16/0206/45

โครงการหาลำดับนิวคลีโอไทด์ ในจีโนมของเชื้อมาลาเรีย *Plasmodium falciparum*

The *Plasmodium falciparum* Genome Project

*Tepanata Pumpaiboon**

บทคัดย่อ: เทพนาฏ พุ่มไพบูลย์. 2545. โครงการหาลำดับนิวคลีโอไทด์ในจีโนมของเชื้อมาลาเรีย *Plasmodium falciparum*. วารสารวิจัยวิทยาศาสตร์การแพทย์ 16(2) : 159-171.

บทความนี้ได้รวบรวมความเป็นมาของโครงการหาลำดับนิวคลีโอไทด์ในจีโนมของเชื้อมาลาเรีย *Plasmodium falciparum* ซึ่งถูกตั้งขึ้นในชื่อของ The Wellcome Trust Malaria Genome Mapping Project ในปี พ.ศ. 2536 ด้วยการสนับสนุนจาก Wellcome Trust (อังกฤษ) การศึกษาโครโมโซมเริ่มจากการแยกโครโมโซมของเชื้อมาลาเรียด้วยเทคนิค pulsed-field gel electrophoresis และนำไปโคลนเข้าสู่พลาสมิด เลือกพลาสมิดที่มีชิ้นโครโมโซมแทรกอยู่โดยใช้ sequence tagged site (STS) แล้วนำไปสร้างแผนที่โครโมโซม ด้วยการสนับสนุนด้านการเงินจากแหล่งเงินทุนหลายแหล่ง ทำให้โครงการในปัจจุบันมีขนาดใหญ่ขึ้นครอบคลุมสถาบันวิจัยสี่แห่ง ได้แก่ Sanger Centre (อังกฤษ), Institute for Genomic Research, Naval Medical Research Institute และ Stanford University (สหรัฐอเมริกา) เพื่อร่วมมือกันในการหาลำดับเบสของโครโมโซมทั้ง 14 แท่งของเชื้อมาลาเรีย สถาบันทั้งสี่แห่งได้รายงานแล้วว่า การหาลำดับเบสของโครโมโซมแท่งที่ 2 และ 3 เสร็จสมบูรณ์แล้ว ส่วนโครโมโซมแท่งที่ 1, 4, 9, 10, 11, 12 และ 14 นั้นอยู่ในขั้นตอนสุดท้ายของการหาลำดับเบส ส่วนโครโมโซมที่เหลือ ได้แก่ แท่งที่ 5, 6, 7, 8 และ 13 อยู่ในระหว่างการหาลำดับเบส นักวิทยาศาสตร์ต่างหวังว่าลำดับเบสที่ได้จากการศึกษาตามโครงการนี้จะสามารถช่วยให้การศึกษาเชื้อมาลาเรียได้ถี่ถ้วนขึ้นในยุคหลังการศึกษาจีโนม และเพื่อให้ความรู้เหล่านี้ถูกรวบรวมไว้ในแหล่งเดียวกันจึงมีการสร้างฐานข้อมูลชื่อ PlasmoDB สำหรับเก็บข้อมูลดังกล่าว

คำสำคัญ : มาลาเรีย พลาสโมเดียม ฟัลซิปารัม จีโนม แผนที่โครโมโซม การหาลำดับเบส ยุคหลังการศึกษาจีโนม

*Malaria Research Center, Institute of Health Research, Chulalongkorn University, Bangkok, Thailand
ศูนย์วิจัยมาลาเรีย สถาบันวิจัยวิทยาศาสตร์การแพทย์ จุฬาลงกรณ์มหาวิทยาลัย กรุงเทพมหานคร

Abstract : Tepanata Pumpaiboon. 2002. The *Plasmodium falciparum* genome project. Thai J Health Res 16(2) : 159-171.

The history of the *Plasmodium falciparum* Genome Project is described. The project was started as The Wellcome Trust Malaria Genome Mapping Project in 1993, supported by the Wellcome Trust (England). Using the pulsed-field gel electrophoresis technique, malaria chromosomes were separated and cloned into vector plasmids. Recombinant plasmids containing inserted chromosome fragments were screened with a sequence tagged site (STS) and were constructed as physical maps. At present, with the support of funding agencies, the project has become larger. Four institutes, the Sanger Centre (England), The Institute for Genomic Research, the Naval Medical Research Institute and Stanford University (USA), are responsible for the sequencing of all 14 falciparum chromosomes. All chromosome 2 and 3 sequences have been completed and reported on. Chromosomes 1, 4, 9, 10, 11, 12 and 14 are in the last phase of the sequencing step, while chromosomes 5, 6, 7, 8 and 13 sequencing are progressing. With these genomic sequences, scientists should be able to study the parasite more thoroughly in the Post-genomic era. The database, PlasmoDB, was created to gather together the knowledge from the genome project and those from other post-genomic era studies.

Key words : *Plasmodium falciparum*, genome, physical map, sequencing, post-genomic era

---

บทนำ

การควบคุมโรคมาลาเรียประสบกับปัญหาหลายประการ และอุปสรรคที่สำคัญ คือ เชื้อมาลาเรียดื้อต่อยาที่ใช้ในการรักษาโดยเฉพาะเชื้อพลาสโมเดียม ฟัลซิปารัม (*Plasmodium falciparum*) ซึ่งเป็นเชื้อที่ทำให้เกิดโรคมาลาเรียขึ้นสมองในคน รวมทั้งโครงการควบคุมยุงพาหะประสบความล้มเหลว ทำให้โรคมาลาเรียกลับมาเป็นปัญหาที่สำคัญทางสาธารณสุข จากการสำรวจขององค์การอนามัยโลกพบว่ามีผู้ติดเชื้อประมาณ 300-500 ล้านรายทั่วโลก และมีผู้เสียชีวิตประมาณ 1.5-2.7 ล้านคนต่อปี โดยส่วนใหญ่เป็นผู้ป่วยเด็กที่มีอายุต่ำกว่า 5 ปีในแอฟริกา (World Health Organization,1997) การศึกษาหาแนวทางใหม่ๆในการป้องกันรักษาโรคมาลาเรีย เช่น การพัฒนายาและวัคซีนชนิดใหม่เพื่อป้องกันโรคมาลาเรียจึงมีความจำเป็นอย่างยิ่ง แต่ทั้งนี้ต้องอาศัยความเข้าใจในชีววิทยาของเชื้อซึ่งมีวงชีวิตที่สลับซับซ้อน กล่าวคือ เชื้อมาลาเรียต้องการโฮสต์ 2 ชนิด คือ สัตว์มีกระดูกสันหลังและสัตว์ไม่มีกระดูกสันหลัง (ยุง) และยังมีขั้นตอนการเจริญออกเป็นหลายระยะ แต่ละระยะมีความแตกต่างกันทั้งด้านชีววิทยา ชีวเคมี และสรีร-วิทยา (สดศรี ไทยทอง, 2540) ดังนั้นการศึกษาลำดับนิวคลีโอไทด์ในจีโนมของเชื้อมาลาเรียจึงเป็นข้อมูลสำคัญที่จะทำให้เกิดความรู้ความเข้าใจในกระบวนการต่าง ๆที่เกิดขึ้นตลอดวงชีวิตของเชื้อ รวมไปถึงความสัมพันธ์ระหว่างเชื้อกับโฮสต์ (Host) เพื่อนำไปใช้เป็นแนวทางในการป้องกันรักษาโรคได้อย่างมีประสิทธิ-ภาพต่อไป

