

วัตถุประสงค์

1. วิเคราะห์โครงสร้างและความสัมพันธ์ทางเชื้อสายฝ่ายแม่ระหว่างประชากรที่ศึกษา
2. ประเมินว่าปัจจัยทางด้านภาษาพูดหรือภูมิศาสตร์ส่งผลต่อโครงสร้างและและความสัมพันธ์ระหว่างประชากรที่ศึกษา

วิธีการทดลอง

1. กลุ่มตัวอย่างและวิธีการเก็บตัวอย่าง

การวิจัยครั้งนี้ได้รับการอนุมัติจากคณะกรรมการจริยธรรมการวิจัยในมนุษย์มหาวิทยาลัยขอนแก่น (รหัสโครงการ HE552167) กลุ่มตัวอย่างที่ใช้ศึกษา คือ กลุ่มชาติพันธุ์ที่อาศัยอยู่ในภาคตะวันออกเฉียงเหนือของประเทศไทย จำนวน 10 กลุ่ม คือ มอญ (MON) เขมร (KHM) ส่วย (SUY) โส้ (SOA) ชาวนน (BON) แสก (SAK) ญ้อ (YOH) ผู้ไท (PUT) กะเลิง (KAL) และลาวอีสาน (LAO) โดยประชากรที่ศึกษาสามารถแบ่ง 2 กลุ่มตามภาษาพูด คือ ประชากรที่พูดภาษาตระกูลออสโตรเอเชียติก (AA) (มอญ เขมร ส่วย โส้ และชาวนน) และประชากรที่พูดภาษาไทย-กะได (TK) (แสก ญ้อ ผู้ไท กะเลิง และลาวอีสาน) และแบ่งตามลักษณะภูมิศาสตร์ คือ ประชากรที่อาศัยอยู่ในแอ่งที่ราบสกลนคร (SK) (โส้ แสก ญ้อ ผู้ไท และกะเลิง) และประชากรที่อาศัยอยู่ในแอ่งโคราช (KR) (มอญ เขมร ส่วย ชาวนน และลาวอีสาน) (ตารางที่ 1 และภาพที่ 1)

ก่อนทำการเก็บตัวอย่างจะมีการสัมภาษณ์ประวัติครอบครัว เพื่อคัดกรองอาสาสมัครที่ไม่มี ความเกี่ยวข้องกับทางสายเลือด อย่างน้อย 2 ชั่วรุ่น จากนั้นทำการเก็บตัวอย่างทำโดยใช้แปรงเก็บเยื่อบุช่องปาก (buccal collection brush) (Qiagen, USA) ถูที่เนื้อเยื่อบริเวณกระพุ้งแก้มประมาณ 40 ครั้ง

2. การสกัดดีเอ็นเอและตรวจสอบคุณภาพและปริมาณของดีเอ็นเอ

นำตัวอย่างเยื่อบุช่องปากแก้มที่ได้มาสกัดดีเอ็นเอด้วยชุดสกัดดีเอ็นเอสำเร็จรูป Gentra Puregene Buccal Cell Kit ดังนี้

2.1 นำตัวอย่างเซลล์เยื่อบุช่องปากแก้มซึ่งติดมากับแปรงเก็บตัวอย่าง ใส่ลงในหลอดทดลอง ขนาด 1.5 มิลลิลิตร จากนั้นเติมสารละลาย (cell lysis solution) เพื่อให้เซลล์แตก ปริมาตร 300 ไมโครลิตร

2.2 เติมเอนไซม์ proteinase K ปริมาตร 2 ไมโครลิตร เพื่อย่อยโปรตีน

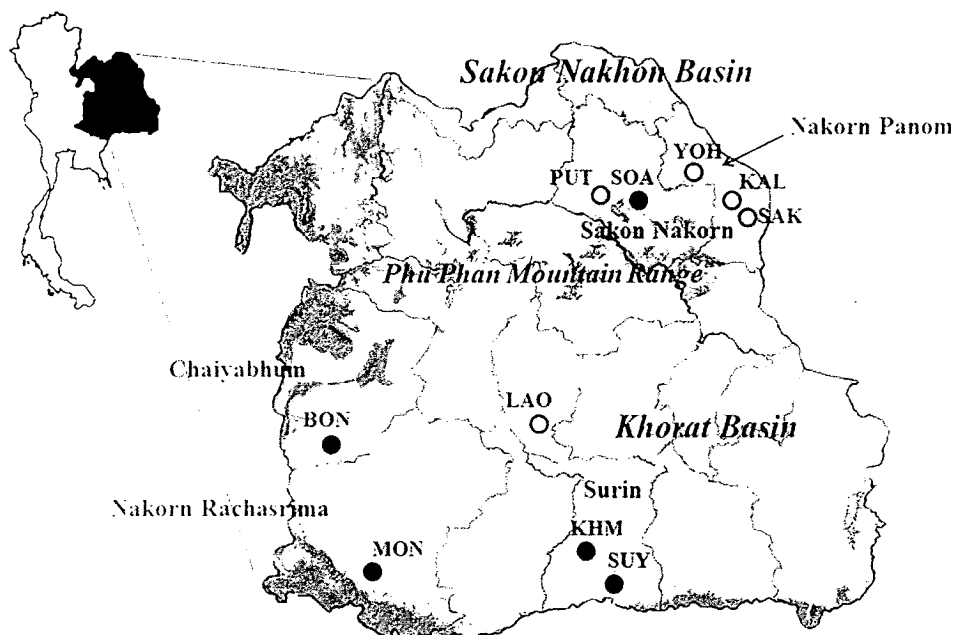
- 2.3 บ่มที่อุณหภูมิ 55 องศาเซลเซียส (ข้ามคืน)
- 2.4 นำแปร่งเก็บตัวอย่างออกจากหลอดทดลอง เติมนิวคลีเอส RNase ปริมาตร 2 ไมโครลิตรเพื่อย่อยอาร์เอ็นเอ
- 2.5 บ่มหลอดทดลองที่อุณหภูมิ 37 องศาเซลเซียส 15 นาที พลิกหลอดกลับไปกลับมา 25 ครั้ง
- 2.6 เติมนิวคลีเอสเพื่อย่อยโปรตีน (protein precipitation) ปริมาตร 100 ไมโครลิตร
- 2.7 เขย่าให้เข้ากันด้วยเครื่อง vortex 20 ครั้ง
- 2.8 บ่มหลอดทดลองในน้ำแข็ง 5 นาที
- 2.9ปั่นเหวี่ยงด้วยความเร็ว 15,000 รอบ / นาที 5 นาที ในขั้นตอนนี้ตะกอนที่ได้ (pellet) จะเป็นโปรตีนและเศษเซลล์
- 2.10 เทส่วนใสด้านบน (supernatant) ใส่หลอดทดลองใหม่ จากนั้นเติม isopropanol พลิกหลอดกลับไปกลับมา 50 ครั้ง
- 2.11 ปั่นเหวี่ยง ด้วยความเร็ว 15,000 รอบ/นาที 5 นาที
- 2.12 เทส่วนใสด้านบนทิ้ง จากนั้นเติม 70% ethanol ปริมาตร 300 ไมโครลิตร พลิกหลอดกลับไปกลับมา 20 ครั้ง
- 2.13 ปั่นเหวี่ยง ด้วยความเร็ว 15,000 รอบ/นาที 5 นาที และทิ้งไว้ให้แห้งที่อุณหภูมิห้อง ในขั้นตอนนี้ตะกอนที่ได้ (pellet) จะเป็นดีเอ็นเอ
- 2.14 เติมนิวคลีเอสละลายดีเอ็นเอ (DNA hydration) ปริมาตร 100 ไมโครลิตร และเก็บสารละลายดีเอ็นเอที่อุณหภูมิ -20 °C
- 2.15 ตรวจสอบคุณภาพและปริมาณดีเอ็นเอที่สกัดได้โดยวัดค่าการดูดกลืนแสงอัลตราไวโอเล็ต ที่ความยาวคลื่น 260 นาโนเมตร และ 280 นาโนเมตร ด้วยเครื่องสเปกโตรโฟโตมิเตอร์และวัดการเรืองแสงของดีเอ็นเอที่จับตัวกับเอธิเดียมโบรไมด์หลังจากแยกขนาดดีเอ็นเอโดยวิธีอิเล็กโตรโฟรีซิส โดยใช้วุ้นอะกาโรสความเข้มข้น ร้อยละ 1

ตารางที่ 1 ข้อมูลทั่วไปของประชากรที่ศึกษาและค่าทางสถิติที่ใช้วิเคราะห์

ประชากร	เขมร	มอญ	ส่วย	ชาวน	ໄສ້
อักษรย่อ	KHM	MON	SUY	BON	SOA
ละติจูด	14.9	14.69	15.01	15.59	17.37
ลองติจูด	103.49	102.06	103.94	101.46	104.3
ภาษา	AA	AA	AA	AA	AA
ภูมิศาสตร์	KR	KR	KR	KR	SN
ตำแหน่งที่ตั้ง (อำเภอ, จังหวัด)	สังขะและชุมพลบุรี, สุรินทร์	ปักธงชัย, นครราชสีมา	ลำโรงทาน, สุรินทร์	เทพสถิตย์, ชัยภูมิ	กุสุมาลย์, สกลนคร
จำนวนตัวอย่าง	68	44	44	42	47
ขนาดประชากร	1,266,828	1,000	407,724	6,283	71,532
แฮปโลไทป์	37	23	22	12	27
ชนิด Unique	24	19	12	10	16
ชนิด Single unique	20	12	11	6	12
ชนิด Multiple unique	4	7	1	4	4
ชนิด Non-unique	13	4	10	2	11
h	0.9583	0.9545	0.9397	0.8583	0.9584
π	0.013	0.0098	0.0143	0.0121	0.0141
Intra MPD	7.3995	5.5254	8.1057	6.8269	8.0324
Polymorphic site	54	40	47	23	48
Tajima's D	-1.0596	-1.3277	-0.9632	1.1116	-0.9327
(p-value)	-0.137	-0.071	-0.173	-0.899	-0.185
Fu's F_s	-17.1136	-8.3834	-3.8913	1.4256	-8.3333
(p-value)	0	-0.008	-0.113	-0.761	-0.008
r	0.0204	0.0195	0.0332	0.0572	0.0154
AA = Austro-Asiatic linguistic family; TT = Tai-Kadai linguistic family; KR= Khorat Basin; SN = Sakon Nakorn Basin ^a Population size estimated in Northeastern Thailand h = haplotype diversity; π = nucleotide diversity; r = a raggedness index value					

ตารางที่ 1 (ต่อ) ข้อมูลทั่วไปของประชากรที่ศึกษาและค่าทางสถิติที่ใช้วิเคราะห์

ประชากร	ลาวอีสาน	ภูไท	ญ้อ	แสก	กะเลิง
อักษรย่อ	LAO	PUT	YOH	SAK	KAL
ละติจูด	15.62	17.28	17.55	17.45	17.33
ลองจิจูด	103.5	103.65	104.09	104.74	104.59
ภาษา	TT	TT	TT	TT	TT
ภูมิศาสตร์	KR	SN	SN	SN	SN
ตำแหน่งที่ตั้ง (อำเภอ, จังหวัด)	เกษตรวิสัย, ร้อยเอ็ด	วาริชภูมิ, สกลนคร	นาหว้า, สกลนคร	เมือง, นครพนม	กุรุคุ, นครพนม
จำนวนตัวอย่าง	35	38	41	28	46
ขนาดประชากร	11,135,493	457,411	406,738	3,535	68,431
แฮปโลไทป์	30	23	20	11	21
ชนิด Unique	21	14	9	6	11
ชนิด Single unique	17	10	7	4	9
ชนิด Multiple unique	4	4	2	2	2
ชนิด Non-unique	9	9	11	5	10
h	0.9899	0.9573	0.9402	0.792	0.9063
π	0.0149	0.0153	0.0131	0.0114	0.0115
Intra MPD	8.4924	8.6956	7.4317	6.4929	6.5266
Polymorphic site	54	47	39	33	35
Tajima's D	-1.3016	-0.8134	-0.6458	-0.8067	-0.5512
(p-value)	-0.078	-0.227	-0.29	-0.22	-0.317
Fu's F_s	-19.0744	-5.6044	-3.3709	0.3691	-4.3474
(p-value)	0	-0.04	-0.119	-0.563	-0.076
r	0.0095	0.0108	0.0203	0.0694	0.0399
AA = Austro-Asiatic linguistic family; TT = Tai-Kadai linguistic family; KR= Khorat Basin; SN = Sakon Nakorn Basin ^a Population size estimated in Northeastern Thailand h = haplotype diversity; π = nucleotide diversity; r = a raggedness index value					



