้ได้ ผู้ใช้งานห้องทดลองและผู้ดูแลระบบเท่านั้นที่จะสามารถเข้าถึงข้อมูลการแสดงผลนี้ได้ ในด้านบนของตารางแสดง Referenced Pattern ที่เป็นรูปแบบอ้างอิงที่ระบบใช้ในการวิเคราะห์เปรียบเทียบลำดับเบสดีเอ็นเอรวมถึงมีข้อมูลทิศทางของรูปแบบอ้างอิงดังกล่าว ซึ่งอาจเป็นแบบ Forward (ลำดับเบสดีเอ็นเออ้างอิงอยู่ในทิศทางจากซ้ายไปขวา หรือจาก 5' ไปยัง 3') หรือ Reverse (ลำดับเบสดีเอ็นเออยู่ในทิศทางจากขวามาซ้าย หรือจาก 3' ไปยัง 5')

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Sample ID	Sample Year	Allele	Summary	Sequence																
2	58D261-F	2017	16.0	[TGCC]6 [TTCC]10	TGCC	TGCC	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC
3	59D087-F	2017	16.0	[TGCC]7 [TTCC]9	TGCC	TGCC	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC
4	59D108-F	2017	16.0	[TGCC]7 [TTCC]9	TGCC	TGCC	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC
5	60D119-M	2018	16.0	[TGCC]7 [TTCC]9	TGCC	TGCC	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC
6	60D184-M	2017	17.0	[TGCC]5 [TTCC]12	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC
7	062F	2016	17.0	[TGCC]6 [TTCC]11	TGCC	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC
8	073F	2016	17.0	[TGCC]6 [TTCC]11	TGCC	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC
9	115F	2016	17.0	[TGCC]6 [TTCC]11	TGCC	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC
10	277F	2016	17.0	[TGCC]6 [TTCC]11	TGCC	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC
11	292M	2016	17.0	[TGCC]6 [TTCC]11	TGCC	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC
12	58D026-F	2017	17.0	[TGCC]6 [TTCC]11	TGCC	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC
13	58D076-F	2017	17.0	[TGCC]6 [TTCC]11	TGCC	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC
14	58D111-F	2017	17.0	[TGCC]6 [TTCC]11	TGCC	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC
15	58D124-F	2017	17.0	[TGCC]6 [TTCC]11	TGCC	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC
16	58D136-F	2017	17.0	[TGCC]6 [TTCC]11	TGCC	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC
17	58D219-F	2017	17.0	[TGCC]6 [TTCC]11	TGCC	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC
18	59D010-F	2017	17.0	[TGCC]6 [TTCC]11	TGCC	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC
19	59D014-F	2017	17.0	[TGCC]6 [TTCC]11	TGCC	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC
20	59D021-F	2017	17.0	[TGCC]6 [TTCC]11	TGCC	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC
21	59D027-F	2017	17.0	[TGCC]6 [TTCC]11	TGCC	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC
22	59D104-F	2017	17.0	[TGCC]6 [TTCC]11	TGCC	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC
23	59D115-F	2017	17.0	[TGCC]6 [TTCC]11	TGCC	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC
24	59D122-F	2017	17.0	[TGCC]6 [TTCC]11	TGCC	TGCC	TGCC	TGCC	TGCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC	TTCC

รูปที่ 49 ไฟล์ข้อมูลผลการวิเคราะห์เปรียบเทียบลำดับเบสดีเอ็นเอที่นำออกในรูปแบบไฟล์เอ็กเซล

4.4. ส่วนต่อประสานโปรแกรมประยุกต์สำหรับบริการภายนอก

ระบบที่นำเสนอมีส่วนต่อประสานโปรแกรมประยุกต์สำหรับบริการภายนอก (Public API) ให้ระบบภายนอกสามารถใช้ข้อมูลในระบบที่นำเสนอได้ โดยมีหน้าเว็บที่อธิบายการใช้งานในรูปแบบของ Swagger (รูปที่ 50) ซึ่งสามารถเข้าถึงได้ผ่านเบราว์เซอร์ที่ /swagger-ui/index.html โดยระบบให้บริการ 3 บริการหลัก ได้แก่

1. การสืบค้นแบบโปรไฟล์

เป็นบริการที่ให้ระบบภายนอกสามารถสืบค้นแบบโปรไฟล์ โดยประกอบด้วยส่วนต่อประสานโปรแกรมประยุกต์ 3 เส้นทาง ประกอบด้วย

- GET /public-api/core-loci/countries วิทยาลัย

ใช้สำหรับการเรียกดูชื่อของชุดโลกส์หลักมาตรฐานที่มีอยู่ในระบบ (รวมของทุกประเทศ) โดยผลลัพธ์จะได้ออกมาเป็นรายการของชื่อของชุดโลกส์หลักมาตรฐานทั้งหมดในระบบ

- GET /public-api/core-loci

ใช้สำหรับการเรียกดูโลกส์ของชุดโลกส์หลักมาตรฐานที่มีอยู่ในระบบ โดยต้องระบุชื่อของชุดโลกส์หลักมาตรฐาน ในรูปแบบของคิวรีพารามิเตอร์ (Query parameter) ชื่อว่า country เช่น /public-api/core-loci?country=USA โดยผลลัพธ์จะได้ออกมาเป็นชุดรายการของโลกส์ และ isRequired โดย isRequired ใช้สำหรับบอกว่าโลกส์แต่ละรายการจำเป็นสำหรับการสืบค้นแบบโปรไฟล์ใน country ที่ระบุหรือไม่

- POST /public-api/profile-search/search

ใช้สำหรับการสืบค้นแบบโปรไฟล์โดยผลลัพธ์ที่ได้อยู่ในรูปแบบอ็อบเจกต์ที่ประกอบด้วย ตารางสรุปของแต่ละประเทศ และค่าผลลัพธ์สุดท้ายของการสืบค้นนี้

2. การให้ข้อมูลสรุปเชิงสถิติตามโลคัส

เป็นบริการที่ให้ระบบภายนอกสามารถขอข้อมูลสถิติโดยระบุจากโลคัสที่ต้องการได้ ประกอบด้วยส่วนต่อประสานโปรแกรมประยุกต์ 2 เส้นทาง ดังนี้

- GET /public-api/loci/all

ใช้สำหรับดึงข้อมูลรายการของโลคัสทั้งหมดที่สามารถดูข้อมูลสถิติได้ของระบบ โดยผลลัพธ์ เป็นรายการของชื่อโลคัสทั้งหมดและถูกแบ่งตามโครโมโซม

- GET /public-api/forensiq-sequences/graph

ใช้สำหรับดึงข้อมูลทางสถิติจากโลคัสที่ต้องการโดยต้องระบุโลคัสที่ต้องการในรูปแบบของควิรีพารามิเตอร์ชื่อ locus เช่น /public-api/forensiq-sequences/graph?locus=FGA โดยผลลัพธ์ เป็นข้อมูลของกราฟแท่งความถี่แอลลีล ค่าสถิติของโลคัส และลำดับเบสดีเอ็นเอจำแนกตามแอลลีล ของทุกรูปแบบที่ไม่ซ้ำกัน

3. การให้ข้อมูลสรุปเชิงสถิติตามภูมิภาค

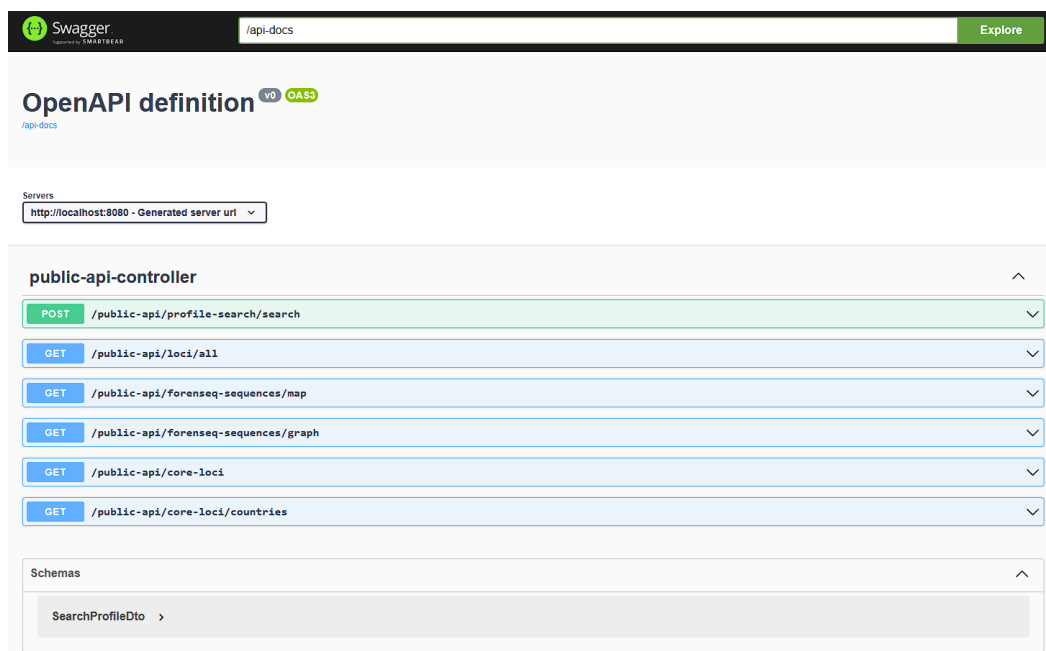
เป็นบริการที่ให้ระบบภายนอกสามารถขอข้อมูลความถี่แอลลีลจำแนกตามจังหวัดโดย สามารถระบุจากโลคัสที่ต้องการได้ ประกอบด้วยส่วนต่อประสานโปรแกรมประยุกต์ 2 เส้นทางดังนี้