จีโนมของเชื้อ *P. falciparum* มีขนาดประมาณ 30 ล้านคู่เบส ประกอบด้วยโครโมโซมทั้งหมด 14 แท่ง และดีเอ็นเอที่อยู่นอกโครโมโซม (extrachromosomal DNA) คือ ดีเอ็นเอเกลียวคู่เส้นยาว (linear DNA) ซึ่งอยู่ภายในไมโตคอนเดรีย (mitochondria) (Feagin, 1992) และดีเอ็นเอเกลียวคู่รูปวงแหวน (circular DNA) ในออร์แกนเนลที่เรียกว่า apicoplast (Howe, 1992) สำหรับเชื้อ *P. falciparum* สายพันธุ์บริสุทธิ์ที่ใช้ในการศึกษาหาลำดับนิวคลีโอไทด์ คือ สายพันธุ์บริสุทธิ์ 3D7 เนื่องจากเป็นสายพันธุ์ที่สามารถพัฒนาให้ระยะต่าง ๆได้ครบทั้งวงชีวิต และยังมีผลการศึกษาทางพันธุกรรมมาก่อนหน้านี้แล้วจากนักวิจัยหลายกลุ่ม นอกจากนั้นยังสามารถเพิ่มจำนวนเชื้อได้โดยเพาะเลี้ยงในห้องปฏิบัติการด้วยวิธี *in vitro* cultivation

โครงการสร้างแผนที่โครโมโซมของเชื้อมาลาเรีย

ในปี พ.ศ. 2536 นักวิจัยจากสถาบันต่าง ๆ (ตารางที่ 1) ร่วมมือกันก่อตั้งโครงการ The Wellcome Trust Malaria Genome Mapping Project โดยมีเป้าหมายที่จะสร้างแผนที่โครโมโซม (physical map) ในแต่ละแท่งของเชื้อ *P. falciparum* โดยได้รับทุนสนับสนุนจาก The Wellcome Trust ประเทศอังกฤษ (Dame *et al.*,1996) การทำแผนที่โครโมโซมเริ่มจากแยกโครโมโซมของ *P. falciparum* ออกจากกันด้วยเทคนิค pulse-field gel electrophoresis (PFGE) จากนั้นนำเอาโครโมโซมแต่ละแท่งที่แยกได้มาตัดด้วยเอนไซม์ตัดจำเพาะชนิดต่าง ๆ (restriction enzymes) แล้วจึงโคลน (clone) แต่ละส่วนเข้าสู่ยีสต์โดยใช้ pYAC4 vector ซึ่งการโคลนชิ้นส่วนของดีเอ็นเอของ *P. falciparum* ซึ่งมีองค์ประกอบนิวคลีโอไทด์ A+T สูงถึงร้อยละ 80 เข้ากับ YAC (Yeast Artificial Chromosome) นั้นมีความเสถียรกว่าการโคลนชิ้นส่วนดีเอ็นเอเข้ากับพลาสมิดของแบคทีเรีย โดยเฉพาะในกรณีที่ชิ้นดีเอ็นเอมีขนาดใหญ่ ซึ่งดีเอ็นเอมักจะเกิด

การขาดหาย (deletion) และเกิดการจัดเรียงลำดับใหม่ (rearrangement) YAC clone นี้จะถูกคัดเลือกไปใช้ในการทำแผนที่โครโมโซมต่อไป

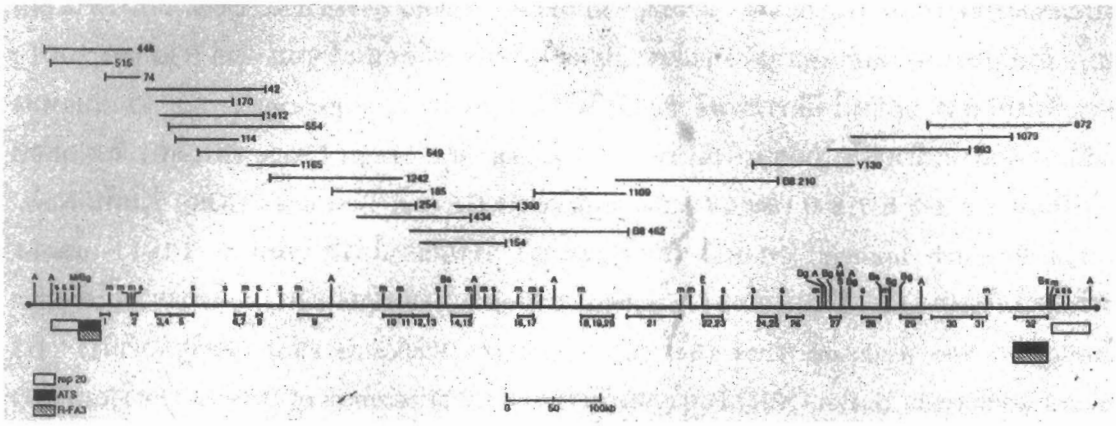
ตารางที่ 1 สถาบันที่เข้าร่วมในโครงการทำแผนที่โครโมโซม (Dame *et al.*, 1996)

Institution	Major areas of contribution
Sloan Kettering Institute (SKI)	RM, YAC, C1, C2, C7, C8
Walter and Eliza Hall Institute of Medical Research (WEHI)	YAC, C3, C4, C5, C12
Oxford University (OX)	RC, YAC, C6, C14
Menzies School of Health Research (MSHR)	C9, C13
National Institute for Medical Research (NIMR)	C10, C11
University of Wurzburg (UW)	C8
National Institutes of Health (NIH)	RM, MS, C7
University of Florida (UF)	GST
University of Edinburgh (UE)	YAC, CA
Monash University (MU)	I
Whitehead Institute (WI)	I

หมายเหตุ C1-C14, Contig Assembly of YACS for Chromosome #; CA, Chromosome Assignment; GST, Gene Sequence Tags; I, Information; MS, Microsatellite loci; RC, Resource Center; RM, Restriction Maps; YAC, YAC Libraries