ภาพที่ 1 แผนที่ภาคตะวันออกเฉียงเหนือของประเทศไทย และที่ตั้งของประชากรที่ศึกษา ตัวย่อของประชากรแสดงในตารางที่ 1 ● แสดงประชากรที่พูดภาษาตระกูลออสโตรเอเชียติก ○ แสดงประชากรที่พูดภาษาตระกูลไท-กะได

3. การเพิ่มขยายชิ้นส่วนของดีเอ็นเอไมโทคอนเดรียด้วยปฏิกิริยาลูกโซ่พอลิเมอเรส (พีซีอาร์)

เพิ่มปริมาณดีเอ็นเอ ไมโทคอนเดรียบริเวณ D-loop ด้วยปฏิกิริยาลูกโซ่พอลิเมอเรส โดยใช้เอนไซม์ *Pfu* DNA polymerase (Enzymomics, Daejeon, Korea) โดยใช้ไพรเมอร์ 1 คู่ (Schurr et al., 1999) จำเพาะกับบริเวณ D-loop ซึ่งมีลำดับเบสดังนี้

ไพรเมอร์ LHmt: 430 CTG TTA AAA GTG CAT ACC GCC 410

ไพรเมอร์ LLmtA: 15704 CAT AGC CAA TCA CTT TAT TG 15723

โดยมีสารในปฏิกิริยาดังนี้

	ปริมาตร (ไมโครลิตร)
10 X nPfu-Forte PCR buffer (รวม $MgCl_2$)	5.00
200 ไมโครโมล dNTP	5.00
5 ไมโครโมล ไพรเมอร์ LHmt	2.50
5 ไมโครโมล ไพรเมอร์ LLmtA	2.50
2.5 U/μl <i>Pfu</i> polymerase	0.50
50 นาโนกรัม/ไมโครลิตร ดีเอ็นเอ	0.50
น้ำกลั่นปราศจากเชื้อ	34.00
รวม	50.00

ทำปฏิกิริยาอุณหภูมิสูงในเครื่องควบคุมอุณหภูมิอัตโนมัติ (Thermal cycler) โดยใช้อุณหภูมิต่างๆ ดังนี้

ช่วงที่ 1 (1 รอบ)	95 องศาเซลเซียส	2 นาที
ช่วงที่ 2 (35 รอบ)	95 องศาเซลเซียส	30 วินาที (denaturation)
	56 องศาเซลเซียส	1 นาที (annealing)
	72 องศาเซลเซียส	1 นาที (extension)
ช่วงที่ 3 (1 รอบ)	72 องศาเซลเซียส	5 นาที

จากนั้นตรวจสอบผลผลิตพีซีอาร์ ซึ่งมีขนาดประมาณ 1,200 คู่เบส ด้วยวิธีอิเล็กโทรโฟรีซิส โดยใช้วุ้นอะกาโรสความเข้มข้น ร้อยละ 1 โดยเทียบขนาดผลผลิตพีซีอาร์กับดีเอ็นเอมาตรฐาน (DNA ladder) (Norgen Biotek Corp, Thorold Ontario, Canada)

4. หาลำดับเบส

ทำการหาลำดับเบสของผลผลิตพีซีอาร์ โดยใช้ไพรเมอร์ที่จำเพาะอีกคู่ ที่มีลำดับเบสดังนี้

ไพรเมอร์ Forward 15897 (5') GTATAAACTAATACACCAGTCTTGT-15921(3')

ไพรเมอร์ Reverse 100 (5') CAGCGTCTCGCAATGCTATCGCGTG-76(3')

การหาลำดับเบสจะทำการหาลำดับเบสของดีเอ็นเอทั้งสองสาย (สาย H และ สาย L) เพื่อเป็นการยืนยันชนิดของเบส โดยส่งผลผลิตพีซีอาร์ไปหาลำดับเบสยังหน่วยบริการ MacroGen กรุงโซล ประเทศเกาหลี โดยใช้ชุดน้ำยาสำเร็จรูป BigDye Terminator Cycle Sequencing Kit v3.1 (Applied Biosystems, USA) และเครื่องหาลำดับเบสอัตโนมัติ รุ่น ABI3730 (Applied Biosystems) ลำดับเบสบริเวณ HVRI จำนวน 433 ตัวอย่างของประชากรที่ศึกษาได้ถูกส่งไปเก็บไว้ในฐานข้อมูล NCBI (The National Center for Biotechnology Information) (accession numbers KJ205639-KJ206068).

5. การวิเคราะห์ข้อมูล

5.1 ทำการ assembly และ alignment ลำดับเบสของดีเอ็นเอไมโทคอนเดรียสาย H และ L ด้วยโปรแกรม Seqscape v.2.7 demo (Applied Biosystem)

5.2 วิเคราะห์ลำดับเบสที่มีความหลากหลาย (polymorphic site) ที่ต่างจากลำดับอ้างอิง (Revised Cambridge reference sequence) วิเคราะห์ความหลากหลายของนิวคลีโอไทด์ (nucleotide diversity, π) ด้วยโปรแกรม DNASP v.5 (Librado and Rozas, 2009)

5.3 วิเคราะห์ความหลากหลายของแฮปโลไทป์ (haplotype diversity, h) ชนิดของแฮปโลไทป์ ค่า mean number of pairwise difference (MPD) ค่าพารามิเตอร์ของการเพิ่มจำนวนประชากร (demographic expansion) เช่น raggedness index value (r) (Harpending, 1994) และ neutrality test คือ Fu's F_s (Fu, 1997) และ Tajima's D (Tajima, 1989)

5.4 คำนวณหาระยะห่างทางพันธุกรรมระหว่างประชากรแบบ pairwise difference (F_{st}) และทดสอบความมีนัยสำคัญด้วยค่า permutations จำนวน 1000 ครั้ง ด้วยโปรแกรม ARLEQUIN v.3.5

และนำเมทริกซ์ของระยะห่างทางพันธุกรรมมาสร้างเป็นแผนภูมิแสดงความสัมพันธ์ทางพันธุกรรมแบบ Multidimensional Scaling (MDS) ด้วยโปรแกรม STATISTICA v.7 (StateSoft Software Ltd.)

5.5 ทำการวิเคราะห์การจัดกลุ่มของประชากรด้วยข้อมูลลำดับเบสและพิกัดทางภูมิศาสตร์โดยวิธี Spatial analysis of molecular variance (SAMOVA) ด้วยโปรแกรม SAMOVA v.1.0 (Reference)

5.6 วิเคราะห์โครงสร้างทางพันธุกรรมของประชากร ด้วยวิธี Analysis of Molecular Variance (AMOVA) (Excoffier *et al.*, 1992) โดยหาค่าความผันแปรทางพันธุกรรมของประชากรทั้ง 3 ระดับ คือ ระหว่างประชากร ระหว่างกลุ่มย่อยในประชากรเดียวกัน และภายในกลุ่มย่อย โดยการกำหนดประชากร ตามภาษาพูดและลักษณะทางภูมิศาสตร์ แล้วทดสอบ ค่าความความมีนัยสำคัญทางสถิติด้วยวิธี non parametric permutation ด้วยโปรแกรม ARLEQUIN v.3.5 (Excoffier and Lischer, 2010) (ตารางที่ 1)

5.7 การทดสอบเมนเทล (Mantel test)

ทำการทดสอบความสัมพันธ์ระหว่างเมทริกซ์จำนวน 3 คู่ คือระยะห่างทางพันธุกรรมและระยะห่างทางภูมิศาสตร์ ระยะห่างทางพันธุกรรมและระยะห่างทางภาษา และระยะห่างทางภูมิศาสตร์และระยะห่างทางภาษา ด้วยวิธีของเมนเทล (Mantel, 1967) ตารางที่ 2 จะแสดงระยะห่างทางภูมิศาสตร์และระยะห่างทางภาษา และตารางที่ 3 แสดงระยะห่างทางพันธุกรรมแบบ (F_{st})

ตารางที่ 2 ครึ่งล่างซ้ายแสดงระยะห่างทางภาษา (dLAN) ระหว่างคู่ของประชากร ซึ่งถูกกำหนดด้วยหมายเลข 4 ถึง 1 ตามการจัดกลุ่มภาษาของ Ethnologue (Lewis, 2009) โดยค่า dLAN 4 จะเป็นระยะห่างทางภาษาของตระกูลภาษาที่ต่างกัน (AA and TK); dLAN 3 จะเป็นระยะห่างทางภาษาของตระกูลภาษาเดียวกัน แต่ต่างกิ่ง (branch) เช่น MON และ SUY; dLAN 2 จะเป็นระยะห่างทางภาษาของกิ่งเดียวกัน แต่ต่างกิ่งย่อย (sub-branch); dLAN 1 จะเป็นระยะห่างทางภาษาที่อยู่กิ่งย่อยเดียวกัน ครึ่งบนขวาแสดงระยะห่างทางภูมิศาสตร์ (great-circle distance) ระหว่างคู่ของประชากรโดยจะใช้พิกัดของละติจูดและลองจิจูดของที่ตั้งหมู่บ้านของแต่ละประชากร

	KHM	MON	SUY	BON	SOA	LAO	PUT	YOH	SAK	KAL
KHM		155.63	50.07	237.46	286.91	79.85	264.69	300.36	312.93	293.54
MON	3		205.48	118.45	380.8	185.71	333.54	383.36	409.11	397.83
SUY	2	3		274.47	264.15	82.47	253.75	281.65	283.79	265.91
BON	3	3	3		361.54	219.26	300.24	354.87	406.89	385.84
SOA	2	3	1	3		211.57	69.04	29.61	48.38	31.53
LAO	4	4	4	4	4		184.99	222.63	242.41	221.96
PUT	4	4	4	4	4	1		54.84	117.38	99.7
YOH	4	4	4	4	4	1	1		70.31	58.55
SAK	4	4	4	4	4	2	2	2		21.41
KAL	4	4	4	4	4	1	1	1	1	

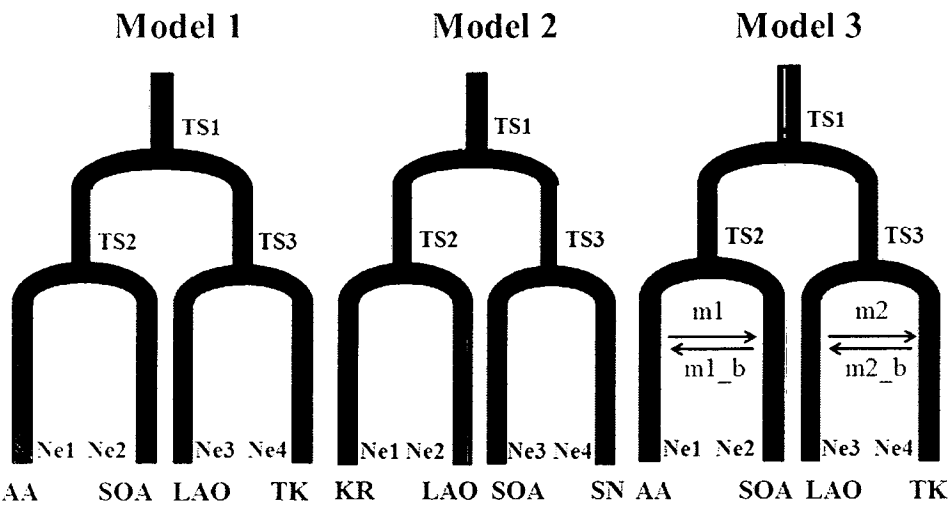
5.8 การเลือกโมเดลด้วย Approximate Bayesian Computation