- GET /public-api/loci/all

ใช้สำหรับการดึงข้อมูลรายการของโลคัสทั้งหมดที่สามารถดูข้อมูลสถิติได้ของระบบ โดยผลลัพธ์เป็นรายการของชื่อโลคัสทั้งหมดและถูกแบ่งตามโครโมโซม

- GET /public-api/forensiq-sequences/map

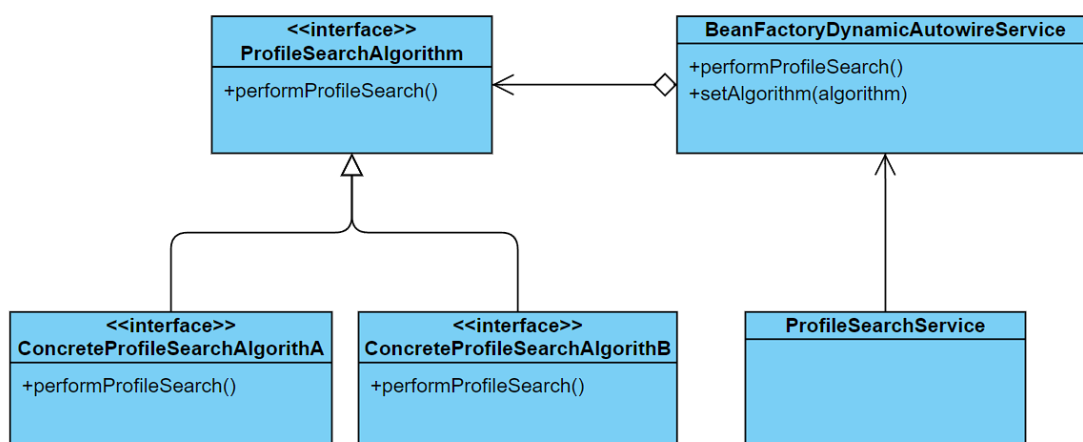
ใช้สำหรับการดึงข้อมูลทางสถิติจากโลคัสที่ต้องการโดยต้องระบุโลคัสที่ต้องการในรูปแบบของควิรีพารามิเตอร์ชื่อ locus เช่น /public-api/forensiq-sequences/map?locus=FGA โดยผลลัพธ์เป็นข้อมูลของกราฟแผนที่ของความถี่แอลลีลแบ่งตามจังหวัด



รูปที่ 50 หน้าเว็บอธิบายการใช้งานส่วนต่อประสานโปรแกรมประยุกต์สำหรับบริการภายนอก

4.5. แนวทางการพัฒนาต่อยอดอัลกอริทึมวิเคราะห์ข้อมูลในระบบ

สำหรับการสืบค้นแบบโปรไฟล์ระบบอนุญาตให้ผู้พัฒนาสามารถเพิ่มวิธีการวิเคราะห์ข้อมูล หรือเพิ่มบริการใหม่ที่ใช้ข้อมูลภายในระบบได้โดยใช้ภาษาจาวา ระบบใช้แบบรูป Strategy (รูปที่ 51) เพื่อให้ผู้ใช้สามารถเปลี่ยนอัลกอริทึมการวิเคราะห์ขณะทำงานได้



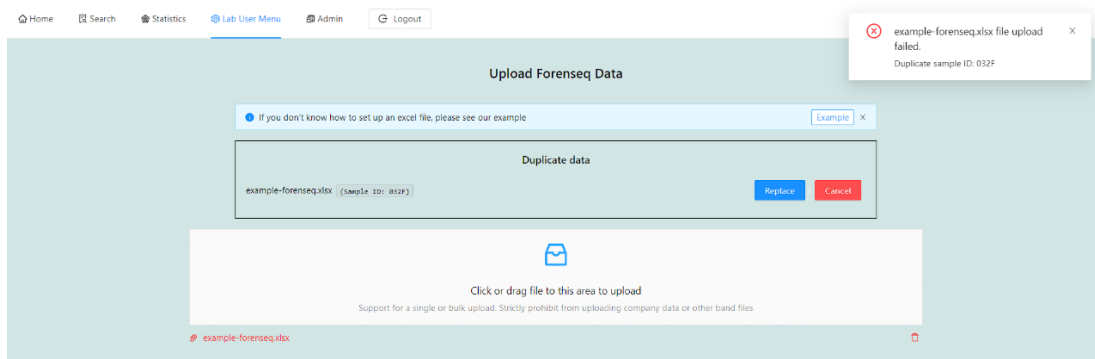
รูปที่ 51 การประยุกต์ใช้แบบรูป Strategy เพื่อสนับสนุนการเพิ่มรูปแบบการสืบค้นข้อมูลแบบโปรไฟล์

วิธีการเพิ่มอัลกอริทึมสำหรับการสืบค้นแบบโปรไฟล์มีขั้นตอนดังนี้

1. สร้างคลาสของภาษาจาวาโดยที่คลาสดังกล่าวต้องอิมพลีเมนต์อินเตอร์เฟซ *ProfileSearchAlgorithm* จากแพ็คเกจ *th.ac.chula.fims.services.interfaces.searchprofile.ProfileSearchAlgorithm* และต้องเขียนทับ (Override) เมธอดชื่อ *performProfileSearch*
2. คลาสดังกล่าวต้องประกาศ Annotation *@Service* ไว้ที่หัวคลาสและใส่อาร์กิวเมนต์ที่เป็นชื่ออัลกอริทึมตามด้วย *ProfileSearchAlgorithm* ตัวอย่างเช่น ถ้าชื่ออัลกอริทึมคือ *norman* จะประกาศ *@Service("normanProfileSearchAlgorithm")*
3. โดยสามารถเรียกใช้ผ่านส่วนติดต่อผู้ใช้งานของเว็บผ่าน POST เมธอดที่มีที่อยู่คือ */api/profile-search/search* และตามด้วยใส่ชื่ออัลกอริทึมเป็นพารามิเตอร์ของคิวรี (Query parameter) เช่น */api/profile-search/search?algorithm=norman*

4.6. การจัดการข้อมูลของผู้ใช้งานห้องทดลอง

นอกจากผู้ใช้งานห้องทดลองสามารถใช้งานฟังก์ชันต่างๆ ได้เหมือนกับผู้ใช้แล้ว ผู้ใช้งานห้องทดลองยังสามารถนำเข้าข้อมูลเอสทีอาร์ของตัวอย่างที่อยู่ในรูปแบบไฟล์เอ็กเซลได้ เมื่อผู้ใช้งานห้องทดลองนำเข้าข้อมูลทีอาร์ ระบบจะตรวจสอบ sample ID ของไฟล์ว่าตรงกับที่มีในระบบหรือไม่ ในกรณีที่มีข้อมูลซ้ำ ผู้ใช้งานสามารถยกเลิกการนำเข้าข้อมูลของตัวอย่างนั้นหรือเขียนทับข้อมูลเดิมได้อย่างใดอย่างหนึ่ง (รูปที่ 52)