คัดเลือก YAC clone โดยใช้ลำดับเบสระบุตำแหน่ง (sequence tagged site, STS) ซึ่งอาจได้มาจากลำดับเบสของยีนในฐานข้อมูลซึ่งมีผู้ศึกษาแล้ว หรือได้จาก λgt10 library (Thompson and Cowman,1997) หรือหาลำดับเบสของชิ้นดีเอ็นเอของ *P. falciparum* ที่เชื่อมอยู่กับ YAC โดยใช้วิธี inverted polymerase chain reaction (IPCR) (Vaudin *et al.*,1995) STS เหล่านี้จะถูกนำไปจับคู่ (hybridize) กับโครโมโซมที่แยกด้วย PFGE เสียก่อนเพื่อยืนยันว่า STS นี้อยู่บนโครโมโซมแท่งนั้นจริง นอกจากนั้นยังต้องเปรียบเทียบลำดับเบสของ STS กับข้อมูลในฐานข้อมูล GenBank และฐานข้อมูล EMBL (European Molecular Biology Laboratory) โดยใช้โปรแกรม BLAST (Basic Local Alignment Search Tool) ซึ่งผลที่ได้นี้จะทำให้ทราบว่าลำดับเบสของ STS นั้นตรงกับยีนใดในฐานข้อมูล จากนั้นจึงออกแบบ primer จากลำดับเบสของ STS แล้วใช้เทคนิค PCR คัดเลือก YAC clone ที่ต้องการ YAC clone แต่ละโคลนจะมีดีเอ็นเอแต่ละส่วนของโครโมโซม ซึ่งจะมีลำดับเบสเหลื่อมซ้อนกัน เรียกแต่ละโคลนว่า YAC contig ซึ่งเมื่อนำมาเรียงลำดับต่อกันโดยอาศัยตำแหน่งของเอนไซม์ตัดจำเพาะที่อยู่บนชิ้นส่วนของดีเอนเอนั้น (restriction map) ก็จะได้เป็นโครโมโซมทั้งหมด (contig หมายถึง ชิ้นดีเอ็นเอที่ถูกตัดด้วยเอนไซม์ตัดจำเพาะซึ่งจะมีขนาดแตกต่างกันไปขึ้นอยู่กับตำแหน่งของเอนไซม์ตัดจำเพาะนั้น) อย่างไรก็ตามในบางครั้งอาจเกิดส่วนที่เป็นช่องว่าง (gap) ซึ่งไม่สามารถหาโคลนใดใน library ของ *P. falciparum* clone 3D7 ที่มีลำดับเบสตรงกับ STS ในตำแหน่งนั้นๆ ในกรณีนี้จึงต้องเลือก YAC library จาก *P. falciparum* clone อื่น ซึ่งจะต้องพิจารณาเปรียบเทียบตำแหน่งที่ตัดด้วยเอนไซม์ตัดจำเพาะ (restriction site) ด้วย (Thompson and Cowman,1997) เมื่อจัดเรียงลำดับชิ้นส่วนดีเอ็นเอใน YAC clone เรียบร้อย

แล้วก็จะได้แผนที่โครโมโซมซึ่งมีรายละเอียดทั้งตำแหน่งที่ตัดด้วยเอนไซม์ตัดจำเพาะ ตำแหน่งของ STS บนโครโมโซม และมี YAC contig กำกับอยู่ ดังแสดงในรูปที่ 1



รูปที่ 1 แผนที่โครโมโซมแท่งที่ 3 ของ *Plasmodium falciparum* (physical map) แสดงให้เห็น YAC contig, STS และตำแหน่งเอนไซม์ตัดจำเพาะบนโครโมโซม (high resolution restriction map) ของเอนไซม์ *Apa* I (A), *Bgl* I (Bg), *Bss* HII (Bs), *Eag* I (E), *Sgr* AI (Sg), Msc I (m) และ Stu I (s) (Thompson and Cowman, 1997)

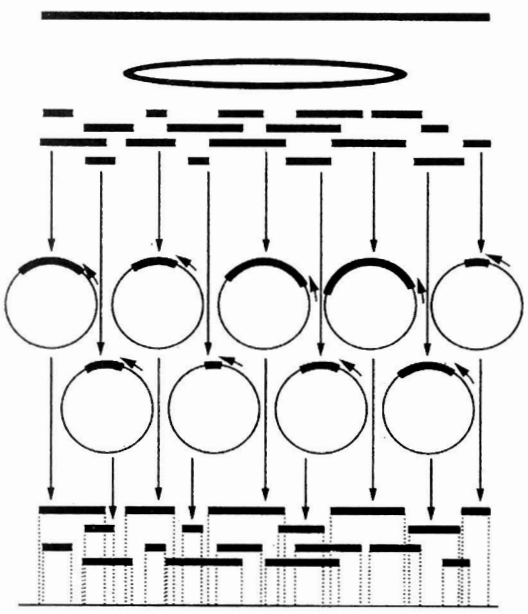
นอกจากใช้ STS ที่กล่าวข้างต้นมาศึกษาแล้วยังมี STS อื่นๆอีก คือ gene sequence tag (GST) ที่ได้จากการใช้เอนไซม์นิวคลีเอสจากถั่วเขียว (mung bean nuclease) ซึ่งมีคุณสมบัติในการตัด genomic DNA ที่ตำแหน่งก่อนและหลัง structural gene รวมทั้งภายในส่วน intron บนสายดีเอ็นเอ ชิ้นดีเอ็นเอที่ได้จะถูกโคลนเข้ากับพลาสมิด แล้วนำเข้าสู่เซลล์ *Escherichia coli* ต่อจากนั้นจะหาลำดับเบสของชิ้นดีเอ็นเอโดยการเลือกโคลนแบบสุ่มและเลือกชิ้นดีเอ็นเอที่มีขนาดเล็ก (Reddy *et al.*,1993)

expressed sequence tag (EST) ก็เป็น STS อีกกลุ่มหนึ่งที่ได้จากการใช้เทคนิค reverse transcriptase polymerase chain reaction (RT-PCR) คือการเปลี่ยน mRNA ของเชื้อในระยะเม็ดเลือดแดง (asexual blood stage) เป็น cDNA แล้วนำไปโคลนเพื่อหาลำดับเบสต่อไป (Dame,1996) STS เหล่านี้มีส่วนช่วยทำให้การหาลำดับเบสบนแผนที่โครโมโซมง่ายขึ้น

โครงการหาลำดับนิวคลีโอไทด์ในจีโนมของเชื้อมาลาเรีย

ในปี พ.ศ. 2538 ได้มีการริเริ่มโครงการหาลำดับเบสในจีโนมของเชื้อแบคทีเรียที่ทำให้เกิดโรคไข้หวัดใหญ่ (*Haemophilus influenzae*) สิ่งมีชีวิตชนิดจึงนี้เป็นสิ่งมีชีวิตชนิดแรกที่เราทราบลำดับนิวคลีโอไทด์ทั้งหมดในจีโนม (Fleischmann *et al.*,1995) การหาลำดับเบสนี้ใช้เทคนิค shotgun sequencing เริ่มด้วยการตัด genomic DNA ให้เป็นชิ้นเล็กๆ ขนาด 1-2 พันคู่เบส ด้วยเอนไซม์ตัดจำเพาะ หรืออาศัยแรงกลอื่นๆ เช่น คลื่นเสียงความถี่สูง เมื่อได้ชิ้นดีเอ็นเอแล้วจึงโคลนเพื่อหาลำดับนิวคลีโอไทด์ต่อไป จากงานวิจัยชิ้นนี้ทำให้นักวิทยาศาสตร์เกิดความสนใจที่จะศึกษาลำดับนิวคลีโอไทด์ในจีโนมของแบคทีเรียก่อโรคชนิดอื่นๆ เพื่อให้เกิดความเข้าใจทั้งด้านชีววิทยา ชีวเคมี และกลไกในการก่อโรค หลังจากนั้นไม่นานกลุ่มนักวิจัยที่ทำการศึกษาเชื้อ *P. falciparum* เริ่มศึกษาความเป็นไปได้ในการหาลำดับนิวคลีโอไทด์ทั้งจีโนมของเชื้อนี้ ซึ่งในขณะนั้นเป็นงานที่ค่อนข้างเป็นไปได้ยากเนื่องจากจีโนมของ *P. falciparum*