สร้างโมเดลแสดงความสัมพันธ์ทางวิวัฒนาการระหว่างประชากร ดังภาพที่ 2 โดยพิจารณาจากตำแหน่งที่ตั้งทางภูมิศาสตร์และภาษาพูด โดยประชากรที่เป็นตัวแปรคือโล (SOA) และ ลาวอีสาน (LAO) เนื่องจาก SOA เป็นประชากรที่พูดภาษาตระกูลออสโตรเอเชียติก (AA) เพียงประชากรเดียวที่อาศัยอยู่ในแอ่งที่ราบสกลนคร (SN) และ LAO เป็นประชากรที่พูดภาษาตระกูลไท-กะได (TK) เพียงประชากรเดียวที่อาศัยอยู่ในแอ่งที่ราบโคราช (KR)

โมเดลที่ 1 แสดงความสัมพันธ์ระหว่างประชากรโดยปัจจัยด้านภาษาส่งผลต่อรูปแบบความสัมพันธ์ โดยประชากร AA และ TK แยกออกจากกันเมื่อระยะเวลา Ts1 และหลังจากนั้นประชากร SOA แยกจากประชากร AA เมื่อระยะเวลา Ts2 และประชากร LAO แยกจาก TK เมื่อระยะเวลา Ts3

โมเดลที่ 2 แสดงความสัมพันธ์ระหว่างประชากรโดยปัจจัยด้านภูมิศาสตร์ส่งผลต่อรูปแบบความสัมพันธ์ โดยประชากรที่อาศัยอยู่ในแอ่ง SN และ KR แยกออกจากกันเมื่อระยะเวลา Ts1 และหลังจากนั้นประชากร SOA แยกจากประชากรที่อาศัยอยู่ในแอ่ง SN เมื่อระยะเวลา Ts2 และประชากร LAO ลาวอีสานแยกจากประชากรที่อาศัยอยู่ในแอ่ง KR เมื่อระยะเวลา Ts3

โมเดล 3 จะเพิ่มเติมจากโมเดล 1 โดยหลังจากที่ประชากร SOA แยกจาก AA และ LAO แยกจาก TK จากนั้นประชากรที่อยู่ในแอ่งเดียวกันจะมีการอพยพเข้าและออกระหว่างกัน กล่าวคือ LAO และ AA จะมีอัตราการอพยพพลงที่ (m1 และ m1_b) และ SOA และ TK จะมีอัตราการอพยพพลงที่ (m2 และ m2_b)



ภาพที่ 2 โมเดลแสดงความสัมพันธ์ระหว่างประชากร ซึ่งได้รับอิทธิพลของภาษา (โมเดล 1) ภูมิศาสตร์ (โมเดล 2) และ การอพยพ (โมเดล 3) Ne Ts และ m คือ effective population sizes ระยะเวลาในการแยกระหว่างประชากร และ อัตราการกลายพันธุ์ ตามลำดับ อักษรย่อของประชากรแสดงในตารางที่ 1

ทำการ simulate ข้อมูลของลำดับเบสดีเอ็นเอไมโทคอนเดรียและ prior distributions (ตารางที่ 3) โดยอาศัยทฤษฎีของ coalescent เพื่อให้ได้ค่าทางสถิติที่มีค่าใกล้เคียงกับค่าสังเกตมากที่สุด (มี Euclidean distance น้อยที่สุด) จากนั้นคำนวณ posterior probabilities ด้วยวิธี Approximate Bayesian Computation (ABC) (Bertorelle *et al.*, 2010)

การศึกษานี้จะใช้ ABC จำนวน 2 รูปแบบ คือ acceptance-rejection procedure (AR) and weighted multinomial logistic regression (LR) (Pritchard *et al.*, 1999; Beaumont, 2008) ในการประเมินความเสถียรของค่า posterior probabilities จะทำการ simulate ข้อมูลหลายครั้ง โดยรูปแบบ AR จะ simulate จำนวน 100 200 300 และ 500 ครั้ง ในขณะที่รูปแบบ LR จะ simulate จำนวน 25000 50000 75000 และ 100000 ครั้ง การ simulate ข้อมูล จะอาศัยโปรแกรม ABC tool box (Wegmann, 2010) จากนั้นทำการเขียน R scripts ซึ่งดัดแปลงมาจาก SG (<http://code.google.com/p/popabc/source/browse/#svn%2Ftrunk%2Fscripts>) เพื่อคำนวณค่า posterior probabilities ของแต่ละโมเดล

จากนั้นทำการประเมินความถูกต้องในการคัดเลือกโมเดลด้วย Type I error โดยจะกำหนดระดับความน่าจะเป็นของการคัดเลือก (decision probability thresholds) คือ 0.5 0.6 0.7 0.8 และ 0.9 เพื่อประสิทธิภาพของวิธี ABC ในการคัดเลือกโมเดล

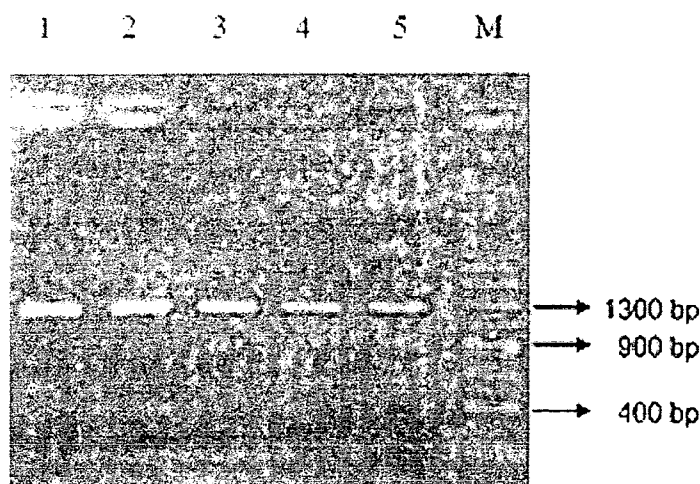
ตารางที่ 3 Prior distributions ของพารามิเตอร์ในแต่ละโมเดล โดย N แสดง effective population sizes T แสดงระยะเวลาในการแยกระหว่างประชากร หน่วยเป็นชั่วรุ่น และ m แทนอัตราการกลายพันธุ์ หน่วยเป็นต่อเบสต่อชั่วรุ่น

Parameter	Distribution	Lower limit	Upper limit
N0	logunif	50,000	500,000
N1	logunif	1,000	100,000
N2	logunif	1,000	100,000
N3	logunif	50,000	500,000
T2	unif	10	20
T3	unif	10	20
T1	unif	100	200
Mutation	unif	0.000000878	0.000021
m1	unif	0.001	0.1
m2	unif	0.001	0.1
m1_b	unif	0.001	0.1
m2_b	unif	0.001	0.1

ผลการทดลอง

1. ผลผลิตจากการเพิ่มปริมาณดีเอ็นเอ

จากการเพิ่มปริมาณดีเอ็นเอไมโทคอนเดรียบริเวณ D-loop ได้ขนาดผลผลิตจากการทำปฏิกิริยา ลูกโซ่พอลิเมอร์สประมาณ 1,200 คู่เบส เทียบขนาดกับ 100 bp DNA Ladder (Norgen Biotek Corp, Thorold Ontario, Canada) (ภาพที่ 3)



ภาพที่ 3 ผลผลิตจากการทำปฏิกิริยาลูกโซ่พอลิเมอร์ส ช่องที่ 1-5 และ คือ ดีเอ็นเอขนาด 1,200 คู่เบส หลังจากการทำปฏิกิริยาลูกโซ่พอลิเมอร์ส ช่อง M คือ 100 bp DNA Ladder

2. ความหลากหลายทางพันธุกรรม

เมื่อนำผลผลิตจากการทำปฏิกิริยาลูกโซ่พอลิเมอร์สไปหาลำดับเบส บริเวณ HVR-I ทั้งสาย H และ L และทำการรวมลำดับเบสทั้งสองสายแล้ว จะได้ลำดับเบสที่มีความยาว 596 คู่เบส (ตำแหน่งที่ 16001 ถึง 16596) จากตัวอย่างดีเอ็นเอที่ใช้ศึกษาจำนวน 433 ตัวอย่าง พบแฮปโลไทป์ที่แตกต่างกันทั้งหมด 173 แฮปโลไทป์ จากตำแหน่งที่เกิดความผันแปร (polymorphic site) ทั้งหมด 135 ตำแหน่ง (ภาคผนวก 1) โดยเป็นแฮปโลไทป์ที่พบเพียงตัวอย่างเดียว (single unique) จำนวน 108 แฮปโลไทป์ และพบแฮปโลไทป์ที่พบมากกว่า 1 ตัวอย่าง แต่พบเพียงประชากรเดียว (multiple unique) จำนวน 34 ตัวอย่าง ส่วนที่เหลืออีก 34 แฮปโลไทป์ จะเป็นแฮปโลไทป์ที่พบมากกว่า 1 ประชากร (shared haplotypes) โดยประชากรที่มี shared haplotypes จำนวนมากที่สุด คือ ประชากร SOA-PUT จำนวน 6 แฮปโลไทป์ ในขณะที่ไม่พบแฮปโลไทป์ชนิด shared ระหว่างคู่ของประชากรจำนวน 5 คู่ ดังนี้ MON-

BON, MON-LAO, MON-PUT, BON-KHM และ BON-SOA (ตารางที่ 3) และแฮปโลไทป์ชนิดที่พบมากที่สุดคือแฮปโลไทป์ที่ 25 (hap25) และ 68 (hap68) ซึ่งพบแฮปโลไทป์ทั้ง 2 จำนวน 20 และ 19 ตัวอย่าง (ภาคผนวก 1 และ 2)

ตารางที่ 4 ครึ่งล่างซ้ายแสดงระยะห่างทางพันธุกรรมระหว่างประชากร แบบ pairwise F_{st} และครึ่งล่างขวาแสดงจำนวนแฮปโลไทป์ที่เหมือนกันระหว่างประชากร อักษรย่อของประชากรแสดงในตารางที่ 1

	KHM	MON	SUY	BON	SOA	LAO	PUT	YOH	SAK	KAL
KHM		1	5	0	3	3	3	4	0	5
MON	0.1517		1	0	1	0	0	1	1	1
SUY	0.0469	0.1061		1	2	3	1	4	1	3
BON	0.0628	0.1537	0.0403		0	1	1	1	1	1
SOA	0.1006	0.1103	0.1306	0.1792		4	6	5	1	2
LAO	0.0264	0.0897	0.0488	0.0686	0.0513		3	2	2	2
PUT	0.0467	0.0916	0.0744	0.1055	0.0396	0.0260		3	1	1
YOH	0.0537	0.0775	0.0461	0.0851	0.0401	0.0233	0.0326		2	5
SAK	0.2280	0.2979	0.2827	0.3316	0.0632	0.1720	0.1581	0.1781		2
KAL	0.0516	0.2007	0.1230	0.1624	0.0406	0.0503	0.0519	0.0539	0.1184	

ตัวหนาแสดงระดับนัยสำคัญทางสถิติ ที่ $P < 0.01$

เมื่อกำหนดความหลากหลายของแฮปโลไทป์ (h) ซึ่งมีค่าสูงสุดในประชากร LAO (0.9899) และต่ำสุดในประชากร SAK (0.7920) ความหลากหลายของนิวคลีโอไทด์ (IT) มีค่าสูงสุดในประชากร PUT (0.0153) และต่ำสุดในประชากร MON (0.0098) ดังแสดงในตารางที่ 1 ค่าความหลากหลายทั้ง 2 อยู่ในช่วงเดียวกับ ประชากรอื่นในประเทศไทยที่เคยมีการศึกษามาก่อนหน้านี้ (Fucharoen *et al.*, 2001; Besaggio *et al.*, 2007; Lertrit *et al.*, 2008; Kutanan *et al.*, 2011a; Kutanan *et al.*, 2011b)

จากการคำนวณค่า intra-MPD ของประชากร จะมีค่าต่ำสุด ในประชากร MON (5.5254) และ สูงสุดในประชากร PUT (8.6956) (ตารางที่ 3) ซึ่งแสดงถึงการมีความเหมือนกันทางพันธุกรรม (genetic homogeneity) หรือการเกิด recent divergence ในตัวอย่างประชากรชาวมอญ ในขณะที่ตัวอย่างใน ประชากรชาวภูไทมีความผันแปรทางพันธุกรรม (genetic heterogeneity) มากที่สุด จากนั้นเมื่อ พิจารณาค่า inter-MPD ซึ่ง อยู่ในช่วงระหว่าง 7.2024 ถึง 10.2660 ช่วงดังกล่าวมีค่าสูง เมื่อ เปรียบเทียบกับผลการรายงานจากงานวิจัยก่อนหน้านี้ (Bodner *et al.*, 2011) ซึ่งแสดงถึงการมี โครงสร้างทางพันธุกรรมที่แตกต่างกันระหว่างประชากร

ตารางที่ 5 ค่า intra-MPD (เส้นทแยงมุม กรอบสี่เหลี่ยม) ค่า corrected MPD (ครึ่งล่างซ้าย) และ ค่า inter-MPD (ครึ่งบนขวา).