รูปที่ 52 หน้าเว็บการนำเข้าข้อมูลที่ซ้ำซ้อนสู่ระบบ

หลังจากนำเข้าข้อมูลเอสทีอาร์ของตัวอย่างแล้ว ผู้ใช้งานห้องทดลองสามารถนำเข้าข้อมูลส่วนบุคคลของแต่ละตัวอย่างได้ (รูปที่ 53) โดยระบบจะนำข้อมูลในแต่ละแถวไปเชื่อมโยงกับข้อมูลเอสทีอาร์ของตัวอย่างเดียวกันในฐานข้อมูล โดยอ้างอิงจากคอลัมน์ SampleId ค่าที่เป็นไปได้สำหรับคอลัมน์ Gender คือ MALE และ FEMALE สำหรับคอลัมน์ Province, Region, Country และ Race ระบบจะนำค่าในแต่ละคอลัมน์ไปเชื่อมโยงกับข้อมูลภายในระบบที่ต้องมีการนำเข้าข้อมูลมาก่อนหน้าโดยผู้ดูแลระบบ ในกรณีที่ระบบหาค่าในคอลัมน์ใดไม่เจอ ระบบจะเว้นการเชื่อมโยงคอลัมน์นั้นไว้ ยกตัวอย่างเช่น ในแถวข้อมูลที่มี SampleId เป็น 038M ระบบสามารถค้นหาค่า Thai จากคอลัมน์ Race ค่า Thailand จากคอลัมน์ Country และค่า Central จากคอลัมน์ Region ได้ แต่ไม่สามารถหาค่า Bangkok ของคอลัมน์ Province ได้ ระบบจะเชื่อมโยงข้อมูล 038M ให้ทั้งหมดทุกคอลัมน์ยกเว้นคอลัมน์ Province ที่ไม่พบข้อมูล โดยการค้นหาค่าจะเป็นแบบการไม่บังคับใช้ตัวพิมพ์เล็ก/ตัวพิมพ์ใหญ่ (Case-insensitive)

	A	B	C	D	E	F
1	Gender	Province	Region	Country	Race	SampleId
2	MALE	Bangkok	Central	Thailand	Thai	038M
3	FEMALE	Bangkok	Central	Thailand	Thai	046M

รูปที่ 53 ไฟล์เอ็กเซลข้อมูลส่วนบุคคลของตัวอย่างในระบบ

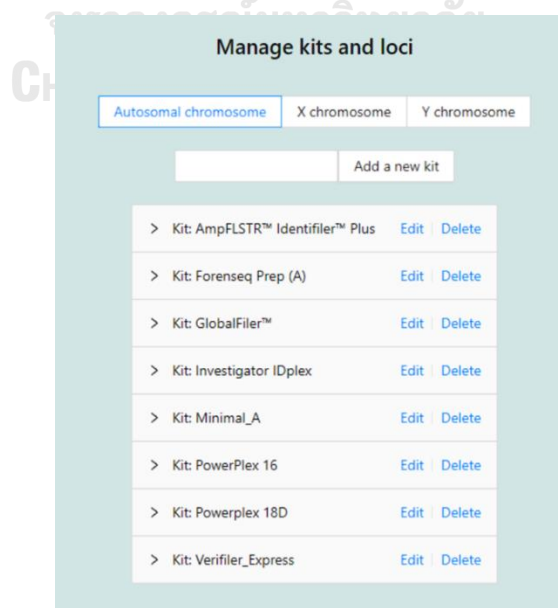
4.7. การจัดการระบบโดยผู้ดูแลระบบ

ผู้ดูแลระบบสามารถใช้งานฟังก์ชันได้เหมือนกับผู้ใช้งานห้องทดลอง นอกจากนี้ผู้ดูแลระบบยังสามารถปรับแต่งระบบได้ ตัวอย่างเช่น การเปลี่ยนการแสดงผลของแผนที่ การจัดการผู้ใช้ในระบบ การจัดการชุดคิทและโลคัสที่เกี่ยวข้อง โดยที่การติดตั้งครั้งแรก ระบบจะมีรหัสเริ่มต้นของผู้ดูแลระบบ จัดเตรียมไว้ให้เพื่อเข้าสู่ระบบครั้งแรกและเปลี่ยนรหัสผ่าน

ผู้ดูแลระบบสามารถแก้ไขแผนที่ที่ใช้แสดงผลการกระจายของแอลลีลตามภูมิศาสตร์ (Allele distribution by geographic) (รูปที่ 44) โดยสามารถนำเข้าแผนที่ในรูปแบบ GeoJSON และสามารถแก้ไขการแสดงจุดศูนย์กลางโดยใช้ระบบละติจูดและลองจิจูด รวมทั้งสามารถแก้ไขการย่อขยายแผนที่เพื่อให้สามารถใช้งานได้กับแผนที่ของแต่ละประเทศที่ผู้ดูแลระบบนำเข้ามาใหม่ (รูปที่ 54) โดยการแสดงผลนี้ใช้ไลบรารีของรีแอคชื่อ React Simple Maps [64] ซึ่งการใช้ไฟล์แผนที่ที่มีรายละเอียดมากจะทำให้การแสดงผลตอบสนองช้า ผู้ดูแลระบบสามารถดูตัวอย่างไฟล์แผนที่ได้จาก [65] ในกรณีที่ผู้ดูแลระบบมีไฟล์แผนที่และต้องการจะแก้ไขหรือลดความละเอียดของไฟล์แผนที่ก็สามารถทำได้ผ่าน [66]



รูปที่ 54 หน้าเว็บสำหรับการจัดการแผนที่

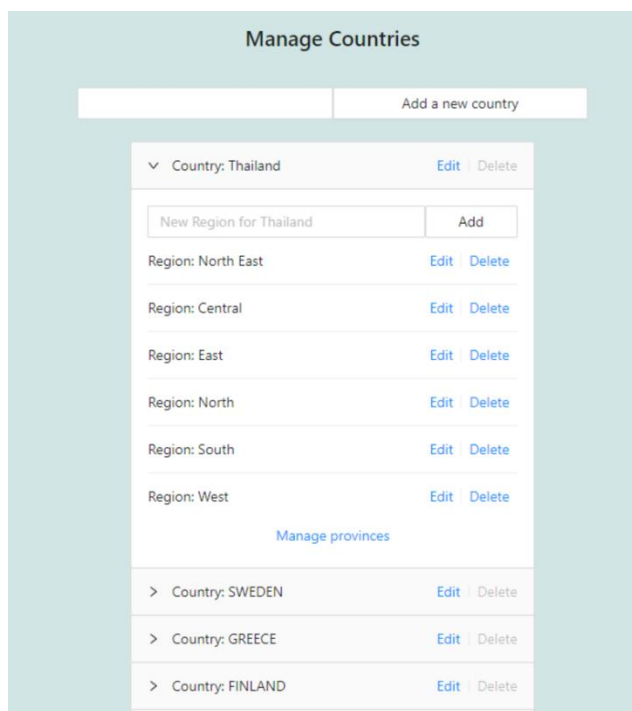


รูปที่ 55 หน้าเว็บการจัดการโลคัสและชุดคิท

นอกจากนี้ผู้ดูแลระบบยังสามารถเพิ่มโลโก้และชุดคิใหม่ (รูปที่ 55) เข้าสู่ระบบได้ มีสิทธิในการจัดการเชื้อชาติ (Race) จังหวัด (Province) (รูปที่ 56) ภูมิภาค (Region) และประเทศ (Country) (รูปที่ 57) ในระบบ ซึ่งข้อมูลเหล่านี้จะถูกค้นหาเมื่อมีการนำข้อมูลส่วนบุคคล (รูปที่ 53) และจะถูกนำไปแสดงผลในส่วนของการแสดงผลการกระจายของแอลลีลตามภูมิภาค

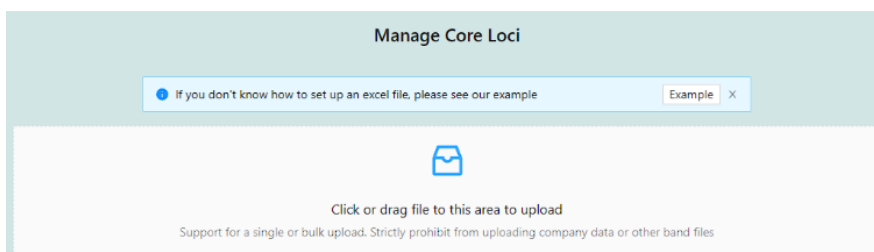
Manage Provinces						
Thailand (ID: 1)						
Add bulk						
Edit Delete						
<input type="checkbox"/> v	ID	Province	Native name	Latitude	Longitude	Region
<input type="checkbox"/>	78	Amnat Charoen	อำนาจเจริญ	15.8585	104.6288	North East
<input type="checkbox"/>	79	Ang Thong	อ่างทอง	14.5883	100.4528	Central
<input type="checkbox"/>	80	Bangkok	กรุงเทพ	13.7539	100.5014	Central
<input type="checkbox"/>	81	Bueng Kan	บึงกาฬ	18.363	103.6519	North East
<input type="checkbox"/>	82	Buri Ram	บุรีรัมย์	14.993	103.1029	North East
<input type="checkbox"/>	83	Chachoengsao	ฉะเชิงเทรา	13.6904	101.0779	East
<input type="checkbox"/>	84	Chai Nat	ชัยนาท	15.1863	100.1235	Central
<input type="checkbox"/>	85	Chaiyaphum	ชัยภูมิ	15.8104	102.0288	North East
<input type="checkbox"/>	86	Chanthaburi	จันทบุรี	12.6096	102.1044	East