มีขนาดใหญ่กว่าจีโนมของ *H. influenzae* ถึง 15 เท่า นั่นคือจะต้องมี shotgun sequence ประมาณ 500,000 sequence การเก็บรวบรวมลำดับ นิวคลีโอไทด์จำนวนมากนี้จำเป็นต้องมีการพัฒนาซอฟท์แวร์ที่มีประสิทธิภาพกว่าที่มีอยู่ในขณะนั้น และปัญหาอีกประการหนึ่งคือ การใช้ YAC clone ในการหาลำดับเบสนั้น ถ้าเป็นโคลนที่มีดีเอ็นเอสายสั้น ๆแล้วจะมีนิวคลีโอไทด์ของยีสต์ปนเปื้อนมาด้วย (Gardner,2001) จนกระทั่งในปี พ.ศ. 2539 ได้มีการรวมตัวของนักวิจัย 3 กลุ่มเพื่อเป็น sequencing center ร่วมกับแหล่งทุนเพื่อการหาลำดับนิวคลีโอไทด์ของทั้งจีโนม *P. falciparum* โดย Sanger Centre (อังกฤษ) รับผิดชอบโครโมโซมที่ 1 3 4 5 6 7 8 9 และ 13 The Institute for Genomic Research (TIGR) ร่วมกับ Naval Medical Research Institute (NMRI) (สหรัฐอเมริกา) รับผิดชอบโครโมโซมที่ 2 10 11 และ14 สำหรับโครโมโซมที่ 12 เป็นหน้าที่ของ Stanford University (สหรัฐอเมริกา) โครงการนี้ได้รับทุนสนับสนุนจาก The Wellcome Trust (อังกฤษ) Burroughs Wellcome Fund (สหรัฐอเมริกา) US National Institutes of Health (NIH) (สหรัฐอเมริกา) และ US Department of Defense (สหรัฐอเมริกา) (Hoffman *et al*.,1997) งานเริ่มต้นด้วยการหาแนวทางในการหาลำดับเบสและพัฒนาทรัพยากรต่าง ๆที่มีอยู่เพื่อช่วยให้ได้ลำดับเบสที่สมบูรณ์ ในที่สุดได้เลือกวิธี shotgun sequencing เริ่มจากใช้เทคนิค PFGE แยกโครโมโซมแต่ละแท่งออกจากกัน ต่อจากนั้นสกัดโครโมโซมโดยย่อยเจลด้วยเอนไซม์ agarase chromosamal DNA ที่ได้จะถูกตัดเป็นชิ้นขนาด 1-2 พันคู่เบส แล้วโคลนเข้ากับพลาสมิดหรือ M13 vector ซึ่งเป็น double stranded vector สุ่มเลือกโคลนเพื่อหาลำดับเบสของชิ้นดีเอ็นเอทั้งสองทิศทางคือจากซ้ายไปขวาและจากขวาไปซ้าย เพื่อให้ได้ข้อมูลลำดับเบสของดีเอ็นเอทั้งสองสาย ซึ่งจะสามารถนำไปใช้ประโยชน์ในการเติมเบสที่หายไปหรือเติมช่องว่าง (gap) ได้อีกด้วย ลำดับเบสของแต่ละชิ้นดีเอ็นเอจะถูกรวบรวมโดยใช้โปรแกรม phrap หรือ TIGR Assembler (Gardner,2001)