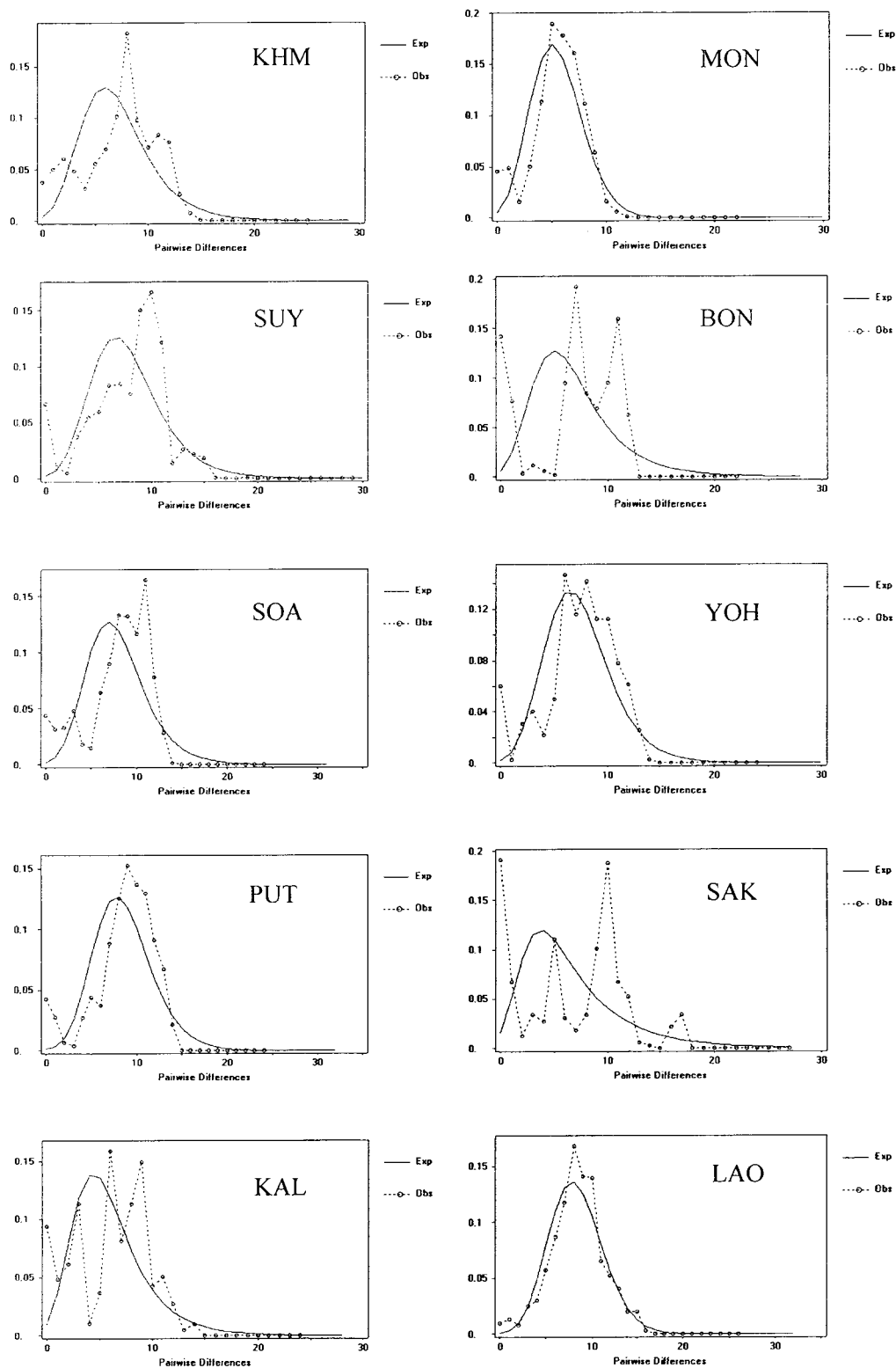
	KHM	MON	SUY	BON	SOA	LAO	PUT	YOH	SAK	KAL
KHM	7.1242	7.4937	7.9806	7.4471	8.4136	8.0042	8.2767	7.6869	8.8677	7.2024
MON	1.1689	5.5254	7.6245	7.2944	7.6267	7.6734	7.8098	7.0183	8.5008	7.5420
SUY	0.3656	0.8089	8.1057	7.7808	9.2810	8.7221	9.0730	8.1452	10.2660	8.3389
BON	0.4715	1.1182	0.3145	6.8270	9.0613	8.2143	8.6686	7.7921	9.9841	7.9695
SOA	0.8353	0.8478	1.2119	1.6316	8.0324	8.7058	8.7055	8.0566	7.7794	7.5883
LAO	0.1959	0.6645	0.4230	0.5546	0.4434	8.4924	8.8233	8.1484	9.0857	7.8932
PUT	0.3668	0.6993	0.6723	0.9073	0.3415	0.2293	8.6956	8.3331	9.0682	8.0172
YOH	0.4089	0.5398	0.3765	0.6628	0.3245	0.1864	0.2695	7.4317	8.4986	7.3743
SAK	2.0591	2.4917	2.9667	3.3242	0.5167	1.5931	1.4740	1.5364	6.4929	7.3849
KAL	0.3770	1.5160	1.0228	1.2927	0.3089	0.3837	0.4061	0.3952	0.8751	6.5266

3. ค่าพารามิเตอร์ของการเพิ่มจำนวนประชากร

การกระจายของจำนวนเบสที่แตกต่างกัน (mismatch distribution) เป็นการวิเคราะห์การเพิ่มจำนวนของประชากรอย่างรวดเร็วในอดีต (rapid population expansion) โดยแสดงจากกราฟการกระจายของจำนวนเบสที่แตกต่างกันระหว่างแฮปโลไทป์แต่ละคู่ในประชากรนั้น ถ้ากราฟจะเป็นรูประฆังคว่ำ มีจุดสูงสุดเพียงจุดเดียว (unimodal, smooth bell shape) และการมีค่า raggedness index น้อยกว่าหรือเท่ากับ 0.03 รวมทั้งการมีค่า neutrality ติดลบ จะหมายถึงการมีจำนวนประชากรเพิ่มขึ้นอย่างรวดเร็วในประชากรนั้น แต่ถ้าลักษณะของกราฟมีการกระจายตัวแบบไม่สม่ำเสมอ (multimodal, ragged) มีจุดสูงสุดมากกว่า 1 จุด มีค่า raggedness index มากกว่า 0.03 และค่า neutrality เป็นบวก จะบ่งบอกถึงการมีจำนวนประชากรที่คงที่มาเป็นระยะเวลายาวนาน

ภาพที่ 4 แสดงกราฟการกระจายของจำนวนเบสที่แตกต่างกัน ซึ่งเป็นกราฟจะเป็นรูประฆังคว่ำ มีจุดสูงสุดเพียงจุดเดียว ของประชากรชาว KHM MON SOA PUT และ LAO และค่า raggedness index ของประชากรทั้ง 3 มีค่าน้อยกว่า 0.03

Neutrality test เป็นการวิเคราะห์การคงที่และเพิ่มขยายของขนาดประชากร โดยมีการตั้งสมมุติฐาน (null hypothesis) ที่ว่าประชากรมีขนาดคงที่ (constant effective population size) ผลจากการวิเคราะห์ neutrality test ด้วยค่า Fu's F_s และ Tajima' D พบว่าค่า Fu's F_s ในประชากร KHM MON SOA LAO และ PUT มีค่าติดลบอย่างมากและมีนัยสำคัญทางสถิติ ($P < 0.05$) (ตารางที่ 1) จากการคำนวณ neutrality test พบว่าปฏิเสธ null hypothesis แม้ค่า Tajima' D จะไม่มีนัยสำคัญทางสถิติ (ดูอักษรย่อของประชากรในตารางที่ 1)



ภาพที่ 4 กราฟการกระจายของจำนวนเบสที่แตกต่างกันในแต่ละประชากรภายใต้โมเดล population growth-decline

4. ความสัมพันธ์ระหว่างประชากร

ตารางที่ 3 แสดงระยะห่างทางพันธุกรรมแบบ pairwise difference (F_{st}) และความแตกต่างอย่างมีนัยสำคัญทางสถิติ จากการเปรียบเทียบประชากรจำนวน 45 คู่ พบว่ามีจำนวน 36 คู่ (80%) ที่แตกต่างอย่างมีนัยสำคัญทางสถิติ ($P < 0.01$) ค่า F_{st} ที่มีค่าสูงและมีนัยสำคัญทางสถิติแสดงถึงความแตกต่างทางพันธุกรรมระหว่างคู่ของประชากรนั้น ซึ่งพบว่ามีค่า F_{st} ระหว่างประชากร MON BON และ SAK กับประชากรอื่นที่เหลือมีค่าสูงและมีนัยสำคัญทางสถิติ ในขณะที่ค่า F_{st} ระหว่างประชากร KHM กับประชากรอื่นที่เหลือมีค่าสูงและมีนัยสำคัญทางสถิติ ยกเว้นระหว่างประชากร KHM กับ LAO นอกจากนั้น SUY และ BON มีค่า F_{st} ต่ำและไม่มีความนัยสำคัญทางสถิติ ซึ่งแสดงถึงการมีโครงสร้างทางพันธุกรรมที่เหมือนกันระหว่างประชากร KHM กับ LAO และ SUY กับ BON เป็นที่น่าสนใจว่าประชากรที่อาศัยอยู่ในแอ่งที่ราบสกลนคร (SK) มีค่า F_{st} ที่แตกต่างกันอย่างไม่มีนัยสำคัญ SOA PUT และ YOH ซึ่งแสดงถึงการมีโครงสร้างทางพันธุกรรมที่เหมือนกัน ซึ่งอาจเกิดจากยีนโฟลว์ระหว่างประชากรในแอ่งที่ราบสกลนคร

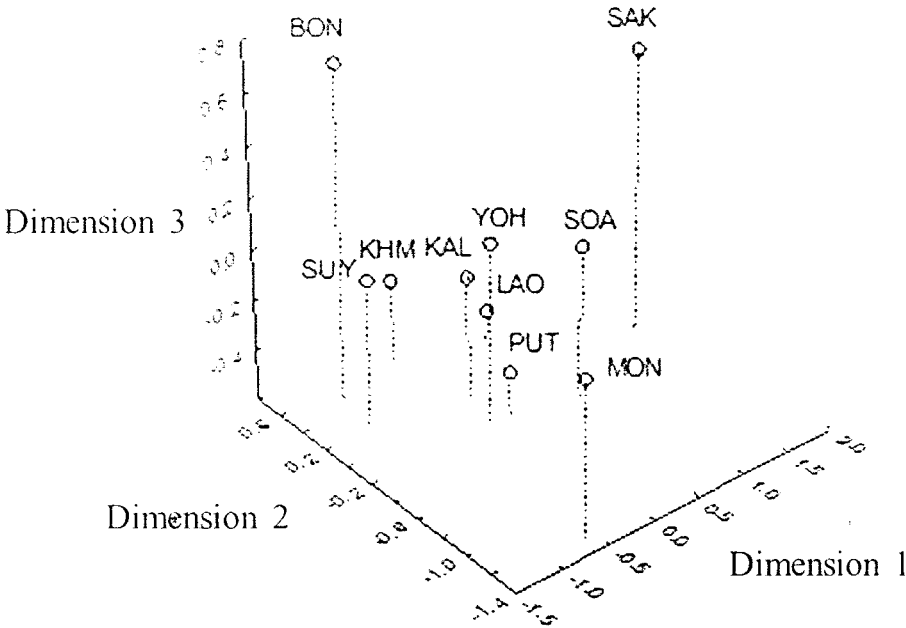
รูปแบบของความแตกต่างทางพันธุกรรมระหว่างประชากร จากการคำนวณค่า F_{st} และ corrected MPD มีความสอดคล้องกัน (ตารางที่ 4) โดยชาว SAK มีโครงสร้างทางพันธุกรรมที่แตกต่างจากประชากรอื่นมากที่สุด รองลงมาคือ BON MON และ SUY ตามลำดับ