รูปที่ 56 หน้าเว็บสำหรับการจัดการจังหวัด



รูปที่ 57 หน้าเว็บสำหรับการจัดการประเทศและภูมิภาค

ผู้ดูแลระบบยังสามารถจัดการส่วนการสืบค้นแบบโปรไฟล์ได้ โดยสามารถเพิ่มหรือลดโลคัสหลักของแต่ละประเทศได้ (รูปที่ 58) สามารถแก้ไขข้อมูลทางสถิติในฐานข้อมูลที่นำมาใช้คำนวณในการสืบค้นได้ ในส่วนของผลการวิเคราะห์เปรียบเทียบลำดับเบสดีเอ็นเอ (Pattern alignment) ผู้ดูแลระบบมีหน้าที่ในการนำเข้รูปแบบเอสทีอาร์อ้างอิง (Reference STR patterns) (รูปที่ 33) เพื่อการวิเคราะห์เปรียบเทียบลำดับเบสดีเอ็นเอ และผู้ดูแลระบบมีหน้าที่ในการเรียกใช้งานการวิเคราะห์ ซึ่งผลลัพธ์จะถูกจัดเก็บเข้าฐานข้อมูลโดยอัตโนมัติ เมื่อผู้ใช้งานห้องทดลองนำเข้ข้อมูลตัวอย่างใหม่หรือผู้ดูแลระบบนำเข้รูปแบบอ้างอิงใหม่เข้าสู่ระบบ ผู้ดูแลระบบสามารถเรียกใช้การวิเคราะห์เปรียบเทียบลำดับเบสดีเอ็นเอใหม่อีกครั้ง เพื่อให้ได้ผลลัพธ์ที่ตรงกับข้อมูลในปัจจุบัน



รูปที่ 58 หน้าเว็บการจัดการโลคัสหลักของการสืบค้นแบบโปรไฟล์

4.8. ที่มาของข้อมูลตัวอย่างภายในระบบ

ข้อมูลตัวอย่างในระบบเป็นข้อมูลที่ไม่ระบุตัวตนทั้งหมด 125 ตัวอย่าง ข้อมูลที่นำมาใช้ในระบบข้อมูลจะถูกสลับจีโนไทป์แบบสุ่มในโลคัสเดียวกันทั้ง 125 ตัวอย่าง โดยที่รายละเอียดของแอลลีลในแต่ละจีโนไทป์จะยังคงสัมพันธ์กับจีโนไทป์ตัวเดิม ตัวอย่างเช่น ในโลคัสเอ ตัวอย่าง ก มีจีโนไทป์-1ก ตัวอย่าง ข มีจีโนไทป์-1ข และตัวอย่าง ค มีจีโนไทป์-1ค การสลับแบบสุ่มของจีโนไทป์ทั้ง 3 ตัวอย่างอาจจะได้ผลลัพธ์เป็นดังนี้ ตัวอย่าง ก มีจีโนไทป์-1ข ตัวอย่าง ข มีจีโนไทป์-1ค และตัวอย่าง ค มีจีโนไทป์-1ก และทำซ้ำแบบนี้ในทุก ๆ โลคัส ซึ่งการสลับในรูปแบบดังกล่าวทำให้ข้อมูลตัวอย่างทั้ง 125 ตัวอย่างเป็นข้อมูลใหม่ทั้งหมดที่ไม่สะท้อนถึงตัวบุคคลในชีวิตจริง แต่ยังคงความถูกต้องของข้อมูลไว้

4.9. การตรวจสอบความถูกต้องของการวิเคราะห์ข้อมูลภายในระบบ

ในการตรวจสอบความถูกต้องของค่าทางสถิติทั้งหมด ได้แก่ Polymorphic Information Content (PIC), Match Probability (PM), Discrimination (PD) และ Power of Exclusion (PE) ผู้พัฒนาระบบใช้ตัวอย่างข้อมูลทั้งหมด 125 ตัวอย่าง ในการคำนวณค่าสถิติดังกล่าวและนำมาเทียบกับค่าทางสถิติที่ใช้ตัวอย่างข้อมูลเดียวกันแต่ถูกคำนวณโดยโปรแกรม STRAF [8] ซึ่งผลการเปรียบเทียบ ระหว่างผลลัพธ์ที่ได้จากระบบที่พัฒนาขึ้น ได้ผลตรงกับผลลัพธ์ทั้งหมดที่ได้จากโปรแกรม STRAF ยกเว้นค่า Gene Diversity (GD) ที่โปรแกรม STRAF ใช้สมการจาก [67] ส่วนระบบที่พัฒนาขึ้นใช้สมการจาก [68] ทำให้ผลลัพธ์ที่ได้ออกมาแตกต่างกัน จึงละเว้นการเปรียบเทียบค่า Gene Diversity สำหรับการตรวจสอบความถูกต้องของค่าทางสถิติ

สำหรับการตรวจสอบการวิเคราะห์เปรียบเทียบลำดับเบสดีเอ็นเอ ได้ทำการสุ่มตัวอย่าง และนำมาเปรียบเทียบกับผลลัพธ์ที่ได้จาก STRait Razor [9] ในทุก ๆ โลคัส และเปรียบเทียบเฉพาะส่วนที่เป็นลำดับซ้ำเท่านั้น (Repeat Region) ไม่เปรียบเทียบในส่วนที่เป็น แพลงค์กิง (Flanking) (ส่วนต้นก่อนและส่วนหลังท้ายลำดับซ้ำ) เนื่องจากข้อมูลเอสทิอาร์ของตัวอย่างที่นำเข้าอาจมีความแตกต่างในส่วนของแพลงค์กิงเมื่อเทียบกับผล STRait Razor ตั้งแต่ต้น ผลการเปรียบเทียบการวิเคราะห์เปรียบเทียบลำดับเบสดีเอ็นเอ ที่ได้จากระบบ ตรงกับที่ได้จาก STRait Razor ทั้งหมด

4.10. เทคโนโลยีที่ใช้ในการพัฒนา

เทคโนโลยีที่ใช้ในการพัฒนาระบบประกอบด้วย 4 ส่วน ได้แก่ ระบบหน้าบ้าน ระบบหลังบ้าน ฐานข้อมูล และการเปิดใช้งานระบบ

ระบบหน้าบ้านใช้ไลบรารีชื่อ รีแอกต์ [69] ของภาษาจาวาสคริป (JavaScript) เพื่อช่วยในการสร้างส่วนต่อประสานกับผู้ใช้ ซึ่งไลบรารีนี้ใช้การ ดอมจำลอง (Virtual DOM) เพื่อสืบค้นการเปลี่ยนแปลงในหน้าเว็บและนำมาเทียบกับดอมจริง จากนั้นทำการอัปเดตเฉพาะส่วนที่มีการเปลี่ยนแปลง ทำให้บราวเซอร์สามารถวาด (Render) หน้าเว็บได้รวดเร็วขึ้น นอกจากนี้ยังมีการใช้แบบรูปคอมโพเนนท์ทำให้คอมโพเนนท์ที่ถูกสร้างก่อนหน้าสามารถนำมาใช้ซ้ำในคอมโพเนนท์ใหม่ ๆ ได้

ระบบหลังบ้านใช้เฟรมเวิร์ค Spring Boot [70] ของภาษาจาวา (Java) เฟรมเวิร์คนี้ใช้ทรัพยากรน้อย (Light weight) และรวดเร็ว ช่วยในการเขียนโค้ดที่สามารถปรับขยายระบบได้ (Scalability) มีการใช้แบบรูปติเพนเดนซีอินเจกชันเพื่อลดความขึ้นต่อกัน (Loose coupling) ระหว่างคลาส มีปลั๊กอินที่ช่วยในการทดสอบ ไลบรารีที่ช่วยในการจัดการฐานข้อมูล และไลบรารีที่ใช้ในการรักษาความปลอดภัยให้กับระบบที่เป็นมาตรฐาน รวมทั้งมีการเขียนโปรแกรมที่เป็นแบบเชิงวัตถุ ซึ่งทำให้โค้ดบำรุงรักษาได้ง่าย (Maintainable)