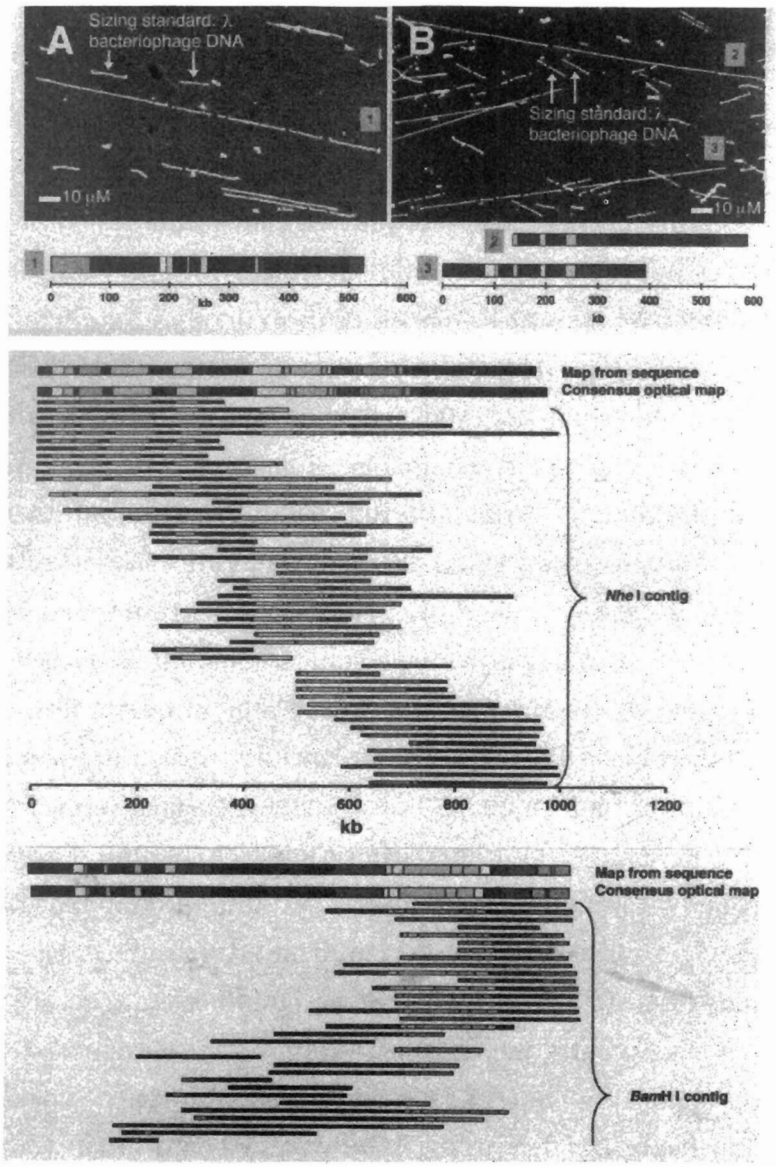


รูปที่ 2 แบบแผนในการทำ shotgun sequencing

นอกจากนั้นกลุ่มของ Sanger Centre และ Stanford University ยังได้หาลำดับเบสใน shotgun library ที่ได้จาก YAC clone ซึ่งทราบตำแหน่งบนโครโมโซมแล้ว ลำดับเบสที่ได้จาก YAC clone นี้จะช่วยในการจัดแบ่งลำดับเบสของแต่ละโคลน (จาก chromosomal DNA) ใน library ออกเป็นกลุ่มย่อย จากนั้นลำดับเบสของแต่ละโคลนที่จัดเก็บไว้ก็จะถูกเรียงเข้าด้วยกันโดยพิจารณาจากลำดับเบสทั้งสองทิศทาง ที่ตรงกันก็จะนำมาต่อเข้าด้วยกันจนได้ลำดับเบสของสายดีเอ็นเอที่ยาวขึ้นเรื่อย ๆ ลำดับเบสที่ได้นี้จะถูกนำไปเปรียบเทียบกับลำดับเบสของ STS หรือ microsatellite marker เพื่อให้ทราบว่าชิ้นดีเอ็นเอนี้อยู่บริเวณใดในแผนที่โครโมโซม นอกจากนั้นยังต้องเทียบกับ optical restriction map ของโครโมโซมนั้น ๆอีกด้วย เพื่อเป็นการยืนยันว่าข้อมูลของลำดับเบสมีความถูกต้อง (รูปที่ 3) สำหรับส่วนของโครโมโซมที่ยังไม่ทราบลำดับเบสที่เหลืออยู่นั้นใช้เทคนิค primer walking (ใช้ universal primer ของ vector เป็น primer ในการทำ DNA sequencing จะสามารถอ่านลำดับเบสได้คราวละ 500 เบส จากนั้นจึงใช้ลำดับเบสที่ได้มาออกแบบ primer ต่อไปจนกระทั่งหาลำดับเบสครบทั้งชิ้น DNA) หรือใช้ลำดับเบสจากบริเวณข้างเคียงในการออกแบบ primer จากนั้นทำ PCR และหาลำดับเบสตามลำดับ (Gardner *et al.*,1998; Bowman *et al.*,1999) ส่วนบริเวณที่มีเบส A และ T เป็นองค์ประกอบสูง (เช่น บริเวณที่เป็น centromere บนโครโมโซมที่ 2 และ 3) ซึ่งมีสูงถึงร้อยละ 97 หรือบริเวณที่มีเบส A และ T ต่อกันเป็นสายยาวซึ่งจะทำให้การหาลำดับเบสให้ถูกต้องแม่นยำเป็นไปได้ยาก ในกรณีเหล่านี้จะแก้ปัญหาโดยการแทรก transposon เข้าไป หรือตัดสายดีเอ็นเอให้มีขนาดเล็กประมาณ 50-500 เบส แล้วโคลนเข้า m13mp18 vector ซึ่งเรียกวิธีนี้ว่า microlibrary technique (Bowman *et al.*,1999) เพื่อทำลายโครงสร้างทุติยภูมิ (secondary structure) ของดีเอ็นเอที่จะรบกวนการหาลำดับเบส วิธีนี้จะช่วยให้การหาลำดับเบสมีความถูกต้องแม่นยำมากขึ้น แต่กระบวนการเหล่านี้ต้องใช้เวลาจึงเป็นสาเหตุให้การหาลำดับนิวคลีโอไทด์ของเชื้อ *P. falciparum* มีความล่าช้า เมื่อการหาลำดับเบสทั้งโครโมโซมเสร็จสมบูรณ์ก็จะเปรียบเทียบกับ optical restriction map ว่าลำดับเบสของโครโมโซมนี้มีความถูกต้องหรือไม่ และข้อมูลของลำดับเบสที่ได้จะเผยแพร่ทาง website ของ NCBI (The National Coalition Building Institute) เป็นระยะ ๆ ตั้งแต่เริ่มต้นโครงการ และเมื่อไม่นานมานี้ได้มีการสร้างฐานข้อมูล PlasmoDB สำหรับรวบรวมข้อมูลจีโนมของเชื้อมาลาเรีย (malaria genome) ทั้งหมดลงในฐานข้อมูลนี้

การหาลำดับเบสของโครโมโซมแท่งที่ 2 และแท่งที่ 3 ได้เสร็จสมบูรณ์และมีการตีพิมพ์เผยแพร่ไปแล้ว สำหรับโครโมโซมแท่งอื่น มีรายงานความก้าวหน้าในเดือนมิถุนายน พ.ศ. 2544 ว่าการหาลำดับเบสของโครโมโซมแท่งที่ 1 4 9 10 11 12 และ 14 กำลังดำเนินการอยู่ในช่วงสุดท้ายของการเติมลำดับเบสในส่วนที่เป็น gap ซึ่งก็เหลืออยู่ไม่มากนัก สำหรับโครโมโซมแท่งที่ 5 6 7 และ 8 ซึ่งเป็นกลุ่มของโครโมโซมที่แยกจากกันไม่ชัดเจนบนเจล (PFGE) และโครโมโซมแท่งที่ 13 การเติม gap ทำได้ช้าเนื่องจากยังมี marker ไม่เพียงพอ (Gardner,2001)

เมื่อขั้นตอนการหาลำดับเบสของโครโมโซมแต่ละแท่งเสร็จสิ้นลง ต่อไปจะเป็นขั้นตอนของการทำ annotation ซึ่งหมายถึงการวิเคราะห์หาลำดับเบสในบริเวณต่าง ๆบนโครโมโซมที่อาจมีความสัมพันธ์กับกระบวนการ (function) ต่าง ๆที่สำคัญ ของเชื้อ *P. falciparum* รวมถึงบริเวณที่เกิดการแปลรหัสเป็นโปรตีน (protein-coding region) บริเวณที่ถอดรหัสเป็น tRNA และบริเวณที่มีลำดับเบสซ้ำ ๆกัน (repetitive sequence)



รูปที่ 3 Optical map ของโครโมโซมแท่งที่ 2 ของ *Plasmodium falciparum* ซึ่งได้จากการตัดแท่งโครโมโซมที่แยกด้วย PFGE แล้วตัดด้วยเอนไซม์ตัดจำเพาะ ปฏิกิริยานี้เกิดขึ้นบนแผ่นสไลด์ จากนั้นหาตำแหน่งที่ตัดและขนาดชิ้นดีเอ็นเอที่ได้ด้วย fluorescent microscopy ความเข้มข้นของแสงฟลูโอเรสเซนต์จะบอกถึงขนาดของชิ้นดีเอ็นเอ

บน : โมเลกุลของดีเอ็นที่ได้จากการตัดด้วยเอนไซม์ตัดจำเพาะ *Nhel* B (A) และ *Bam* HI (B)

ล่าง : แผนที่โครโมโซม (Jing *et al.*, 1999)

ในการ annotate นั้นจะมีโปรแกรมคอมพิวเตอร์ช่วยค้นหาตำแหน่งของยีนบนโครโมโซม เช่น Hexamer/Genefinder (Horrocks *et al.*,2000) GlimmerM และ phat ซึ่งสองโปรแกรมหลังนี้นักวิจัยใช้ในการ annotate ยีนบนโครโมโซมแท่งที่ 2 และแท่งที่ 3 ให้สำเร็จลุล่วงไปด้วยดี แต่ในบางกรณีแต่ละโปรแกรมที่นำมาใช้ในการวิเคราะห์แสดงรูปแบบของยีนที่แตกต่างกันและในบริเวณนั้นมีลำดับเบสที่ไม่ตรงกับลำดับเบสของ EST (expressed sequence tag)ใดเลยในฐานข้อมูลทั้งหมดที่มีอยู่ ในกรณีนี้ ผู้ที่ทำหน้าที่วิเคราะห์จะต้องตัดสินใจเลือกรูปแบบของยีนที่น่าจะเป็นรูปแบบที่ถูกต้องที่สุด ด้วยเหตุนี้จึงมี