เมื่อนำเมทริกซ์ของระยะห่างทางพันธุกรรมมาสร้างเป็นแผนภูมิแสดงความสัมพันธ์ทางพันธุกรรมแบบ multidimensional scaling (MDS) แบบ 3 มิติ เพื่อหาความสัมพันธ์ทางเชื้อสายระหว่างประชากร (ภาพที่ 5) พบว่าประชากรที่อาศัยอยู่ในแอ่งที่ราบสกลนครเกือบทั้งหมด ยกเว้นชาว SAK (YOH PUT SOA และ KAL) ถูกจัดอยู่ในกลุ่มเดียวกัน ซึ่งอยู่ตรงตำแหน่งตรงกลางของแผนภูมิ การที่ประชากรมีตำแหน่งที่ใกล้กันในแผนภูมิ แสดงถึงการมีความใกล้ชิดกันทางพันธุกรรม ประชากร SAK ที่มีตำแหน่งบนแผนภูมิห่างออกไป ซึ่งแสดงถึงการมีโครงสร้างทางพันธุกรรมที่แตกต่างไปอย่างมาก อย่างไรก็ตาม SAK แสดงความใกล้ชิดทางพันธุกรรมระหว่างประชากรในแอ่งที่ราบสกลนครมากกว่าแอ่งที่ราบโคราช (KR) เป็นที่น่าสนใจว่าแม้ว่าประชากร LAO จะอาศัยอยู่ในแอ่งที่ราบโคราช แต่ชาว LAO กลับถูกจัดอยู่ในกลุ่มเดียวกับประชากรในแอ่งที่ราบสกลนคร (SN) เมื่อพิจารณาประชากรที่อาศัยอยู่ในแอ่งที่ราบโคราชพบว่า KHM มีพันธุกรรมใกล้ชิดกับกับชาว SUY และ BON ในขณะที่ MON มีตำแหน่งที่ไกลออกไปจากประชากรอื่นในแผนภูมิ ซึ่งแสดงถึงการมีพันธุกรรมที่แตกต่างออกไป

ผลการวิเคราะห์การจัดกลุ่มประชากรด้วยวิธี SAMOVA (ตารางที่ 5) พบว่าจากการกำหนดกลุ่มของประชากร ด้วยการเพิ่มจาก 2 กลุ่ม ไปจนถึง 6 กลุ่ม ประชากรที่แยกออกมาจากประชากรอื่นประชากรแรก คือ SAK (2 กลุ่ม) จากนั้นเป็นประชากร MON (3 กลุ่ม) SUY (4 กลุ่ม) BON (5 กลุ่ม) และ KHM (6 กลุ่ม) ซึ่งการที่ประชากรใดแยกออกมาประชากรแรกจะแสดงการมีความแตกต่างทางพันธุกรรมจากประชากรอื่นอย่างมาก จากนั้นเมื่อเพิ่มจำนวนกลุ่มสูงขึ้นประชากรที่แยกออกมาก็จะมีความแตกต่างทางพันธุกรรมรองลงมาตามลำดับ จากการวิเคราะห์ความผันแปรทางพันธุกรรมระหว่างกลุ่มประชากร (F_{ct}) พบว่ามีค่าสูงสุดเมื่อมีการแบ่งกลุ่มประชากรที่ 6 กลุ่ม (0.0728, $P < 0.01$) ซึ่งกลุ่ม

ทั้ง 6 ประกอบด้วย กลุ่ม 1: SAK กลุ่ม 2: MON กลุ่ม 3: SUY กลุ่ม 4: BON กลุ่ม 5: KHM กลุ่ม 6: YOH PUT SOA KAL และ LAO

ซึ่งการที่ประชากร SAK MON SUY BON และ KHM แยกอยู่คนละกลุ่มแสดงถึงการมีพันธุกรรมที่แตกต่างกัน ในขณะที่ประชากร YOH PUT SOA KAL และ LAO ถูกจัดอยู่ในกลุ่มเดียวกัน แสดงถึงการมีพันธุกรรมที่คล้ายกัน ซึ่งผลของ SAMOVA มีความสอดคล้องกับ MDS



ภาพที่ 5 แผนภูมิแสดงความสัมพันธ์ทางพันธุกรรมแบบ multidimensional scaling (MDS) ที่สร้างจากระยะห่างทางพันธุกรรมแบบ pairwise difference (F_{st}) อักษรย่อของประชากรแสดงในตารางที่ 1

5. ความสัมพันธ์ระหว่างพันธุกรรม ภูมิศาสตร์ และภาษาพูด

เมื่อวิเคราะห์โครงสร้างและความแตกต่างทางพันธุกรรมด้วยวิธี AMOVA (ตารางที่ 6) โดยทำการแบ่งกลุ่มประชากรตามภาษาพูด (AA และ TK) และลักษณะทางภูมิศาสตร์ (SN และ KR) เมื่อพิจารณาค่าความผันแปรทางพันธุกรรมระหว่างกลุ่มประชากร พบว่าความผันแปรระหว่างกลุ่มประชากรสอดคล้องกับการแบ่งกลุ่มตามลักษณะทางภูมิศาสตร์ ($F_{ct} = 0.0468$, $P < 0.01$) เมื่อพิจารณาประชากรในแต่ละกลุ่มภูมิศาสตร์พบว่าความผันแปรระหว่างประชากร (among populations, F_{st}) ในแอ่งที่ราบสกลนคร ($F_{st} = 0.06902$, $P < 0.01$) และแอ่งโคราช ($F_{st} = 0.07900$, $P < 0.01$) มีค่าน้อยกว่าความผันแปรระหว่างประชากรรวม (overall F_{st} , 0.09889, $P < 0.01$) จากผลดังกล่าวแสดงถึงประชากรที่อาศัยอยู่ในแต่ละภูมิศาสตร์มีความใกล้ชิดทางพันธุกรรม โดยประชากรในแอ่งที่ราบสกลนครมีความใกล้ชิดกันมากกว่าประชากรในแอ่งที่ราบโคราช

แต่เมื่อพิจารณากลุ่มประชากรตามภาษาพูด พบว่าความผันแปรระหว่างกลุ่มประชากรไม่สอดคล้องกับการแบ่งกลุ่มตามภาษาพูด ($F_{ct} = 0.00913$, $P > 0.01$) ซึ่งความผันแปรทางพันธุกรรมส่วนใหญ่ (ร้อยละ 89.74) เกิดขึ้นภายในกลุ่มประชากร ($F_{sc} = 0.09434$, $P < 0.01$) ในขณะที่ ความผันแปรทางพันธุกรรมอีกร้อยละ 9.35 เกิดขึ้น ระหว่างประชากร ($F_{st} = 0.10260$, $P < 0.01$) โดยประชากร AA มีค่าความผันแปรทางพันธุกรรม ($F_{st} = 0.10681$, $P < 0.01$) มากกว่า TK ($F_{st} = 0.07820$, $P < 0.01$) ซึ่งแสดงถึงการมีโครงสร้างทางพันธุกรรมที่ต่างกันอย่างมากของประชากรที่พูดภาษาตระกูลมอญเขมรมากกว่าไท-กะได

การทดสอบแมนเทิลเพื่อวิเคราะห์ correlation test และ partial correlation test ระหว่างเมตริกซ์ของระยะห่างทางพันธุกรรม (ตารางที่ 3) ระยะห่างทางภาษา และระยะห่างทางภูมิศาสตร์ (ตารางที่ 2) ผลการศึกษาพบว่า ระยะห่างทางพันธุกรรมสอดคล้องกับระยะห่างทางภูมิศาสตร์ [correlation test ($r = 0.4713$, $P < 0.01$) และ partial correlation test ($r = 0.4449$, $P < 0.01$)] แต่พบว่า ระยะห่างทางพันธุกรรมไม่สอดคล้องกับระยะห่างทางภาษา [correlation test ($r = 0.1735$, $P > 0.01$) และ partial correlation test ($r = 0.0008$, $P > 0.01$)] นอกจากนั้นยังพบว่า ระยะห่างทางภาษา และระยะห่างทางภูมิศาสตร์ไม่สอดคล้องกัน [(correlation ($r = 0.3667$, $P > 0.01$) และ partial correlation ($r = 0.3281$, $P > 0.01$)]

ตารางที่ 7 การวิเคราะห์ AMOVA อักษรย่อของประชากรแสดงในตารางที่ 1

	จำนวน กลุ่ม	จำนวน ประชากร	ระดับความผันแปร (ร้อยละ)			F_{st}	F_{sc}	F_{ct}
			a	b	c			
ภูมิศาสตร์								
ทุกประชากร	1	10	90.11	9.89		0.09889		
SN	1	5	93.10	6.90		0.06902		
KR	1	5	92.10	7.90		0.07900		
SN/KR	2	10	88.235	7.081	4.684	0.11765	0.07429	0.04684
ภาษา								
ทุกประชากร	1	10	90.11	9.89		0.09889		
TT	1	5	92.18	7.82		0.07820		
MK	1	5	89.32	10.68		0.10681		
TT/MK	2	10	89.74	9.35	0.91	0.10260	0.09434	0.00913

ตัวหนาแสดงระดับนัยสำคัญทางสถิติ ที่ $P < 0.01$

a = ภายในแต่ละประชากร; b = ระหว่างประชากรภายในกลุ่มเดียวกัน; c = ระหว่างกลุ่มของประชากร

F_{st} = Fixation index among populations and groups

F_{sc} = Fixation index among populations but within groups

F_{ct} = Fixation index among groups

AA = Austro-Asiatic linguistic family; TT = Tai-Kadai linguistic family; KR= Khorat Basin; SN = Sakon Nakorn Basin

6. การคัดเลือกโมเดล

ตารางที่ 8 แสดงค่า posterior probability ของแต่ละโมเดลซึ่งแสดงความสัมพันธ์ทางวิวัฒนาการของประชากรที่ศึกษา (ภาพที่ 2) โดยโมเดล 2 ซึ่งเป็นโมเดลที่อธิบายถึงความผันแปรทางพันธุกรรมของประชากรที่ศึกษาได้รับอิทธิพลจากภูมิศาสตร์ มีค่า posterior probability สูงสุด (มากกว่าร้อยละ 87) เมื่อคำนวณด้วยวิธี AR และ LR ในทุกระดับ threshold ซึ่งแสดงถึงการสนับสนุนโมเดลที่ 2

ตารางที่ 8 ค่า posterior probabilities ในแต่ละโมเดล จากการคำนวณด้วยวิธี acceptance-rejection procedure (AR) และ weighted multinomial logistic regression (LR)

Threshold	Model 1	Model 2	Model 3
AR			
100	0.090	0.910	0.000
200	0.070	0.910	0.020
300	0.077	0.907	0.017
500	0.078	0.904	0.018
LR			
25000	0.009	0.873	0.118
50000	0.006	0.870	0.124
75000	0.005	0.883	0.112
100,000	0.004	0.906	0.090

ผลการคำนวณ type I error เพื่อประเมินว่าข้อมูลและวิธีการที่น่าคัดเลือกโมเดลน่าเชื่อถือได้หรือไม่ เนื่องจากโมเดลทั้ง 3 แบบ ค่อนข้างคล้ายกันและข้อมูลพันธุศาสตร์มีเพียงดีเอ็นเอไมโทคอนเดรีย ซึ่งมีการถ่ายทอดในรูปแบบแฮปโลไทป์จึงถูกพิจารณาเป็นเพียงเครื่องหมายทางพันธุกรรม 1 ตำแหน่ง โดยจะคำนวณ type I error ทั้งวิธี AR และ LR และจะใช้ระดับของ probability threshold ตั้งแต่ 0.5 ถึง 1 ผลการศึกษา (ตารางที่ 8) จากการทำ simulation จำนวน 50,000 ครั้ง พบว่าค่า probability of recognize the right model ของวิธี LR สูงกว่า AR โดยเฉพาะเมื่อเพิ่มระดับของ probability threshold ให้สูงขึ้น แต่อย่างไรก็ตามทั้ง 2 วิธี (LR และ AR) จะมีค่า probability of recognize the right model ระดับสูง (มากกว่า 0.6) เมื่อค่า decision probability threshold เท่ากับ 0.5 และเมื่อโมเดลที่ถูกต้องไม่ถูกเลือก (not assigned) ค่า probability of recognize the right model มีค่าสูงมากโดยเฉพาะเมื่อระดับ decision probability threshold มีค่าสูง (0.9) ซึ่งแสดงถึงประสิทธิภาพของวิธี ABC ที่ใช้ว่าสามารถเลือกโมเดลได้ถูกต้อง

ตารางที่ 8 ผลการคำนวณ type I error สำหรับโมเดลทางวิวัฒนาการที่ได้จากวิธี ABC

AR					LR				
probability threshold	probability of recognize the right model				probability threshold	probability of recognize the right model			
	Model					Model			
	1	Model	Model	Not		Model	Model	Model	Not
	(true)	2	3	Assigned		1 (true)	2	3	Assigned
>0.5	0.49	0.1	0.01	0.4	>0.5	0.59	0.12	0.06	0.23
>0.6	0.4	0.02	0.01	0.57	>0.6	0.54	0.07	0.03	0.36
>0.7	0.33	0	0	0.67	>0.7	0.45	0.05	0.01	0.49
>0.8	0.2	0	0	0.8	>0.8	0.35	0.03	0	0.62
>0.9	0.08	0	0	0.92	>0.9	0.19	0	0	0.81
	Model					Model			
	Model	2	Model	Not		Model	2	Model	Not
	1	(true)	3	Assigned		1	(true)	3	Assigned
>0.5	0.08	0.45	0.06	0.41	>0.5	0.07	0.61	0.18	0.14
>0.6	0.03	0.33	0.01	0.63	>0.6	0.04	0.5	0.07	0.39
>0.7	0.01	0.26	0	0.73	>0.7	0.01	0.41	0.03	0.55
>0.8	0	0.11	0	0.89	>0.8	0	0.32	0.02	0.66
>0.9	0	0.05	0	0.95	>0.9	0	0.17	0	0.83
	Model					Model			
	Model	Model	3	Not		Model	Model	3	Not
	1	2	(true)	Assigned		1	2	(true)	Assigned
>0.5	0.02	0.08	0.59	0.31	>0.5	0.04	0.09	0.7	0.17
>0.6	0	0.05	0.37	0.58	>0.6	0.02	0.06	0.61	0.31
>0.7	0	0.02	0.16	0.82	>0.7	0	0.02	0.49	0.49
>0.8	0	0.01	0.04	0.95	>0.8	0	0.01	0.38	0.61
>0.9	0	0	0	1	>0.9	0	0	0.23	0.77

สรุปและวิจารณ์ผลการทดลอง

จากข้อมูลลำดับเบสของดีเอ็นเอไมโทคอนเดรียบริเวณ HVR1 ของการศึกษารั้วนี้ทำให้สามารถศึกษาความผันแปรทางพันธุกรรมในระดับภูมิภาค (micro-geographic level) และปัจจัยที่อาจส่งผลต่อความผันแปรทางพันธุกรรม ในกลุ่มชาติพันธุ์ที่หลากหลายของภาคตะวันออกเฉียงเหนือ โดยปัจจัยทางด้านภูมิศาสตร์ และ/หรือ ภาษา ที่อาจส่งผลต่อความผันแปรทางพันธุกรรมของประชากรในระดับภูมิภาค ทวีป และทั่วโลก ยังคงเป็นคำถามที่ท้าทายต่อนัก molecular anthropologist และ นักพันธุศาสตร์มนุษย์ (Helgason *et al.*, 2004; Relethford, 2004; Jay *et al.*, 2013) ซึ่งโจทย์ปัญหาดังกล่าวยังไม่มีรายงานการศึกษาในภาคตะวันออกเฉียงเหนือ ผลการวิเคราะห์ Mantel test AMOVA SAMOVA และ ABC ซึ่งมีความสอดคล้องกันแสดงว่าปัจจัยด้านภูมิศาสตร์มีอิทธิพลต่อโครงสร้างและความสัมพันธ์ระหว่างประชากรที่ศึกษาตามโมเดล isolation by distance (IBD) ซึ่งกล่าวว่าประชากรที่อาศัยอยู่ในพื้นที่เดียวและลักษณะภูมิศาสตร์ที่คล้ายกันจะมีพันธุกรรมที่เหมือนกันมากกว่าประชากรที่อาศัยในพื้นที่ไกลออกไป (Wright, 1943; Slatkin, 1993) โดยลักษณะภูมิศาสตร์ของภาคตะวันออกเฉียงเหนือของประเทศไทย สามารถแบ่งเป็น 2 ส่วน คือ แอ่งที่ราบโคราช (KR) ซึ่งเป็นบริเวณที่ประชากรที่พูดภาษาตระกูลออสโตรเอเชียติก (AA) อาศัยอยู่เป็นจำนวนมาก ในขณะที่บริเวณแอ่งที่ราบสกลนคร (SN) จะมีกลุ่มประชากรที่พูดภาษาตระกูลไท-กะได (TK) ซึ่งอพยพมาจากประเทศลาวและเวียดนาม อาศัยอยู่เป็นจำนวนมาก

ในแอ่งที่ราบสกลนคร ประชากรที่ศึกษาประกอบด้วย ชาวโซ่ (SOA) ภูไท (PUT) ไทแสก (SAK) กะเลิง (KAL) และไทญ้อ (YOH) ซึ่ง SOA เป็นประชากรเดียวในแอ่งที่ราบสกลนครที่มีภาษาพูดจัดอยู่ในกลุ่ม AA ในขณะที่ประชากรที่เหลือพูดภาษา TT ชาวโซ่มีถิ่นฐานเดิมอยู่ในบริเวณแขวงคำม่วน ประเทศลาว ในปี ค.ศ. 1844 ชาวโซ่บางส่วนได้อพยพ เข้าสู่ดินแดนของประเทศไทย อาศัยอยู่ในเขต อ. กุสุมาลย์ จากผลการวิเคราะห์ ABC (ภาพที่ 2) โมเดลทางวิวัฒนาการที่เสนอ พบว่าชาวโซ่มีโครงสร้างทางพันธุกรรมที่คล้ายกับประชากรอื่นในแอ่งที่ราบสกลนคร ซึ่งสาเหตุดังกล่าวอาจเกิดจากการผสมผสานทางพันธุกรรมระหว่างชาวโซ่และประชากรข้างเคียงหลังจากการอพยพเข้ามาอาศัยในประเทศไทย เมื่อประมาณ 200 ปีที่ผ่านมา นอกจากนั้นการที่ SOA กับ PUT และ SOA กับ YOH มีแฮปโลไทป์ที่เหมือนกัน ถึง 6 และ 5 แฮปโลไทป์ ตามลำดับ (ตารางที่ 2) จะสนับสนุนโครงสร้างทางพันธุกรรมที่คล้ายกันของ SOA และประชากรข้างเคียง

ในขณะที่ประชากรในแอ่งที่ราบสกลนครส่วนใหญ่มีโครงสร้างทางพันธุกรรมที่คล้ายกัน ยกเว้น SAK ซึ่งมีโครงสร้างทางพันธุกรรมที่แตกต่างออกไปอย่างมาก ซึ่งสนับสนุนจากผลการวิเคราะห์

pairwise F_{st} (ตารางที่ 3) MPD (ตารางที่ 4) และ SAMOVA (ตารางที่ 5) จากหลักฐานตามประวัติศาสตร์ ชาวไทแสกอพยพมาจากประเทศเวียดนาม เข้าสู่ประเทศลาว ในบริเวณแขวงคำม่วน เมื่อประมาณ 380 ปีที่ผ่านมา จากนั้น SAK ได้อพยพข้ามแม่น้ำโขงเข้าสู่เขตจังหวัดนครพนม ประเทศไทย เมื่อประมาณ 200 ปีที่ผ่านมา โครงสร้างทางพันธุกรรมที่แตกต่างของ SAK อาจเกิดเนื่องจากอิทธิพลของปรากฏการณ์คอขวด (bottleneck effect) (Davis *et al.*, 2011) ซึ่งเป็นเจเนติกดริฟท์ชนิดหนึ่ง โดยประชากรนี้อาจมีการลดจำนวนเพศหญิงลงอย่างฉับพลันในระหว่างการอพยพ จึงทำให้ประชากรสะสมความแตกต่างและความหลากหลายทางพันธุกรรมลดลง (มีค่าความหลากหลายของแฮปโลไทป์และ intra-MPD ต่ำที่สุด) นอกจากนั้นข้อถกเถียงสำหรับการจัดกลุ่มภาษาของ SAK โดยมีนักภาษาศาสตร์กลุ่มเก่าเชื่อว่าภาษาพูดของชาว SAK เป็นภาษา AA ตระกูลย่อยมอญ-เขมร ในขณะที่นักภาษาศาสตร์กลุ่มใหม่จัดภาษาแสดให้อยู่ในตระกูล TK ตระกูลย่อยไทเหนือ (Smalley, 1994; Schliesinger, 2000) ผลจากการวิเคราะห์ดีเอ็นเอไมโทคอนเดรีย พบว่า SAK มีความสัมพันธ์ทางเชื้อสายใกล้เคียงกับชาว SOA (พูด AA) มากที่สุด จากการศึกษาถึงความสัมพันธ์ระหว่างภาษาและพันธุกรรมจำนวนมาก ทำให้สามารถอนุมานว่าประชากรที่มีภาษาพูดใกล้เคียงกันจะมีโครงสร้างทางพันธุกรรมที่คล้ายกัน (Barbujani and Sokal, 1990; Cavalli-Sforza *et al.*, 1992; Barbujani and Pilastro, 1993; Boattini *et al.*, 2011) ดังนั้นจากการศึกษาครั้งนี้พบว่าการจัดกลุ่มภาษาของชาวแสดไม่สอดคล้องกับโครงสร้างทางพันธุกรรม อย่างไรก็ตามภาพรวมของโครงสร้างทางพันธุกรรมของ SAK อาจได้รับอิทธิพลจากปัจจัยอื่นที่มีอิทธิพลมากกว่า เช่น ภูมิศาสตร์

ในแอ่งที่ราบโคราช ประชากรที่ศึกษามีทั้งหมด 5 ประชากร คือเขมร (KHM) ชาวนน (BON) ส่วย (SUY) มอญ (MON) ซึ่งเป็นประชากรที่พูดภาษาตระกูลมอญ-เขมร (AA) และชาวอีสาน (LAO) ซึ่งพูดภาษาตระกูลไท-กะได ชาวลาวอีสานหมายถึงกลุ่มคนที่มีเชื้อชาติลาว แต่มีสัญชาติไทย โดยชาวอีสานเป็นกลุ่มคนหลักที่อาศัยอยู่ในภาคตะวันออกเฉียงเหนือของประเทศไทย ตามหลักฐานทางประวัติศาสตร์ ชาวอีสานอพยพมาจากดินแดนของประเทศลาวในปัจจุบัน ในช่วงระหว่าง ค.ศ. 1827 ถึง 1870 แม้ว่าหมู่บ้านชาวลาวจะตั้งอยู่ในแอ่งที่ราบโคราช แต่ประชากรดังกล่าวกลับมีความสัมพันธ์ทางเชื้อสายใกล้ชิดกับประชากรในแอ่งที่ราบสกลนคร รูปแบบความสัมพันธ์ดังกล่าวอาจเกิดมาจากการมีต้นกำเนิดเดียวกัน และหลังจากที่อพยพเข้ามาสู่ประเทศไทย ชาวอีสานและประชากรไท-กะไดในแอ่งที่ราบสกลนครยังคงรักษาพันธุกรรมไว้เหมือนเดิม การศึกษาครั้งนี้ยังขาดประชากรชาวลาวอีสานที่อยู่ในแอ่งที่ราบสกลนคร ดังนั้นการศึกษาพันธุกรรมของประชากร LAO ในแอ่งที่ราบสกลนครจะทำให้ทราบพลวัตของประชากรชาว LAO ได้ชัดเจนขึ้น

เป็นที่น่าสนใจว่า LAO และ KHM มีโครงสร้างทางพันธุกรรมที่ไม่แตกต่างกัน แสดงจากค่า pairwise F_{st} ที่แตกต่างกันอย่างไม่มีนัยสำคัญ (ตารางที่ 2) ผลการศึกษาดังกล่าวสอดคล้องกับงานวิจัย

ก่อนหน้านี้ ด้านพันธุศาสตร์ (Lertrit *et al.*, 2008) และสังคมวิทยา (Smalley, 1998; Khanittanan, 2001; Talbot and Janthed, 2002) แม้ว่าผลการศึกษาครั้งนี้จะพบอิทธิพลของภูมิศาสตร์ที่มีต่อโครงสร้างทางพันธุกรรมของประชากร แต่สำหรับในพื้นที่แอ่งที่ราบโคราช ประชากรที่ศึกษามีความแตกต่างทางพันธุกรรมในระดับหนึ่ง ซึ่งอาจอธิบายได้โดยการที่ภูมิศาสตร์ได้กำหนดให้ประชากรในแอ่งที่ราบโคราชมีพันธุกรรมที่เหมือนกัน แต่เมื่อระยะเวลาผ่านไปปัจจัยอื่นๆ อาจส่งผลให้ประชากรเหล่านั้นสะสมความแตกต่างของพันธุกรรม เช่น ปัจจัยทางด้านวัฒนธรรม ภาษา และแรงผลักดันทางวิวัฒนาการเช่น การเกิดเจเนติกดริฟท์ การผสมเลือดชิด และการผสมผสานทางพันธุกรรม

ตัวอย่างของประชากรที่อาจเกิดการผสมเลือดชิด คือชาวบน (BON) หรือ ฮัญญู นักภาษาศาสตร์เชื่อว่าชาวบนเป็นกลุ่มชนดั้งเดิมที่อาศัยอยู่ในพื้นที่แห่งนี้ก่อนที่ชาวเขมรและกลุ่มคนที่พูดภาษาตระกูลไท-กะไดจะครอบครองดินแดนแห่งนี้ ชาวบนมีถิ่นที่อยู่อาศัยในประเทศไทยเท่านั้น ในเขต จ.ชัยภูมิ จ.เพชรบูรณ์ และ จ.นครราชสีมา โดยตัวอย่างของประชากรที่ศึกษาเป็นชาวบนจากบ้านหวังอ้ายโพธิ์ อ.เทพสถิต จ.ชัยภูมิ ซึ่งเป็นหมู่บ้านชาวบนที่ยังคงมีภาษาและวัฒนธรรมที่เป็นเอกลักษณ์ (Premsrirat, 2002; Prasert, 2009) ผลการศึกษาของพารามิเตอร์ของความหลากหลายทางพันธุกรรมที่มีค่าต่ำ เช่น h S และ intra-MPD (ตารางที่ 1) ซึ่งอาจเกิดจากสาเหตุการผสมเลือดชิดเนื่องจากประเพณีนิยมการแต่งงานภายในกลุ่ม อย่างไรก็ตาม ในปัจจุบันชาวบนมีการผสมผสานกับชาวลาวอีสานมากยิ่งขึ้น ก็อาจทำให้การมีโครงสร้างทางพันธุกรรมที่จำเพาะนี้จางหายไป การศึกษาครั้งนี้จึงเป็นการศึกษาด้านพันธุศาสตร์ครั้งแรกที่รายงานความหลากหลายทางพันธุกรรมที่ต่ำในชาวบนจากบ้านหวังอ้ายโพธิ์ นอกจากนั้นนักภาษาศาสตร์ระบุว่าภาษาพูดของชาวบนเป็นภาษาเดียวกับภาษามอญโบราณ ตั้งแต่สมัยทวารวดี ดังนั้นถ้าใช้หลักว่าประชากรที่มีภาษาพูดใกล้เคียงกันจะมีโครงสร้างทางพันธุกรรมที่คล้ายกัน ก็คาดว่า BON และ MON จะมีความสัมพันธ์ทางเชื้อสายใกล้ชิดกัน แต่ผลการศึกษาครั้งนี้กลับตรงกันข้าม กล่าวคือ จะพบว่า BON และ MON มีระยะห่างทางพันธุกรรมที่ห่างกันจึงสามารถสรุปได้ว่าชาวมอญ จาก จ.นครราชสีมา และชาวบนจาก จ.ชัยภูมิไม่มีความสัมพันธ์ทางเชื้อสายกัน

ชาวมอญ (MON) เป็นกลุ่มประชากรหนึ่งที่มีความเก่าแก่ที่สุดในประวัติศาสตร์ของภูมิภาคเอเชียตะวันออกเฉียงใต้ ซึ่งต้นกำเนิดของชาวมอญยังไม่สามารถระบุได้ชัดเจน เชื่อว่าในสมัยก่อนชาวมอญอาศัยอยู่ในบริเวณพื้นที่ประเทศจีนตอนใต้ในปัจจุบัน จากนั้นได้อพยพมาครอบครองดินแดนตอนบนของประเทศพม่าในปัจจุบัน ในช่วงต้นของคริสต์ศักราช จากนั้นด้วยเหตุผลทางการเมืองได้ถูกขับไล่ ชาวมอญส่วนหนึ่งอพยพลงมาทางตอนใต้และอาศัยอยู่บริเวณเมืองพะโค หรือหงสาวดี (Pegu) และเมืองสะเทิม (Thaton) ของประเทศพม่า และชาวมอญอีกส่วนอพยพมาทางทิศตะวันออก และครอบครองดินแดนทางภาคกลางและภาคใต้ของประเทศไทยในปัจจุบัน อาณาจักรทวารวดีของชาว

มอญ ถูกสร้างขึ้นประมาณศตวรรษที่ 3 และรุ่งเรืองจนถึงศตวรรษที่ 10 ในบริเวณตอนกลางของประเทศ ไทย (Schliesinger, 2000) อาณาจักรทวารวดีได้แพร่ขยายไปทั่วทุกสารทิศ ทั้งภาคเหนือและภาคใต้ ของประเทศไทยในปัจจุบัน จนกระทั่งในปี ค.ศ. 1775 ชาวมอญกลุ่มหนึ่งได้อพยพมาจากประเทศพม่า เข้ามาอาศัยอยู่ในบริเวณจ.นครราชสีมา จนกระทั่ง ในปี ค.ศ. 1793 จ.นครราชสีมา มีชาวมอญอาศัยอยู่ ประมาณ 2,500 คน จากประวัติศาสตร์การอพยพ MON ที่ศึกษาครั้งนั้นมาจากประเทศพม่า ดังนั้นการ ไม่พบความสัมพันธ์ทางเชื้อสายระหว่าง MON และ BON จึงน่าจะมีสาเหตุมาจากชาวมอญในพม่าและ ชาวมอญในอาณาจักรทวารวดี มีพันธุกรรมที่แตกต่างกันมาตั้งแต่อดีต สิ่งที่น่าสนใจคือผลการศึกษา ทางด้านพันธุศาสตร์ครั้งนี้นักพบว่าชาว MON มีค่า π intra-MPD และจำนวนของแฮปโลไทป์เฉพาะที่ ต่ำ (ตารางที่ 1) และยังแสดงการมีการเพิ่มขยายขนาดของประชากร (ตารางที่ 1 และภาพที่ 4) ดังนั้น ประชากร MON จึงน่าจะเกิดปรากฏการณ์คอขวด (bottleneck) และหลังจากนั้นมีการเพิ่มขยายขนาด ของประชากรอย่างรวดเร็ว

ชาวส่วย (SUY) เป็นกลุ่มชาติพันธุ์ที่พูดภาษาออสโตรเอเชียติก กลุ่มย่อยอมอญ-เขมรอีกกลุ่ม หนึ่ง ในบริเวณภาคตะวันออกเฉียงเหนือตอนใต้ ผลการศึกษาพบว่า SUY มีความใกล้ชิดทางพันธุกรรม กับ KHM และ BON (ตารางที่ 3 และ ภาพที่ 5) สอดคล้องกับหลักฐานทางประวัติศาสตร์ ภาษา และ สังคม ที่ระบุความสัมพันธ์ที่ใกล้ชิดกันระหว่าง SUY และ KHM ชาวส่วยหรือกวย (Kui) หรือ ชาวเขมร เรียกว่าเขมร์-บอเรน (แปลว่าเขมรโบราณ) เป็นประชากรดั้งเดิมที่อาศัยอยู่ในพื้นที่ของประเทศไทย ลาว และกัมพูชาในปัจจุบัน ก่อนที่ชาวเขมรและไท-กะไดจะครอบครองดินแดนแห่งนี้ ชาวส่วยที่ศึกษา อาศัยอยู่ในจังหวัดสุรินทร์ซึ่งอพยพมาจากประเทศลาวจำนวน 2 ชั่ว คือ ระหว่างปี ค.ศ. 1656-1688 และ การอพยพอีกจำนวนมาก ในปี ค.ศ. 1760 ในปัจจุบันชาวส่วยที่อาศัยอยู่ในภาคตะวันออกเฉียง เหนือของประเทศไทย ได้รับเอาภาษาพูดของชาวลาวอีสานและเขมร โดยชาวส่วยที่สามารถพูดภาษา ลาวได้ เรียกว่า ลาวส่วย และส่วยที่สามารถพูดภาษาเขมรได้ เรียก เขมรส่วย ดังนั้นการที่ชาวส่วยมี โครงสร้างทางพันธุกรรมใกล้กับชาวเขมรจึงน่าจะเกิดมาจากการผสมผสานระหว่างกลุ่มชาติพันธุ์ อย่างไม่กี่ก็ตามได้มีนักมานุษยวิทยา เสนอถึงความสัมพันธ์ระหว่างกลุ่มชาติพันธุ์ส่วยและชาวน (Sa- ard, 1984) ซึ่งผลการศึกษาครั้งนี้ก็สนับสนุนสมมติฐานดังกล่าว

งานวิจัยนี้ทำการศึกษาโครงสร้างและความผันแปรทางพันธุกรรมของหลายกลุ่มชาติพันธุ์ที่ อาศัยอยู่ในภาคตะวันออกเฉียงเหนือของประเทศไทย ผลการวิเคราะห์ทางพันธุศาสตร์พบว่าปัจจัย ทางด้านภูมิศาสตร์มีอิทธิพลต่อความผันแปรทางพันธุกรรมของประชากรที่ศึกษา โดยประชากรที่อาศัย อยู่ในแอ่งที่ราบเดียวกันจะมีความสัมพันธ์ทางเชื้อสายใกล้ชิดกัน โดยเฉพาะประชากรที่อาศัยอยู่ในแอ่งที่ รามสกลนคร อย่างไรก็ดีตามภายในแอ่งที่ราบเดียวกัน โครงสร้างทางพันธุกรรมของประชากรได้ถูก

กำหนดโดยปัจจัยทางด้านภาษาและวัฒนธรรม รวมทั้งแรงผลักดันทางวิวัฒนาการ เช่น เจเนติกดริฟท์ การผสมเลือดชิด และการผสมผสานทางพันธุกรรม

ข้อเสนอแนะสำหรับงานวิจัยในอนาคต

การศึกษาครั้งนี้อาศัยเพียงเครื่องหมายทางพันธุกรรมจากดีเอ็นเอไมโทคอนเดรีย ซึ่งมีการถ่ายทอดผ่านทางฝ่ายหญิงเท่านั้น ในอนาคตผลการศึกษาจากเครื่องหมายทางพันธุกรรมชนิดอื่น เช่น โครโมโซมวาย และดีเอ็นเอบนออโตโซมจะช่วยให้การศึกษาพันธุศาสตร์ประชากรของมนุษย์ในประชากรภาคตะวันออกเฉียงเหนือของประเทศไทยมีความสมบูรณ์มากยิ่งขึ้น

เอกสารอ้างอิง

- Alfonso-Sánchez, M.A., Cardoso, S., Martínez-Bouzas, C., Peña, J.A., Herrera, R.J., Castro, A. *et al.* Mitochondrial DNA haplogroup diversity in Basques: a reassessment based on HVI and HVII polymorphisms. *Am. J. Hum. Biol.* **20**(2), 154-164 (2008).
- Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M. & Howell, N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147 (1999).
- Barbujani, G. & Sokal, R.R. Zones of sharp genetic change in Europe are also linguistic boundaries. *Proc. Natl. Acad. Sci. USA.* **87**, 1816-1819 (1990).
- Barbujani, G. & Pilastro, A. Genetic evidence on origin and dispersal of human populations speaking languages of the Nostratic Macro family. *Proc. Natl. Acad. Sci. USA.* **90**(10), 4670-4673 (1993).
- Barbujani, G. Geographic patterns: how to identify them and why. *Hum. Biol.* **72**, 133-153 (2000).
- Beaumont, M. Joint determination of topology, divergence time and immigration. in *Simulation, Genetics, and Human Prehistory* (eds Matsumura, S., Forster, P. & Renfrew, C.) 135-154 (McDonald Institute for Archaeological Research, Cambridge, England, 2008)
- Bertorelle, G., Benazzo, A. and Mona, S. ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Mol. Ecol.* **19**(13), 2609-2625 (2010).

- Besaggio, D., Fuselli, S., Srikummool, M., Kampuansai, J., Castri, L., Tyler-Smith, C. *et al.* Genetic variation in Northern Thailand Hill Tribes: origins and relationships with social structure and linguistic differences. *BMC Evol. Biol.* **7 (Suppl 2)**, S12 (2007).
- Boattini, A, Griso, C. & Pettener, D. Linguistic versus genetic isolation. The strange case of the Walser from Upper Lys Valley (Italian Western Alps). *J. Anthropol. Sci.* **89**, 161-175 (2011).
- Boonsoda, P., Srithawong, S., Srikuka, S., Kutanan, W. 2013. Mitochondrial DNA variation of the Khmer in Surin Province, Thailand. *Thai J. Genet.* 6(1): 40-48 (in Thai).
- Budowle, B., Allard, M.W., Wilson, M.R. & Chakraborty, R. Forensics and mitochondrial DNA: applications, debates, and foundations. *Annu. Rev. Genomics. Hum. Genet.* **4**, 119-141 (2003).
- Cavalli-Sforza, L.L., Minch, E. & Mountain, J.L. Coevolution of genes and languages revisited. *Proc. Natl. Acad. Sci. USA.* **89**, 5620-5624 (1992).
- Cavalli-Sforza, L.L., Menozzi, P. & Piazza A. *The history and geography of human genes* (Princeton University Press, Princeton, USA, 1994).
- Coia, V., Boschi, I., Trombetta, F., Cavulli, F., Montinaro, F., Destro-Bisol, G. *et al.* Evidence of high genetic variation among linguistically diverse populations on a micro-geographic scale: a case study of the Italian Alps. *J. Hum. Genet.* **57(4)**, 254-260 (2012).
- Cavalli-Sforza L.L. and Feldman M.W. 2003. The application of molecular genetic approaches to the study of human evolution. *Nat. Genet.* 33(Suppl.): 266-275.
- Davis, M.C., Novak, S.J. & Hampikian, G. Mitochondrial DNA analysis of an immigrant Basque population: Loss of diversity due to founder effects. *Am. J. Phys. Anthropol.* **144(4)**, 516-525 (2011).
- Dupanloup, I., Schneider, S. & Excoffier, L. A simulated annealing approach to define the genetic structure of populations. *Mol. Eco.* **11**, 2571-2581 (2002).
- Eller, E. Population substructure and isolation by distance in three continent regions. *Am. J. Phys. Anthropol.* **108**, 147-159 (1999).
- Excoffier, L., Smouse, P. & Watro, J. Analysis of molecular variance inferred from metric distance among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics.* **131**, 479-491 (1992).

- Excoffier, L. & Lischer, H.E.L. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Eco. Res.* **10**, 564-567 (2010).
- Fucharoen, G., Fucharoen, S. & Horai, S. Mitochondrial DNA polymorphisms in Thailand. *J. Hum. Genet.* **46**, 115-125 (2001).
- Fu, Y.X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics.* **147**, 915-925 (1997).
- Greenberg, B.D., Newbold J.E. and Sugino A. 1983. Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. *Gene*, 21, 33-49.
- Harpending, H.C. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum. Biol.* **66**, 591-600 (1994).
- Helgason, A., Yngvado'ttir, B., Hrafnkelsson, B., Gulcher, J. & Stefa'nnsson, K. An Icelandic example of the impact of population structure on association studies. *Nat. Genet.* **37**, 90-95 (2004).
- Jay, F., Sjödin, P., Jakobsson, M. & Blum, M.G. Anisotropic isolation by distance: the main orientations of human genetic differentiation. *Mol. Biol. Evol.* **30(3)**, 513-525 (2013).
- Khanittanan, W. Khmero-Thai: the great change in the history of Thai Language in the Chao Praya basin. *J. Language and Linguistics.* **19(2)**, 35-50 (2001).
- Kutanan, W., Kampuansai, J., Nakbunlung, S., Lertvicha, P., Seielstad, M., Bertorelle, G. *et al.* Genetic structure of KhonMueang populations along a historical Yuan migration route in Northern Thailand. *Chiang Mai J. Science.* **38(2)**, 295-305 (2011a).
- Kutanan, W., Kampuansai, J., Fuselli, S., Nakbunlung, S., Seielstad, M., Bertorelle, G. *et al.* Genetic structure of the Mon-Khmer speaking groups and their affinity to the neighbouring Tai populations in Northern Thailand. *BMC Genet.* **12**: 56 (2011b).
- Kutanan, W., Srithawong, S., Kamlao, A. & Kampuansai, J. Mitochondrial DNA-HVR1 Variation Reveals Genetic Heterogeneity in Thai-Isan Peoples from the Lower Region of Northeastern Thailand. *Adv. Anthropol.* **4(1)**, 7-12 (2014).
- Lertrit, P., Poolsuwan, S., Thosarat, R., Sanpachudayan, T., Boonyarit, H., Chinpaisal, C. *et al.* Genetic history of Southeast Asian populations as revealed by ancient and modern human mitochondrial DNA analysis. *Am. J. Phys. Anthropol.* **137**, 425-440 (2008).

- Lewis, M.P. *Ethnologue: Languages of the World* 16th edn. (SIL International, Dallas, Texas, USA, 2009) Online version: <http://www.ethnologue.com/>.
- Librado, P. & Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. **25**, 1451-1452 (2009).
- Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res* 27: 209-220.
- Malyarchuk, B.A., Perkova, M.A., Derenko, M.V., Vanecek, T., Lazur, J. & Gomolcak, P. Mitochondrial DNA Variability in Slovaks, with Application to the Roma Origin. *Ann. Hum. Genet.* **72**, 228-240 (2008).
- Nei, M. *Molecular Evolutionary Genetics* (Columbia University Press, New York, USA, 1987)
- Pakendorf, B. & Stoneking, M. Mitochondrial DNA and human evolution. *Ann. Rev. Genomics. Hum. Genet.* **6**, 165-183 (2005).
- Pardiñas, A.F., Roca, A., García-Vazquez, E. & López, B. Assessing the Genetic Influence of Ancient Sociopolitical Structure: Micro-differentiation Patterns in the Population of Asturias (Northern Spain). *PLoS ONE*. **7(11)**, e50206 (2012).
- Prasert, S., Pansila, V. & Lasunon, O. Guidelines and Methods for Conservation, Revitalization and Development of the Traditions and Customs of NyahKur Ethnic Group for Tourism in the Province of Chaiyapum in Northeast Thailand. *The Social Sciences*. **4**, 174-179 (2009).
- Premisrat, S. The Future of Nyah Kur. in *Collected papers on Southeast Asian and Pacific languages* (eds Bauer, R.S.) 155-165 (The Australian University, Canberra, Australia, 2002).
- Pritchard, J.K., Seielstad, M.T., Perez-Lezaun, A. & Feldman, M.W. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol. Biol. Evol.* **16(12)**, 1791-1798 (1999).
- Rogers, A.R. & Harpending, H. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9(3)**, 552-569 (1992).
- Relethford, J.H. Global Pattern of Isolation by distance based on genetic and morphological data. *Hum. Biol.* **76(4)**, 499-513 (2004).

- Sa-ard, O. *Phrase to sentence in Kuay (Surin)* (Mahidol University, Nakorn Pathom, Thailand, 1984).
- Schliesinger, J. *Ethnic groups of Thailand: Non-Tai-speaking peoples* (White Lotus Press, Bangkok, Thailand, 2000)
- Schliesinger, J. *Tai Group of Thailand, Volume 1: Introduction and overview* (White Lotus Press, Bangkok, Thailand, 2001)
- Schurr, T.G., Sukernik, R.I., Starikovskaya, Y.B. & Wallace D.C. Mitochondrial DNA variation in Koryaks and Itel'men: population replacement in the Okhotsk sea-Bering sea region during the Neolithic. *Am. J. Phys. Anthropol.* **108**, 1-39 (1999).
- Slatkin, M. Isolation by distance in equilibrium and nonequilibrium populations. *Evolution.* **47**, 264-279 (1993).
- Smalley, W.A. *Linguistic Diversity and National Unity: Language Ecology in Thailand* (University of Chicago Press, Chicago, USA, 1994).
- Smalley, W. A. 1994. *Linguistic Diversity and National Unity: Language Ecology in Thailand*. University of Chicago Press, Chicago.
- Smalley, W.A. Multilingualism in the Northern Khmer population of Thailand. *Language. Sci.* **10(2)**, 395-408 (1988).
- Tajima, F. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics.* **123**, 585-595 (1989).
- Talbot, S. & Janthred, C. Northeast Thailand before Angkor: Evidence from an Archaeological Excavation at the Prasat Hin Phimai. *Asia Perspectives.* **40 (2)**, 179-194 (2002).
- Tishkoff, S.A. & Kid, K.K. Implications of biogeography of human populations for 'race' and medicine. *Nat. Genet.* **36**, S21-S27 (2004).
- Torrioni, A., Achilli, A., Macaulay, V., Richards, M. & Bandelt, H.J. Harvesting the fruit of the human mtDNA tree. *Trends Genet.* **22**, 339-345 (2006).
- Wegmann, D., Leuenberger, C., Neuenschwander, S. & Excoffier, L. ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC bioinformatics.* **11**, 116 (2010).
- Wongtaed, S. *Explore Isan Society and Cultures* (Art and culture Press, Bangkok, Thailand, 1999) (in Thai)
- Wright, S. Isolation by distance. *Genetics.* **28**, 114-138 (1943).

- Zerjal, T., Beckman, L., Beckman, G., Mikelsaar, A.V., Krumina, A., Kucinskas, V. *et al.* Geographical, Linguistic and Cultural Influences on Genetic Diversity: Y-Chromosomal Distribution in Northern European Populations. *Mol. Biol. Evol.* **8(6)**, 1077-1087 (2001).
- Zerjal, T., Xue, Y., Bertorelle, G., Wells, R.S., Bao, W., Zhu, S. *et al.* The genetic legacy of the Mongols. *Am. J. Hum. Genet.* **72(3)**, 717-721 (2003).
-