ระบบจัดการฐานข้อมูลที่ใช้ในระบบที่นำเสนอ คือ MySQL [71] ซึ่งเป็นระบบฐานข้อมูลเชิงสัมพันธ์ (Relational Database) ที่ได้รับความนิยมใช้ จากบริษัทต่างๆ เช่น Facebook [72], WordPress [73] ทั้งนี้ระบบฐานข้อมูลดังกล่าวมีความน่าเชื่อถือ (Reliable) ความปลอดภัย (Secure) และมีประสิทธิภาพสูง (High performance) ระบบหนึ่ง

ในส่วนการนำไปใช้งาน ระบบที่พัฒนาขึ้นใช้แนวคิด containerized โดยใช้ดี็อกเกอร์เอนจิน (Docker engine) ในการรันระบบ ซึ่งช่วยให้ระบบสามารถรันบนสภาพแวดล้อม (Environment) ที่แตกต่างกันได้โดยง่ายและรวดเร็วสำหรับการเปิดใช้งาน รวมทั้งช่วยให้ระบบสามารถปรับขนาด (Scalability) ยืดหยุ่นตามจำนวนการใช้งาน

บทที่ 5

สรุปผลการวิจัย

5.1. การอภิปรายผล

วิทยานิพนธ์ฉบับนี้นำเสนอระบบสนับสนุนการจัดเก็บและวิเคราะห์ข้อมูลลำดับซ้ำเรียงต่อเนื่องแบบสั้นที่ได้จากเทคโนโลยีเอ็นจีเอสสำหรับนิติวิทยาศาสตร์ ซึ่งอนุญาตให้ผู้ใช้งานสามารถจัดการและจัดเก็บข้อมูลเอสทีอาร์ได้อย่างมีประสิทธิภาพ นอกจากนี้ยังมีระบบการสนับสนุนการวิเคราะห์ข้อมูลแบบอัตโนมัติที่ช่วยให้ผู้ใช้งานสามารถแปลผลและแสดงผลข้อมูลได้สะดวก โดยระบบที่นำเสนอช่วยแก้ปัญหาต่างๆ ดังนี้

ในกรณีที่ผู้ใช้งานมีข้อมูลเอสทีอาร์ที่ได้จาก ForenSeq และต้องการคำนวณค่าทางสถิติของกลุ่มประชากร ผู้ใช้งานสามารถใช้โปรแกรม STRAF ได้ แต่ผู้ใช้งานต้องทำการแปลงข้อมูลทั้งหมดให้อยู่ในรูปแบบที่ STRAF กำหนด ซึ่งกระบวนการแปลงข้อมูลนี้อาจจะใช้เวลาหรือเกิดข้อผิดพลาดขึ้นได้ โดยเฉพาะกับกรณีที่มีจำนวนข้อมูลจำนวนมาก นอกจากนี้ ถ้ามีข้อมูลเอสทีอาร์เพิ่มเข้ามาในระบบ ผู้ใช้งานต้องทำการกระบวนการแปลงข้อมูลใหม่ทั้งหมด แต่ในระบบที่นำเสนอ เมื่อมีการอัปเดตข้อมูลเข้าสู่ระบบ ระบบจะทำการคำนวณและวิเคราะห์ข้อมูลให้ใหม่โดยอัตโนมัติ ซึ่งทำให้ประหยัดเวลาและลดความผิดพลาดที่อาจจะเกิดขึ้นจากมนุษย์

ในอีกกรณีหนึ่งซึ่งผู้ใช้งานต้องการวิเคราะห์เปรียบเทียบลำดับเบสดีเอ็นเอในกลุ่มประชากร ผู้ใช้งานสามารถใช้ STRait Razor ได้เช่นกัน แต่ผู้ใช้งานต้องใช้ไฟล์ FASTQ และนำไฟล์ดังกล่าวมาวิเคราะห์แปลผลข้อมูลผ่านคอมมานด์ไลน์ (Command line) ซึ่งจะได้ผลลัพธ์ที่อยู่ในรูปแบบของไฟล์ข้อความ และนำไฟล์ข้อความดังกล่าวอัปเดตเข้า Excel-based workbook เพื่อแสดงผล ในทางกลับกัน ระบบที่นำเสนออนุญาตให้ผู้ใช้งานสามารถอัปเดตไฟล์ ForenSeq sample detail reports ที่เป็นข้อมูลของเอสทีอาร์ได้โดยตรง ซึ่งหลังการอัปเดตข้อมูลเข้าสู่ระบบ ผู้ใช้งานสามารถเรียกใช้ฟังก์ชันการวิเคราะห์ผ่านระบบได้โดยตรง โดยไม่จำเป็นต้องใช้หลากหลายโปรแกรม

การออกแบบของระบบที่นำเสนอใช้หลักการทางคอมพิวเตอร์ในการพัฒนา เช่น การเขียนโปรแกรมเชิงวัตถุ การใช้แบบรูป เพื่อให้ระบบที่นำเสนอมีความเสถียร ง่ายต่อการต่อยอดและการบำรุงรักษาในอนาคต รวมถึงระบบที่นำเสนออยู่ในรูปแบบของดีออกเกอร์ที่มีสภาพแวดล้อมพร้อมสำหรับการติดตั้งและนำไปใช้งานในระบบปฏิบัติการต่างๆ ซึ่งผู้ใช้งานสามารถดาวน์โหลดและติดตั้งได้ทั้งภายในองค์กร หรือบนคลาวด์ โดยที่ไม่ต้องห่วงเรื่องความเข้ากันได้ของระบบ นอกจากนี้ระบบที่

นำเสนอมีการกำหนดสิทธิผู้ใช้งานตามบทบาท (Role based access control) เพื่อให้ผู้ใช้งานเข้าถึงข้อมูลที่สอดคล้องกับบทบาทเท่านั้น

หนึ่งในข้อจำกัดของระบบที่นำเสนอคือระบบในเวอร์ชันปัจจุบันยังรองรับข้อมูลจากเทคโนโลยีที่จำกัด เช่น ข้อมูลที่ได้จาก ForenSeq signature prep of Illumina platform ซึ่งอาจจะทำให้ห้องทดลองหรือองค์กรที่ไม่มีสิทธิเข้าถึงเทคโนโลยีดังกล่าวใช้งานระบบได้ อย่างไรก็ตาม แนวทางการวิจัยในอนาคตมีการวางแผนเพื่อรองรับข้อมูลที่ได้จากเทคโนโลยีอื่นเพิ่มเติม

5.2. สรุปผลการวิจัย

วิทยานิพนธ์ฉบับนี้นำเสนอระบบสนับสนุนการจัดเก็บและวิเคราะห์ข้อมูลลำดับซ้ำเรียงต่อเนื่องแบบสั้นที่ผลิตจากเทคโนโลยีเอ็นจีเอสสำหรับนิติวิทยาศาสตร์ ซึ่งได้รับการออกแบบและพัฒนาเพื่ออำนวยความสะดวกให้ผู้ที่ทำงานกับข้อมูลดังกล่าว ด้วยบริการการจัดการข้อมูลเอสทีอาร์ที่ได้จากเทคนิคการหาลำดับเบสยุคใหม่ที่เป็นผลลัพธ์จากเทคโนโลยี ForenSeq

ระบบอนุญาตให้ผู้ใช้งานทั่วไปสามารถสืบค้นและดูข้อมูลภาพรวมของระบบได้ เช่น การดูข้อมูลสรุปเชิงสถิติตามภูมิภาค การดูข้อมูลสรุปเชิงสถิติตามโลคัส การสืบค้นฐานข้อมูลโดยใช้มัลเอสทีอาร์ และการสืบค้นแบบโปรไฟล์ อนุญาตให้ผู้ใช้งานระดับห้องปฏิบัติการสามารถใช้งานในลักษณะเดียวกับผู้ใช้ทั่วไป รวมทั้งสามารถใช้เครื่องมือการสนับสนุนการวิเคราะห์ เช่น เพื่อเพิ่มประสิทธิภาพในการใช้งาน เช่น วิเคราะห์เปรียบเทียบลำดับเบสดีเอ็นเอ เพิ่ม-ลดข้อมูลเอสทีอาร์ เพิ่ม-ลดรายละเอียดรายบุคคลของชุดข้อมูล ระบบอนุญาตให้ผู้ดูแลระบบสามารถใช้งานในลักษณะเดียวกับผู้ใช้งานระดับห้องปฏิบัติการ รวมถึงปรับเปลี่ยนการตั้งค่าและตัวแปรของระบบ เช่น เพิ่ม-ลดชุดคิทของระบบหลักและการสืบค้นแบบโปรไฟล์ เพิ่ม-ลดข้อมูลประเทศ ภูมิภาคและจังหวัด เพิ่ม-ลดรูปแบบทั่วไปที่ใช้ในการวิเคราะห์เปรียบเทียบลำดับเบสดีเอ็นเอ เปลี่ยนแผนที่ในหน้าข้อมูลสรุปเชิงสถิติตามภูมิภาค