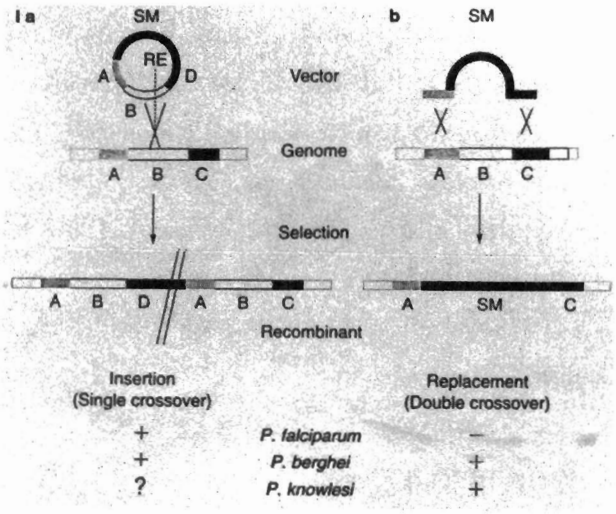
ความพยายามที่จะแก้ไขโดยเพิ่มรูปแบบของยีนที่ได้จากผลการวิจัยที่ลงตีพิมพ์ย้อนหลัง 3 ปีเข้าไปใน Training set ของโปรแกรม GlimmerM นอกจากนั้นยังมีการเพิ่มกลุ่มข้อมูล EST จากสิ่งมีชีวิตอื่น ๆเข้าไปในฐานข้อมูลเพื่อช่วยในการทำนายรูปแบบของยีนหรือส่วนที่เป็น intron อีกด้วย (Gardner *et al.*,2001)

หลังจากที่ได้ทำการวิเคราะห์ยีนต่าง ๆบนโครโมโซมแล้ว ยีนเหล่านี้จะถูกจัดออกเป็นประเภทต่าง ๆตามระบบ Gene Ontology (GO) system ซึ่งระบบนี้ได้นำไปใช้กับฐานข้อมูลของยีสต์และแมลงหวี่แล้ว โดย GO system จะประกอบด้วย 3 ontology คือ 1. molecular function ontology 2. biological process ontology 3. cellular component ontology ในแต่ละ ontology จะมีกลุ่มของคำที่ใช้ในการอธิบายผลิตภัณฑ์ของยีน (gene product) ในสิ่งมีชีวิตใด ๆ กลุ่มนักปรสิตวิทยาได้ร่างกลุ่มของคำใหม่ ๆ รวมเข้าไปใน GO system ด้วย ซึ่งกลุ่มของคำเหล่านี้จะใช้อธิบายในแง่ชีววิทยาของปรสิต เช่น 'rosetting' จะถูกจัดอยู่ใน biological process ontology เป็นต้น การจัดประเภทของผลิตภัณฑ์ของยีน จะทำให้เพิ่มความเข้าใจในชีววิทยาของเชื้อด้วย (Gardner,2001)

งานวิจัยหลังการศึกษาจีโนมของเชื้อมาลาเรีย

เมื่อการหาลำดับเบสเสร็จสิ้นลงและทราบแล้วว่ามียีนใดบ้างบนโครโมโซมแต่ละแท่งของเชื้อมาลาเรีย งานวิจัยที่เกี่ยวข้องกับ functional genomic ซึ่งรวมถึงการแสดงออกของยีนและการสร้างโปรตีนก็เริ่มมีมากขึ้น เทคนิคหนึ่งที่ใช้ในการศึกษาการทำงานของยีน คือ transfection เป็นการนำชิ้นดีเอ็นเอที่สนใจศึกษาจากภายนอก (exogenous DNA) เข้าสู่เซลล์ของสิ่งมีชีวิต โดยส่วนของดีเอ็นเอที่สนใจนั้นจะเชื่อมกับพลาสมิดที่มี reporter gene อยู่ เช่น ยีนที่แปลรหัสได้เอนไซม์ luciferase ซึ่งจะมีคุณสมบัติในการเรืองแสง เมื่อชิ้นดีเอ็นเอเข้าสู่เซลล์ได้สำเร็จ ก็จะแทรกเข้าไปในจีโนม (recombination) ตรงตำแหน่งที่มีลำดับเบสตรงกัน (homologous) ดังแสดงในรูปที่ 4 ส่วนที่แทรกเข้าไปจะขัดขวางการทำงานของยีนหรือทำให้ส่วน open reading frame ของยีนนั้น ๆเปลี่ยนไป มีการนำเทคนิค transfection ไปใช้ในการศึกษาคุณสมบัติการเกาะติดของเม็ดเลือดแดงที่ติดเชื้อ *P. falciparum* กับผนังหลอดเลือด (cytoadhesion) ซึ่งอาศัยส่วนที่เรียกว่า knob ประกอบด้วย knob-associated histidine-rich protein (KAHRP) ซึ่งเป็นโปรตีนที่อยู่ภายในโครงสร้างของส่วนที่เป็น electron-dense knob บนผิวของเม็ดเลือดแดงที่ติดเชื้อ (infected erythrocyte) และ *P. falciparum* erythrocyte membrane protein 1 (PfEMP-1) ซึ่งเป็นโปรตีนที่จับกับ receptor ของเซลล์ผนังหลอดเลือด ผลการทดลองพบว่าเมื่อขัดขวางการทำงานของยีนที่แปลรหัสได้ KAHRP ความสามารถในการเกาะติดของเม็ดเลือดแดงที่ติดเชื้อกับผนังหลอดเลือดจะลดลงเมื่ออยู่ภายใต้สภาวะที่มีการไหลเวียนของกระแสเลือด ดังนั้น KAHRP มีความสำคัญต่อการสร้าง knob และโครงสร้างของ knob นี้เป็นปัจจัยสำคัญที่ช่วยให้โปรตีน PfEMP-1 จับกับ receptor ของเซลล์ผนังหลอดเลือดของโฮสต์ภายใต้สภาวะที่มีการไหลเวียนของกระแสเลือดได้ (Crabb *et al.*,1997) แต่เทคนิคนี้มีข้อจำกัดคือ เชื้อ *P. falciparum* ในระยะที่อยู่ในเม็ดเลือดแดงนั้นมีโครโมโซมเพียงชุดเดียว (haploid) ดังนั้นการศึกษาการทำงานของยีนที่มีความสำคัญต่อการดำรงชีพของเชื้อด้วยการขัดขวางการทำงานของยีนนี้จะมีผลทำให้เชื้อตาย นอกจากนี้การทำ transfection กับเชื้อ *P. falciparum* นั้นจะทำได้ยาก จึงมีการใช้ transfection system ของ *Toxoplasma gondii* เข้ามาช่วยในการวิเคราะห์หน้าที่ของยีนบางยีนของ

P. falciparum แทน (Waller *et al.*,1998) นอกจากเทคนิค transfection แล้วยังมีอีกหลายเทคนิคที่ใช้ในการวิเคราะห์หน้าที่ของยีน เช่น DNA microarrays และ serial analysis of gene expression (SAGE)