นอกจากนี้ระบบถูกออกแบบให้มีความยืดหยุ่นสามารถเพิ่มส่วนการวิเคราะห์และการแสดงผลโดยผู้พัฒนาอื่นได้โดยง่าย เช่น การเพิ่มอัลกอริทึมของการสืบค้นแบบโปรไฟล์ เพิ่มหรือปรับเปลี่ยนส่วนของโลคัสหลักของการสืบค้นแบบโปรไฟล์ และมีส่วนต่อประสานโปรแกรมประยุกต์สำหรับบริการภายนอก เช่น การการสืบค้นแบบโปรไฟล์ ข้อมูลสรุปเชิงสถิติตามโลคัส และข้อมูลสรุปเชิงสถิติตามภูมิภาค ระบบสามารถติดตั้งและใช้งานได้ง่ายผ่านดีออกเกอร์ สามารถใช้งานได้บนเบราว์เซอร์กูเกิลโครม (Google Chrome) และไมโครซอฟต์เอดจ์ (Microsoft Edge)

5.3. แนวทางการวิจัยในอนาคต

แนวทางในการออกแบบและพัฒนาในอนาคต สามารถพัฒนาระบบฐานข้อมูลให้รองรับการเก็บข้อมูลนิติทางวิทยาศาสตร์ประเภทอื่น ๆ เช่น โครมกะโหลก ลายนิ้วมือ เป็นต้น และรองรับข้อมูลที่ได้จากเทคโนโลยีอื่นนอกเหนือจาก ForenSeq รวมทั้งเพิ่มวิธีการวิเคราะห์อื่น ๆ ที่ใช้สำหรับข้อมูลนิติทางวิทยาศาสตร์นั้น ๆ การเชื่อมโยงข้อมูลต่างประเภทเพื่อเป็นหลักฐานสนับสนุนซึ่งกันและกัน รวมทั้งสามารถประยุกต์ใช้แบบจำลองในกลุ่มการเรียนรู้ของเครื่องและการเรียนรู้เชิงลึก เพื่อทำนายฟีโนไทป์จากข้อมูลลำดับซ้ำเรียงต่อเนื่องแบบสั้น





ภาคผนวก ก.

ตารางฐานข้อมูล (Database) ของระบบที่ออกแบบและพัฒนา

พจนานุกรมข้อมูล (Data Dictionary)

1. PERSONS

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
GENDER	Gender	VARCHAR(6)		
COUNTRY_ID	Country ID	INT	FK	COUTRIES
PROVINCE_ID	Province ID	INT	FK	PROVINCES
RACE_ID	Race ID	INT	FK	RACES
REGION_ID	Region ID	INT	FK	REGIONS

2. CONTINENTS

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
CONTINENT	Continent	VARCHAR(255)		

3. COUNTRIES

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
COUNTRY	Country	VARCHAR(255)		

4. REGIONS

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
REGION	Region name	VARCHAR(255)		
COUNTRY_ID	Country ID	INT	FK	COUNTRIES

5. PROVINCES

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
REGION_ID	Region ID	INT	FK	REGIONS
LATITUDE	Latitude	DOUBLE		
LONGITUDE	Longitude	DOUBLE		
NATIVE_NAME	Native province name	VARCHAR(255)		
PROVINCE	Province name	VARCHAR(255)		

จุฬาลงกรณ์มหาวิทยาลัย

6. RACES

CHULALONGKORN UNIVERSITY

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
RACE	Race	VARCHAR(255)		

7. SAMPLES

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
SAMPLE_ID	Sample ID	VARCHAR(255)	FK	SAMPLES
SAMPLE_YEAR	Sample year	INT		
PERSON_ID	Person ID	INT	FK	SAMPLES

8. CE_DATA

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
CHROMOSOME_TYPE	Chromosome	VARCHAR(255)		
GENOTYPE	Genotype	VARCHAR(255)		
LOCUS	Locus	VARCHAR(255)		
QR_INDICATOR	QC indicator	VARCHAR(255)		
SAMPLE_ID	Sample ID	INT	FK	SAMPLES

9. FORENSEQ

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
CHROMOSOME_TYPE	Chromosome	VARCHAR(255)		
GENOTYPE	Genotype	VARCHAR(255)		
LOCUS	Locus	VARCHAR(255)		
QR_INDICATOR	QC indicator	VARCHAR(255)		
SAMPLE_ID	Sample ID	INT	FK	SAMPLES

10. RAZOR

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
CHROMOSOME_TYPE	Chromosome	VARCHAR(255)		
GENOTYPE	Genotype	VARCHAR(255)		
LOCUS	Locus	VARCHAR(255)		
QR_INDICATOR	QC indicator	VARCHAR(255)		
SAMPLE_ID	Sample ID	INT	FK	SAMPLES

11. FORENSEQ_SEQUENCE

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
ALLELE	Allele	VARCHAR(255)		
READ_COUNT	Number of the allele	INT		
REPEATS_PATTERN	Repeat pattern of a sequence	VARCHAR(255)		
SEQUENCE	DNA's Nucleotide sequence	VARCHAR(1000)		
REPEAT_MOTIF	Repeat motif of a sequence	VARCHAR(255)		
FORENSEQ_ID	Forenseq ID	INT	FK	FORENSEQ
SAMPLE_ID	Sample ID	INT	FK	SAMPLES

12. CHANGED_BASES

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
BASED_FROM	Nucleotide which it changes from	CHAR(1)		
POSITION	Position where the base change	INT		
BASED_TO	Nucleotide which it changes to	CHAR(1)		
FS_ID	Forenseq Sequence ID	INT	FK	FORENSEQ_SEQUENCE

13. CORE_LOCI

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
COUNTRY	Country of locus	VARCHAR(255)		
LOCUS	Core locus	VARCHAR(255)		

14. PROFILE_SEARCH_LOCUS

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
LOCUS	Locus	VARCHAR(255)		
COUNTRY_ID	Country ID	INT	FK	COUNTRIES

15. PROFILE_SEARCH_ALLELE

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
ALLELE	Allele	FLOAT		
VALUE	Allele frequency value	DECIMAL(10,10)		
LOCUS_ID	Locus ID	INT	FK	LOCI

16. SEQUENCE_ALIGNMENT_LOCI

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
LOCUS	Locus	VARCHAR(255)		
KIT_ID	Kit ID	INT	FK	KITS

17. REFERENCED_ALLELE

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
ALLELE	Allele of a locus	FLOAT		
SEQ_ALIGN_LOCUS_ID	Locus ID	VARCHAR(255)	FK	SEQUENCE_ALIGNMENT_LOCI

18. MOTIF

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
PATTERN	Motif in sequence	VARCHAR(255)		
RANGE_SEQ	Range of a motif	VARCHAR(255)		
SEQ_NO	Order number	INT		
ALLELE_ID	Allele id	INT	FK	ALLELES

19. SUMMARY_DATA

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
ALLELE	Allele	FLOAT		
BASED	Database of data	VARCHAR(255)		
COUNTRY	Country	VARCHAR(255)		
DATE	When the data were collected	DATETIME		
FREQUENCY	Frequency of an allele	FLOAT		
LOCUS	Locus	VARCHAR(255)		

20. CONFIGURATION

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
CONFIGURATION_KEY	Key of map	VARCHAR(255)		
CONFIGURATION_VALUE	Value of map	VARCHAR(255)		

21. KITS

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
CHROMOSOME_TYPE	Chromosome	VARCHAR(255)		
KIT	Kit's name	VARCHAR(255)		

22. LOCI

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
LOCUS	Locus's name	VARCHAR(255)		
KIT_ID	Kit ID	INT	FK	KITS

23. EVENT_HISTORY

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
EVENT_TYPE	Event's type	ENUM		
FEATURE	Feature of the event	VARCHAR(255)		
NOTE	Optional note	VARCHAR(255)		
EVENT_TIME	The time of the event	DATETIME		
ACTOR_ID	User ID	BIGINT	FK	USERS

24. ROLES

Column Name	Description	Data Type	Key	Reference
ID	Primary key	INT	PK	
NAME	Role's name	VARCHAR(20)		

25. USER_ROLES

Column Name	Description	Data Type	Key	Reference
USER_ID	User ID	BIGINT	FK	USERS
ROLE_ID	Role ID	INT	FK	ROLES

26. USERS

Column Name	Description	Data Type	Key	Reference
ID	Primary key	BIGINT	PK	
EMAIL	Email	VARCHAR(50)		
USERNAME	Username	VARCHAR(20)		
PASSWORD	Password	VARCHAR(12)		
STATUS	Status	INT		



ภาคผนวก ข.

ขั้นตอนการวิเคราะห์เปรียบเทียบลำดับเบสดีเอ็นเอ

1. รหัสเทียมของอัลกอริทึมที่ใช้ในการเปรียบเทียบลำดับเบสดีเอ็นเอของตัวอย่างกับลำดับเบสดีเอ็นเออ้างอิง (Pseudocode)

analyzePatternAlignment(**reference_STR_repeat_motifs**, **sequence**, **allele**)

1. **allele_counter** = 0, **result** = empty string, **single_base_changing** = null, **current_index** = 0
2. for **i** = 0 ; **i** < **reference_STR_repeat_motifs.size()** ; **i** ++
3. if **allele_counter** > **allele**,
4. break the loop.
5. **current_motif** = **reference_STR_repeat_motifs[i]**
6. **next_motif** = null
7. if **i + 1** < **reference_STR_repeat_motifs.size()**
8. **next_motif** = **reference_STR_repeat_motifs[i + 1]**
9. if **current_motif** is all upper case
10. **should_it_count** = true
11. if **current_motif** is uncounted nucleotide tracts
12. **result** = **result** + **current_motif**.
13. skip to next iteration (continue statement)
14. **number_of_repeats** = **current_motif.getNumberOfRepeats()**
15. **repeat_counter** = 0
16. while **current_index** + **current_motif.size()** <= **sequence.size()**
17. if **number_of_repeats** is not null AND **repeat_counter** >= **number_of_repeats**
18. break the while loop.
19. if **allele_counter** >= **allele**
20. break the while loop.
21. if (
 - 21.1. **next_motif** is not null AND
 - 21.2. **sequence.substring(current_index, current_index + current_motif.size())**
 - 21.3. is not equals to **current_motif** AND
 - 21.4. **sequence.indexOf(next_motif) - current_index** < **next_motif.size()**
22.)
 - 22.1. break the while loop
23. **single_base_changing** = null
24. if **should_it_count** is true
25. **single_base_changing** = **extractSingleBaseChanging(**
 - 25.1. **sequence.substring(current_index,**
 - 25.2. **current_index + current_motif.size()),**
 - 25.3. **current_motif**
26.)
27. **result_index** = **sequence.indexOf(current_motif, current_index)**
28. if **single_base_changing** is not null AND
 - 28.1. (**result_index - current_index** >= **current_motif.size()**
 - 28.2. OR **result_index** = -1)
29. break this while loop
30. if **result_index** > **current_index**
 - 30.1. AND **result_index - current_index** < **current_motif.size()**
31. **indel** = **sequence.substring(current_index, result_index)**


```

32.         result = result + indel
33.         current_index = result_index
34.         result = result + sequence.substring(current_index,
35.         current_index + current_motif.size())
36.         if should_it_count is true,
37.             allele_counter = allele_counter + 1
38.         if current_index < sequence.size()
39.             flanking = sequence.substring(current_index, sequence.size())
40.             result = result + flanking
41.         return result, single_base_changing

```

2. รหัสเทียมของฟังก์ชัน extractSingleBaseChanging ที่อยู่ในรหัสเทียมของอัลกอริทึมในการเปรียบเทียบลำดับเบสดีเอ็นเอของตัวอย่างกับลำดับเบสดีเอ็นเออ้างอิง

extractSingleBaseChanging (sequence_motif, referenced_motif)

```

1.  Set from = 'X', to = 'X', changing_count = 0
2.  if (sequence_motif.size() != referenced_motif.size() OR referenced_motif.size() < 3)
3.      return null
4.  for i = 0; i < sequence_motif.size(); i++
5.      if sequence_motif.charAt(i) != referenced_motif.charAt(i)
6.          from = referenced_motif.charAt(i)
7.          to = sequence_motif.charAt(i)
8.          changing_count = changing_count + 1
9.  if changing_count != 1
10.     return null
11. return from, to

```

บรรณานุกรม

1. Tang, H., et al., *Profiling of Short-Tandem-Repeat Disease Alleles in 12,632 Human Whole Genomes*. Am J Hum Genet, 2017. 101(5): p. 700-715.
2. Sun, J.H., et al., *Disease-Associated Short Tandem Repeats Co-localize with Chromatin Domain Boundaries*. Cell, 2018. 175(1): p. 224-238 e15.
3. Dashnow, H., et al., *STRetch: detecting and discovering pathogenic short tandem repeat expansions*. Genome Biol, 2018. 19(1): p. 121.
4. Bruno, A.E., et al., *miRdSNP: a database of disease-associated SNPs and microRNA target sites on 3'UTRs of human genes*. BMC Genomics, 2012. 13: p. 44.
5. Castellanos-Rubio, A. and S. Ghosh, *Disease-Associated SNPs in Inflammation-Related lncRNAs*. Front Immunol, 2019. 10: p. 420.
6. Alwi, Z.B., *The Use of SNPs in Pharmacogenomics Studies*. Malays J Med Sci, 2005. 12(2): p. 4-12.
7. Shastry, B.S., *SNPs in disease gene mapping, medicinal drug development and evolution*. J Hum Genet, 2007. 52(11): p. 871-880.
8. Gouy, A. and M. Zieger, *STRAF-A convenient online tool for STR data evaluation in forensic genetics*. Forensic Sci. Int. Genet., 2017. 30: p. 148-151.
9. Woerner, A.E., J.L. King, and B. Budowle, *Fast STR allele identification with STRait Razor 3.0*. Forensic Sci Int Genet, 2017. 30: p. 18-23.
10. King, J. *STRait Razor v3 Analysis (Excel-based workbook)*. [Internet] [cited 2023 6 Jan]; Available from: <https://www.dropbox.com/s/t3n0d2h6od0gek2/STRait%20Razor%20Analysis%20v3.xlsm>.
11. Burritt, B. *OmniPop 200.1*. [Internet] 23 Feb 2022; Available from: <https://strbase-archive.nist.gov/population/OmniPop200.1.xls>.
12. VEROGEN. *ForenSeq DNA Signature Prep Kit*. 2021 [cited 2021 7 Aug]; Available from: <https://verogen.com/products/forenseq-dna-signature-prep-kit/>.
13. Butler, J.M., *Short tandem repeat typing technologies used in human identity*

- testing*. Biotechniques, 2007. 43(4): p. ii-v.
14. Westen, A.A., et al., *Comparing six commercial autosomal STR kits in a large Dutch population sample*. Forensic Sci Int Genet, 2014. 10: p. 55-63.
 15. Runfola, D., et al., *geoBoundaries: A global database of political administrative boundaries*. PLoS One, 2020. 15(4): p. e0231866.
 16. *All about chromosomes*, chromosomes_under_a_microscope.jpg, Editor., Open University.
 17. Green, E.D. *Chromosome*. [cited 2021 18 July]; Available from: <https://www.genome.gov/genetics-glossary/Chromosome>.
 18. *DNA Structure*, in *Forensic human identification: An introduction*, T. Thompson and S. Black, Editors. 2006, CRC Press. p. 7.
 19. Spunk, *Comparison of a single-stranded RNA and a double-stranded DNA with their corresponding nucleobases*, D.D. RNA-EN.svg, Editor. 2010. p. Comparison of a single-stranded RNA and a double-stranded DNA with their corresponding nucleobases.
 20. Collins, F.S. *Gene*. [cited 2021 July 18]; Available from: <https://www.genome.gov/genetics-glossary/Gene>.
 21. *Gene* gene.jpg, Editor., The National Human Genome Research Institute.
 22. Biesecker, L.G.n.d. *Allele*. [cited 2021 July 18]; Available from: <https://www.genome.gov/genetics-glossary/Allele>.
 23. Bernasconi, A., *Multiple Alleles*, in *Genetics*. Encyclopedia.com.
 24. Walt, J.v.d. and J.M. Vance, *Genotype and Phenotype*, in *Genetics*. Encyclopedia.com.
 25. Maliayee, *ABO blood group phenotypes*, ABO_Blood_Group_Phenotypes.jp, Editor. 2020, Wikimedia: <https://upload.wikimedia.org>.
 26. Institute, N.H.G.R. *HAPLOID*. [Online] 2023 [cited 2023 10 June]; Available from: <https://www.genome.gov/genetics-glossary/haploid>.
 27. Institute, N.H.G.R. *DIPLOID*. [Online] 2023 [cited 2023 10 June]; Available from: <https://www.genome.gov/genetics-glossary/Diploid>.
 28. *Forensic human identification: An introduction*, T. Thompson and S. Black, Editors. 2006, CRC Press.