รูปที่ 4 กลไกการเกิด DNA recombination ในเชื้อ *Plasmodium* (Craig *et al*, 1999)
SM หมายถึง selectable marker หรือ reporter gene

สำหรับเทคนิค DNA microarrays นั้นคือการจุดชิ้นของดีเอ็นเอลงบนแผ่นสไลด์ จำนวนจุดอาจมากกว่า 10,000 จุดต่อสไลด์ ชิ้นดีเอ็นเอที่จุดลงไปนั้นได้จากการเพิ่มจำนวนชิ้นดีเอ็นเอจาก cDNA library หรือ GST library หรือใช้ primer ที่จำเพาะต่อบริเวณที่สนใจจาก genomic DNA แล้วแต่กรณีที่จะศึกษา จากนั้นจับคู่ (hybridize) กับ cDNA ที่ติดฉลากด้วยสารฟลูโอเรสเซนต์ (cDNA นี้ได้จาก RNA ที่มี poly A แล้วทำปฏิกิริยา RT-PCR) แล้วอ่านผลด้วยเครื่องสแกน (scanner) มีการใช้เทคนิคนี้ในการศึกษาการแสดงออกของยีนของเชื้อ *P. falciparum* ในระยะ gametocyte และระยะ trophozoite โดยใช้ DNA microarrays ที่ได้จาก GST library (ใช้ universal primer ของ cloning vector ในการเพิ่มจำนวน DNA) เมื่ออ่านผลก็จะทราบว่ายีนใดมีการถอดรหัส (trascription) เป็น mRNAในระยะใดระยะหนึ่งของการเจริญของเชื้อเท่านั้น (stage-specific gene) (Hayward *et al.*,2000)

อีกเทคนิคหนึ่งที่ใช้ในการศึกษาในระดับการถอดรหัสของยีน คือ SAGE เริ่มจากการสังเคราะห์ cDNA จาก mRNA ตัวอย่างที่สนใจ แล้วใช้เอนไซม์ตัดจำเพาะตัด cDNA ให้ได้ขนาด 10-20 เบส จากนั้นโคลนเข้ากับพลาสมิดและหาลำดับเบสของชิ้นดีเอ็นเอต่อไป ลำดับเบสที่ได้จะนำไปเทียบกับลำดับเบสของทั้งจีโนมในฐานข้อมูล เนื่องจากเชื้อมาลาเรียในระยะเม็ดเลือดแดงเท่านั้นที่สามารถเพาะเลี้ยงเพื่อเพิ่มจำนวนได้มากเพียงพอที่จะใช้ในการศึกษาด้วยเทคนิค DNA microarray หรือ SAGE จึงคาดว่าในอนาคตอาจมีการพัฒนาวิธีการอื่น ๆที่อาศัยพื้นฐานของเทคนิค PCR ในการศึกษาการแสดงออกของยีนของเชื้อในระยะอื่น ๆต่อไป (Carucci,2000)

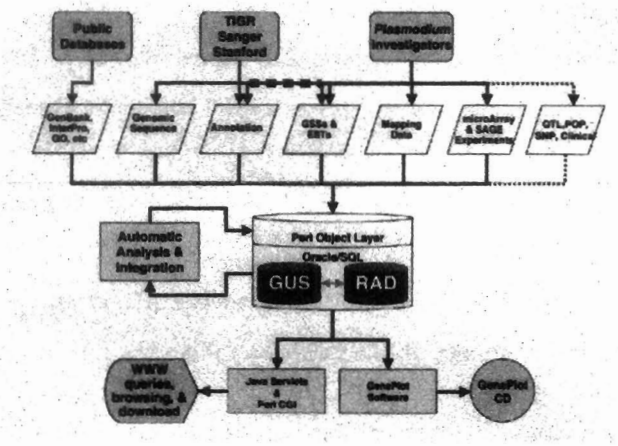
การศึกษาหน้าที่ของยีนในจีโนม (functional genomic) มีความหมายรวมไปถึงการแปลรหัสจนได้โปรตีนซึ่งเป็นผลิตภัณฑ์ของยีนด้วย โปรตีนเหล่านี้อาจมีหน้าที่หนึ่งหน้าที่ใดในกระบวนการสำคัญของเชื้อ

ก็เป็นได้ ดังนั้นการศึกษาเพียงระดับการถอดรหัสจึงไม่เพียงพอ เนื่องจากความสัมพันธ์ระหว่างระดับ mRNA กับระดับของโปรตีนนั้นยังไม่ชัดเจน การศึกษาถึงระดับการแปลรหัสจึงเป็นสิ่งที่ต้องดำเนินการศึกษาต่อไป ซึ่งที่ผ่านมาก็มีการนำข้อมูลลำดับเบสในจีโนมของ *P. falciparum* มาใช้ร่วมกับการวิเคราะห์โปรตีนที่ได้จากการแยกโดยวิธี 2-dimensional electrophoresis (2-DE) เมื่อแยกโปรตีนออกจากกันแล้ว เจลที่มีจุดโปรตีนอยู่จะถูกนำมาวิเคราะห์หาลำดับกรดอะมิโนที่อยู่ในสายเปปไทด์นั้น ๆด้วยเทคนิค mass spectrometry แล้วนำข้อมูลไปเปรียบเทียบกับลำดับกรดอะมิโนในสายเปปไทด์จากฐานข้อมูลที่แปลรหัสจากลำดับเบสของจีโนมของ *Plasmodium falciparum* (Horrocks *et al.*,2000)

PlasmoDB

ข้อมูลต่าง ๆจากโครงการและข้อมูลจากงานวิจัยที่เกี่ยวข้องในขณะนี้ได้เก็บรวมรวมในฐานข้อมูล PlasmoDB ซึ่งประกอบไปด้วยข้อมูลของลำดับเบส (จีโนม GSS และ EST) ข้อมูลแผนที่โครโมโซม (microsatellite และ optical mapping) ข้อมูลของ gene annotation และข้อมูลที่ได้จากการแสดงออกของยีน (EST, SAGE และ microarrays) รวมถึงข้อมูลที่เกิดจากการวิเคราะห์ลำดับเบสในจีโนม ทั้งการทำนายยีน การทำนายรูปร่างของโปรตีน (เช่น signal peptide, secondary structure, amino acid content เป็นต้น) และ Gene Ontology นอกจากนี้ยังมีข้อมูลของเชื้อ *P. yoelii*, *P. chabaudi*, *P. berghei* และ *P. vivax* ด้วย

ฐานข้อมูลนี้แบ่งการจัดเก็บเป็น 2 ระบบ คือ GUS (Genomics Unified Shema) จัดเก็บข้อมูลที่เป็นลำดับเบสซึ่งเป็นข้อมูลพื้นฐาน ส่วน RAD (RNA Abundance Database) จะจัดเก็บข้อมูลการแสดงออกของยีน สามารถใช้ข้อมูลในฐานข้อมูล PlasmoDB ในการจำแนกยีนที่สนใจได้ทาง http://PlasmooDB.org โดยใช้คำสำคัญ (keyword) หรือลำดับเบสที่สนใจ หรือดูลำดับเบสของโครโมโซมที่หาลำดับเบส และ ทำ annotation เสร็จสมบูรณ์แล้ว ซึ่งเมื่อเดือนตุลาคม พ.ศ.2544 มีรายงานว่าการหาลำดับเบสของโครโมโซม *P. falciparum* จากทั้งหมด 14 แท่งเสร็จสมบูรณ์แล้ว 7 แท่ง และได้จัดเก็บในฐานข้อมูลนี้เป็นที่เรียบร้อยแล้ว สำหรับผู้ที่ปฏิบัติงานภาคสนามหรืออยู่ในสถานที่ที่บริการอินเตอร์เนทเข้าไปไม่ถึงก็สามารถเข้าถึงข้อมูลของโครโมโซมที่การหาลำดับเบสและทำ annotation เสร็จสมบูรณ์แล้วได้จากโปรแกรม GenePlot ซึ่งอยู่ในรูปของ CD-ROM (Bahl *et al.*,2002) (รูปที่ 5)



รูปที่ 5
แสดงการทำงานของฐานข้อมูล PlasmoDB

สรุป

การหาลำดับเบสทั้งจีโนมของ *P. falciparum* ได้รุดหน้าไปมาก จากข้อมูลที่เสร็จสมบูรณ์แล้วในโครโมโซมแท่งที่ 2 และแท่งที่ 3 พบว่ามีแอนติเจนยีนใหม่ 2 family และพบยีนที่แปลรหัสให้เอนไซม์ที่ทำหน้าที่ในกระบวนการสังเคราะห์กรดไขมันนั้นเป็นตัวอย่างที่ดีที่ได้จากโครงการนี้ เมื่อโครงการเสร็จสมบูรณ์เราก็จะได้ทำความเข้าใจในกระบวนการสำคัญต่าง ๆที่เกิดขึ้นภายในเซลล์ของเชื้อ ตลอดจนกลไกที่เชื้อจะพัฒนาไปสู่ระยะต่าง ๆในวงชีวิต จากข้อมูลที่มีอยู่ ณ ปัจจุบันนี้ผนวกกับข้อมูลที่จะได้จากการหาลำดับเบสทั้งจีโนมของมนุษย์ (human genome sequence) และการหาลำดับเบสทั้งจีโนมของพาหะนำโรคมาลาเรียที่สำคัญ คือ ยุงก้นปล่อง *Anopheles gambiae* (*Anopheles* genome sequence) ที่คาดว่าจะเสร็จสมบูรณ์ในปี พ.ศ. 2545 จะเป็นโอกาสให้เราได้ศึกษาถึงความสัมพันธ์ระหว่างโฮสต์ พาหะ และเชื้อปรสิต ซึ่งจะเป็นแหล่งข้อมูลสำคัญในการค้นคว้าพัฒนายาและวัคซีนป้องกันโรคมาลาเรียต่อไป

เอกสารอ้างอิง

สดศรี ไทยทอง. 2540. ชีววิทยาและพันธุกรรมของเชื้อมาลาเรีย. ใน : จันทรา เหล่าถาวร ศรชัย หลูอารีย์สุวรรณ บรรณาธิการ. มาลาเรีย. กรุงเทพมหานคร : ศักดิ์โสภาการพิมพ์, หน้า 16-28.

Bahl A., Brunk B., Coppel RL., Crabtree J., Diskin J., Fraunholz MJ. *et al.* 2002. PlasmoDB : the *Plasmodium* genome resource. An integrated database providing tools for accessing, analyzing and mapping expression and sequence data (both finished and unfinished). Nucleic Acids Res. 30, 87-90.

Bowman S., Lawson D., Basham D., Brown D., Chillingworth T., Churcher CM. *et al.* 1999. The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. Nature. 400, 532-538.

Carucci DJ. 2000. Malaria Research in the Post-genomic Era. Parasitol Today. 16, 434-437.

Crabb BS., Cooke BM., Reeder JC., Waller RF., Caruana SR., Davern KM. *et al.* 1997. Targeted gene disruption shows that knobs enable malaria-infected red cells to cytoadhere under physiological shear stress. Cell. 89,287-296.

Craig AG., Waters AP., and Ridley RG.. 1999. Malaria genome project task force: A post-genomic agenda for functional analysis. Parasitol Today. 15, 211-214.

Dame JB., Arnot DE., Bourke PF., Chakrabarti D., Christodoulou Z., Coppel RL. *et al.*1996. Current status of the *Plasmodium falciparum* genome project. Mol Biochem Parasitol. 79,1-12.

Feagin JE. 1992. The 6-kb element of *Plasmodium falciparum* encodes mitochondrial cytochrome genes. Mol Biochem Parasitol. 52,145-148.

Fleischmann RD., Adams MD., White O., Clayton RA., Kirkness EF., Kerlavage AR. *et al.* 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science. 269,496-512.

Gardner MJ. 2001. A status report on the sequencing and annotation of the *Plasmodium falciparum* genome. Mol Biochem Parasitol. 118,133-138.

Gardner MJ., Tettelin H., Carucci DJ., Cummings LM., Aravind L., Koonin EV. *et al.* 1998. Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. Science. 282, 1126-1132.

Hayward RE., DeRisi JL., Alfadhli S., Kaslow DC., Brown PO., and Rathod PK. 2000. Shotgun DNA microarrays and stage-specific gene expression in *Plasmodium falciparum* malaria. Mol Microbiol. 35, 6-14.

Hoffman SL., Bancroft WH., Michael G., James SL., Burroughs ECB., Stephenson JR. *et al.* 1997. Funding for malaria genome sequencing. Nature. 387, 647.

Horrocks P., Bowman S., Kyes S., Waters AP., and Craig A. 2000. Entering the post-genomic era of malaria research. Bull WHO. 78, 1424-1437.

Howe CJ. 1992. Plastid origin of an extrachromosomal DNA molecule from *Plasmodium*, the causative agent of malaria. J Theor Biol. 158, 199-205.

Jing J., Lai Z., Aston C., Lin J., Caricci DJ., Gardner MJ. *et al.* 1999. Optical mapping of *Plasmodium falciparum* Chromosome 2. Genome Res. 9,175-181.

Reddy GR., Chakrabarti D., Schuster SM., Ferl RJ., Almira EC., and Dame JB. 1993. Genes sequence tags from *Plasmodium falciparum* genomic DNA fragments prepared by the "genease" activity of mung bean nuclease. Proc Natl Acad Sci USA. 90, 9867-9871.

Thompson JK., and Cowman AF. 1997. A YAC contig and high resolution restriction map of chromosome 3 from *Plasmodium falciparum*. Mol Biochem Parasitol. 90, 537-542.

Vaudin M., Roopra A., Hillier L., Brinkman R., Sulston J., Wilson RK. *et al.* 1995. The construction and analysis of M13 libraries prepared from YAC DNA. Nucleic Acids Res. 23, 670-674.

Waller RF., Keeling PJ., Donald RGK., Striepen B., Handman E., Unnasch NL. *et al.* 1998. Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. Proc Natl Acad Sci USA. 95, 12352-12357.

World Health Organization. 1997. World malaria situation in 1994-Part I. Wkly Epidemiol Rec. 72, 269.