29. Butler, J.M., *Short tandem repeat typing technologies used in human identity testing*. BioTechniques, 2007. 43: p. Sii-Sv.
30. Sekiyu, *A partial human STR profile obtained using the applied biosystems identifier kit*, Str_profile.jpg, Editor. 2006, wikimedia.
31. Butler, J.M., *Capillary electrophoresis: Principles and instrumentation*, in *Advanced Topics in Forensic DNA Typing: Methodology*, J.M. Butler, Editor. 2012, Academic Press. p. 141-165.
32. McNamara-Schroeder, K., et al., *DNA fingerprint analysis of three short tandem repeat (STR) loci for biochemistry and forensic science laboratory courses*. Biochem Mol Biol Educ, 2006. 34(5): p. 378-83.
33. Apblum, *Diagram of capillary electrophoresis system*, Capillaryelectrophoresis.png, Editor. 2004, Wikimedia: <https://upload.wikimedia.org>
34. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors*. Proc Natl Acad Sci U S A, 1977. 74(12): p. 5463-7.
35. Behjati, S. and P.S. Tarpey, *What is next generation sequencing?* Arch Dis Child Educ Pract Ed, 2013. 98(6): p. 236-8.
36. Vejchaprasit, T. *Single nucleotide polymorphism (SNP)*. 2014 [cited 2021 25 July]; Available from: <http://biology.ipst.ac.th/?p=953>.
37. Moszyński, A., et al., *SNPs in microRNA target sites and their potential role in human disease*. Open Biology, 2017. 7: p. 170019.
38. Bailey-Wilson, J.E. *Haplotype*. [cited 2021 July 20]; Available from: <https://www.genome.gov/genetics-glossary/haplotype>.
39. Goldman, A.D. and L.F. Landweber, *What Is a Genome?* PLOS Genetics, 2016. 12(7): p. e1006181.
40. Kaur, S., et al., *Role of single nucleotide polymorphisms (SNPs) in common migraine*. The Egyptian Journal of Neurology, Psychiatry and Neurosurgery, 2019. 55(1): p. 47.
41. Bachtar, M., et al., *Towards precision medicine: interrogating the human genome to identify drug pathways associated with potentially functional*,

- population-differentiated polymorphisms*. The Pharmacogenomics Journal, 2019. 19(6): p. 516-527.
42. Shastry, B.S., *SNPs in disease gene mapping, medicinal drug development and evolution*. Journal of Human Genetics, 2007. 52(11): p. 871-880.
 43. Sobrino, B., M. Brion, and A. Carracedo, *SNPs in forensic genetics: a review on SNP typing methodologies*. Forensic Sci Int, 2005. 154(2-3): p. 181-94.
 44. Sobrino, B. and A. Carracedo, *SNP typing in forensic genetics: a review*. Methods Mol Biol, 2005. 297: p. 107-26.
 45. Butler, H., et al., *The GeoJSON Format*. Request for Comments, 2016.
 46. International, G.E., *ECMA-404: The JSON data interchange format*. 2013.
 47. Metcalf, M.B.C. *The TopoJSON Format Specification*. 2018 17 Nov 2018 [cited 2021 26 July]; Available from: <https://github.com/topojson/topojson-specification>.
 48. Yug, *Topojson shapes*, T. shapes-en.svg, Editor. 2013, Wikimedia Commons: <https://commons.wikimedia.org/>. p. Core elements of Topojson shapes drawing system : point, line, polygon.
 49. Noback, M. and SpringerLink, *Principles of Package Design : Creating Reusable Software Components*. 1st 2018. ed. 2018, Berkeley, CA: Apress : Imprint: Apress.
 50. Cmendes, *A simple example of several software components - pictured within a hypothetical holiday-reservation system represented in UML 2.0.*, Component-based-Software-Engineering-example2.png, Editor. 2009, Wikimedia. p. A simple example of two software components.
 51. Gamma, E., et al., *Design patterns: elements of reusable object-oriented software*. 1995: Addison-Wesley Longman Publishing Co., Inc.
 52. RegisFrey, *Diagram of interactions within one possible take on the MVC pattern*. 2010, Wikipemedia: wikipedia.org. p. The model, view, and controller (MVC) pattern relative to the user.
 53. RegisFrey, *The model, view, and controller (MVC) pattern relative to the user*, MVC-Process.svg, Editor. 2010, Wikimedia Commons: <https://commons.wikimedia.org/>.
 54. Bodner, M., et al., *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 Genet, 2016. 24: p. 97-102.
55. al, B.e., *STRidER in the field of forensic STR typing*, STRidER_Abb_no_font.svg, Editor. 2016, STRidER.
56. Butler, J.M., et al., *STANDARD REFERENCE DATA*. 2021, National Institute of Standards and Technology: National Institute of Standards and Technology's website. p. Free interface SRD database.
57. Willuweit, S. and L. Roewer, *YHRD : Y-Chromosome STR Haplotype Reference Database*. 2021, Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
58. Cheung, K.H., et al., *ALFRED: a Web-accessible allele frequency database*. Pac Symp Biocomput, 2000: p. 639-50.
59. Heras, J., et al., *A survey of tools for analysing DNA fingerprints*. Brief Bioinform, 2016. 17(6): p. 903-911.
60. Phillips, C., et al., *"The devil's in the detail": Release of an expanded, enhanced and dynamically revised forensic STR Sequence Guide*. Forensic Sci Int Genet, 2018. 34: p. 162-169.
61. STRAF. *Chapter 3 Forensic parameters | The STRAF Book*. [Online] 2021 [cited 2023 10 June]; Available from: <https://agouy.github.io/straf/forensic-parameters.html#power-of-discrimination-pd>.
62. Botstein, D., et al., *Construction of a genetic linkage map in man using restriction fragment length polymorphisms*. Am J Hum Genet, 1980. 32(3): p. 314-331.
63. Liu, B.H., *Statistical Genomics: Linkage, Mapping, and QTL Analysis*. 1998: CRC Press.
64. labs, z.c. *React Simple Maps*. [Internet] 2021 [cited 2022 21 Feb]; Available from: <https://www.react-simple-maps.io/>.
65. Earth, N. *Natural Earth | Downloads - Free vector and raster map data at 1:10m, 1:50m, and 1:110m scales*. [Internet] 2022 [cited 2022 21 Feb]; Available from: <https://www.naturalearthdata.com/downloads/>.

66. Zimmerman, R. *How to convert and prepare TopoJSON files for interactive mapping with d3*. [Internet] 2020 [cited 2022 21 Feb]; Available from: <https://medium.com/hackernoon/how-to-convert-and-prepare-topojson-files-for-interactive-mapping-with-d3-499cf0ced5f>.
67. STRAF. *The STRAF Book*. 2022 6 Jan 2023; Available from: https://agouy.github.io/straf_book/forensic-parameters.html.
68. Liu, B.H., *Statistical Genomics: Linkage, Mapping, and QTL Analysis (1st ed.)*. 1998: CRC Press.
69. Facebook. *React A JavaScript library for building user interfaces*. [Internet] 2021 [cited 2022 21 Feb]; Available from: <https://react.dev/>.
70. Spring. *Spring Boot*. [Internet] 2021 [cited 2022 21 Feb]; Available from: <https://spring.io/projects/spring-boot>.
71. Oracle. *MySQL :: MySQL 8.0 Reference Manual*. [Internet] 2021 [cited 2022 21 Feb]; Available from: <https://dev.mysql.com/doc/refman/8.0/en/>.
72. MySQL. *MySQL Customer: Facebook*. [Online] [cited 2023 9 June]; Available from: <https://www.mysql.com/customers/view/?id=757>.
73. Wordpress. *A Developer's Guide: Ghost vs WordPress*. [Online] [cited 2023 9 June]; Available from: <https://developer.wordpress.com/2022/10/24/a-developers-guide-ghost-vs-wordpress/>.



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ประวัติผู้เขียน

ชื่อ-สกุล	Nuttachai Kulthammanit
วัน เดือน ปี เกิด	24 Nov 1996
สถานที่เกิด	Bangkok
ผลงานตีพิมพ์	Nuttachai Kulthammanit, Tikumphorn Sathirapatya, Poonyapat Sukawutthiya, Hasnee Noh, Kornkiat Vongpaisarnsin, and Duangdao Wichadakul. (2023). STRategy: A support system for collecting and analyzing next-generation sequencing data of short tandem repeats for forensic science. PLOS One Journal on San Francisco, California, and Cambridge, United Kingdom.



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY