รายงานวิจัยฉบับสมบูรณ์

การวิเคราะห์ความหลากหลายทางพันธุกรรมของกลุ่มชาติพันธุ์ในภาคตะวันออกเฉียงเหนือของประเทศไทย

ผู้ช่วยศาสตราจารย์ ดร.วิภู กุตะนันท์

2 กรกฎาคม 2557

สัญญาเลขที่ MRG5580058

รายงานวิจัยฉบับสมบูรณ์

การวิเคราะห์ความหลากหลายทางพันธุกรรมของกลุ่มชาติพันธุ์ในภาคตะวันออกเฉียงเหนือของประเทศไทย

ผู้ช่วยศาสตราจารย์ ดร.วิภู กุตะนันท์

ภาควิชาชีววิทยา คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

สนับสนุนโดยสำนักงานกองทุนสนับสนุนการวิจัย

(ความเห็นในรายงานนี้เป็นของผู้วิจัย
สกว.ไม่จำเป็นต้องเห็นด้วยเสมอไป)

เอกสารแนบหมายเลข 2

บทคัดย่อ

รหัสโครงการ: MRG5580058

ชื่อโครงการ: การวิเคราะห์ความหลากหลายทางพันธุกรรมของกลุ่มชาติพันธุ์ในภาคตะวันออกเฉียงเหนือของประเทศไทย

ชื่อนักวิจัย : ผู้ช่วยศาสตราจารย์ วิภู กุตะนันท์ มหาวิทยาลัยขอนแก่น

E-mail Address : wibhu@kku.ac.th

ระยะเวลาโครงการ: 2 กรกฎาคม 2555 ถึง 2 กรกฎาคม 2557

ภาคตะวันออกเฉียงเหนือของประเทศไทยหรือภาคอีสาน มีอาณาเขตติดต่อกับประเทศลาวและกัมพูชา และมีระยะทางไม่ไกลจากประเทศเวียดนาม จึงทำให้ดินแดนแห่งนี้เป็นเส้นทางการอพยพของกลุ่มชาติพันธุ์ที่หลากหลายตั้งแต่อดีต ทั้งกลุ่มชาติพันธุ์ที่พูดภาษาตระกูลออสโตรเอเชียติค เช่น มอญ เขมร ส่วย ชาวบน และโส้ เป็นต้น และกลุ่มชาติพันธุ์ที่พูดภาษาตระกูลไท-กะได เช่น ผู้ไท ไทญ้อ ไทแสก กะเลิง และลาวอีสาน ภาคอีสานสามารถแบ่งเป็น 2 ส่วนตามลักษณะทางภูมิศาสตร์ คือแอ่งที่ราบสกลนครและแอ่งที่ราบโคราช งานวิจัยทางด้านพันธุศาสตร์ประชากรก่อนหน้านี้ ได้แสดงถึงอิทธิพลของปัจจัยด้านภูมิศาสตร์และภาษาพูดที่มีต่อความผันแปรทางพันธุกรรม แต่การศึกษาปัจจัยดังกล่าวยังไม่เคยมีรายงานในพื้นที่ภาคตะวันออกเฉียงเหนือของประเทศไทย ดังนั้นงานวิจัยนี้จึงมีวัตถุประสงค์ที่จะประเมินปัจจัยที่อาจส่งผลต่อโครงสร้างและความผันแปรของดีเอ็นเอไมโทคอนเดรียในประชากรจำนวน 10 กลุ่มชาติพันธุ์ จำนวน 433 คน โดยใช้ลำดับเบสดีเอ็นเอไมโทคอนเดรียบริเวณที่มีความหลากหลายสูง ความยาว 596 คู่เบส เป็นเครื่องหมายทางพันธุกรรม ผลการศึกษาจากการวิเคราะห์แผนภูมิแสดงความสัมพันธ์ทางพันธุกรรมแบบ multidimensional scaling และการวิเคราะห์ spatial analysis of molecular variance พบความสัมพันธ์ทางเชื้อสายที่ใกล้ชิดกันของประชากรที่อาศัยอยู่ในแอ่งที่ราบสกลนคร ในขณะที่ผลการวิเคราะห์ analysis of molecular variance และ Mantel test แสดงอิทธิพลของปัจจัยด้านภูมิศาสตร์ที่ส่งผลต่อโครงสร้างทางพันธุกรรมของประชากร ทำการสร้างโมเดลแสดงความสัมพันธ์ทางวิวัฒนาการ จำนวน 3 โมเดล ซึ่งได้รับอิทธิพลของภาษา (โมเดล 1) ภูมิศาสตร์ (โมเดล 2) และ การอพยพ (โมเดล 3) จากนั้นใช้วิธีการคำนวณแบบ Approximate Bayesian Computation และ type I error เพื่อคัดเลือกโมเดลที่เหมาะสม ผลการศึกษาคือโมเดลที่ 2 เหมาะสมกับข้อมูลดีเอ็นเอไมโทคอนเดรียในประชากรที่ศึกษา ซึ่งแสดงถึงปัจจัยทางภูมิศาสตร์ที่มีผลต่อความผันแปรของดีเอ็นเอไมโทคอนเดรียในประชากรภาคตะวันออกเฉียงเหนือของประเทศไทย

คำหลัก: ดีเอ็นเอไมโทคอนเดรีย/ แอ่งที่ราบสกลนคร/ แอ่งที่ราบโคราช/ ออสโตรเอเชียติค/ ไท-กะได/ ภาคตะวันออกเฉียงเหนือของประเทศไทย/ Approximate Bayesian Computation

Abstract

Project Code : MRG5580058

Project Title : Deciphering diversity in various ethnic affiliations in Northeastern Thailand

Investigator : Assist.Prof.Dr. Wibhu Kutanan

E-mail Address : wibhu@kku.ac.th

Project Period : 2 July 2013- 2 July 2015

Northeastern Thailand or Isan shares borders with Laos and Cambodia and lies in close proximity to Vietnam, this region has become a crossroad of various Southeast Asian peoples through migration and settlement periods since prehistoric times. Several studies have shown the influence of geographic and linguistic factors in shaping genetic variation. Geographic barriers separate Northeastern Thailand into two wide basins, the Sakon Nakorn Basin and the Korat Basin serving today as home to diverse ethnicities encompassing two different linguistic families, i.e., the Austro-Asiatic; Suay (Kui), Mon, Chaobon (Nyahkur), So and Khmer, and the Tai-Kadai; Saek, Nyaw, Phu Tai, Kaleung and Lao Isan. The present study intends to evaluate the elements responsible for maternal genetic variations, like geography and language, of these ten Northeastern Thai ethnicities. Population history is also reconstructed based on sequencing of a 596-bp segment of the hypervariable region I (HVRI) mtDNA in 433 individuals. Congruent results of three dimentional scaling plot and spatial analysis of molecular variance exhibited relatively close affiliations among population within the Sakon Nakorn Basin, while analysis of molecular variance and Mantel test revealed the predominant geographic factor in determining population affinity. Three demographic evolutionary models described by language (Model 1), geography (Model 2), and recent migration (Model 3) were propose to evaluate whether model was fitted to describe mtDNA data. Approximate Bayesian Computation and a type I error results strongly selected Model 2, supporting that geography is the primary influential factor underlying genetic divergence of studied populations.

Keywords : mtDNA-HVRI/ genetic affinity/ Approximate Bayesian Computation/Austro-Asiatic/ Tai-Kadai/ Sakon Nakorn Basin/Korat Basin/ Northeastern Thailand

Output จากโครงการวิจัยที่ได้รับทุนจาก สกว.

1. ผลงานตีพิมพ์ในวารสารวิชาการนานาชาติ

   Wibhu Kutanan*, Silvia Ghirotto, Giorgio Bertorelle, Suparat Srithawong, Kanokpohn Srithongdaeng, Nattapon Pontham, and Daoroong Kangwanpong. 2014. Geography has more influence than language on maternal genetic structure of various Northeastern Thai ethnicities. Journal of Human Genetics (in revision).

2. การนำผลงานวิจัยไปใช้ประโยชน์
   - ผลงานวิจัยครั้งนี้ จะถูกนำไปใช้ในการเรียนการสอนวิชา Human evolution and Population Genetics ซึ่งเป็นวิชาที่สอนในระดับบัณฑิตศึกษาของภาควิชาชีววิทยา คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น
   - ในขั้นตอนการทำวิจัย ผู้วิจัยได้รับความร่วมมือจากนักศึกษาที่สนใจความรู้ด้านพันธุศาสตร์ประชากรและสามารถสร้างนักวิจัยรุ่นเยาว์ ตั้งแต่ระดับมัธยม (นักเรียน โครงการ พสวท.) จำนวน 1 คน ระดับปริญญาตรี จำนวน 4 คน และระดับปริญญาโท จำนวน 1 คน ดังนั้นผลลัพท์ที่ได้นอกจจากงานวิจัยแล้วยังสามารถสร้างนักวิจัยรุ่นเยาว์ จำนวนรวม 6 คน

กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณอาสาสมัครที่เสียสละเวลาเข้าร่วมโครงการวิจัย ขอขอบคุณทีมงานผู้ช่วยวิจัยในการติดต่อประสานงานและเก็บตัวอย่างประชากร ดังรายนามต่อไปนี้ รตอ.หญิง พิชชาภา บุญโสดา นางสาวศุภรัตน์ ศรีทะวงษ์ นางสาวอาทิตยา คำเหลา นายณัตพล พลธรรม และนางสาวกนกพร ศรีทองแดง

ขอขอบคุณ รองศาสตราจารย์ ดร.ดาวรุ่ง กังวานพงศ์ ภาควิชาชีววิทยา คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ นักวิจัยที่ปรึกษาที่คอยให้คำปรึกษา ในการทำวิจัยอย่างต่อเนื่อง และร่างบทความวิจัย

ขอขอบคุณ Assoc. Prof. Dr. Giogio Bertorelle และ Dr. Silvia Fuselli จาก Department of Life Science, University of Ferrara ประเทศอิตาลี ที่ช่วยเหลือในการวิเคราะห์ข้อมูลทางพันธุศาสตร์ประชากรขั้นสูง

งานวิจัยนี้ได้รับการสนับสนุนจากทุนพัฒนาศักยภาพในการทำงานวิจัยของอาจารย์รุ่นใหม่ สำนักงานกองทุนสนับสนุนการวิจัย (สกว.) ประจำปี พ.ศ. 2555 (Grant No. MRG5580058)

ผู้ช่วยศาสตราจารย์ ดร.วิภู กุตะนันท์

2 กรกฎาคม 2557

สารบัญ

สารบัญตาราง

สารบัญภาพ

บทนำ

ภาคตะวันออกเฉียงเหนือหรือภาคอีสานของประเทศไทยสามารถแบ่งเป็น 2 ส่วน ตามลักษณะทางภูมิศาสตร์ โดยมีเทือกเขาภูพานเป็นแนวกั้น คือ บริเวณอีสานตอนเหนือ หรือ แอ่งที่ราบสกลนคร (Sakon Nakorn Basin) และบริเวณอีสานตอนใต้ หรือแอ่งที่ราบโคราช (Khorat Basin) เนื่องจากดินแดนภาคอีสานมีอาณาเขตติดกับประเทศลาว กัมพูชาและตั้งอยู่ไม่ห่างจากประเทศเวียดนาม และทางตอนใต้ของประเทศจีน ในอดีตจึงมีการอพยพของหลายกลุ่มชาติพันธุ์เข้าสู่ภาคอีสานแห่งนี้ จึงส่งผลให้ภาคอีสานของประเทศไทยในปัจจุบันมีความหลากหลายของกลุ่มชาติพันธุ์ ภาษา และวัฒนธรรม ทำให้พื้นที่แห่งนี้ได้รับความสนใจจากนักวิชาการหลายสาขา เช่น นักภาษาศาสตร์ (Smalley, 1994) นักประวัติศาสตร์ และนักโบราณคดี (Wyatt, 1984) จากจำนวนกลุ่มชาติพันธุ์ที่อาศัยอยู่ในบริเวณภาคอีสานประมาณ 20 กลุ่ม สามารถแบ่งเป็น 2 กลุ่มหลักตามภาษาพูด คือ ประชากรที่พูดภาษาตระกูลออสโตรเอเชียติค (Austroasiatic linguistic family) เช่น มอญ (Mon) เขมร (Khmer) ส่วย (Suay) โส้ (So) และชาวบน (Chaobon) เป็นต้น และประชากรที่พูดภาษาตระกูลไท-กะได (Tai-Kadai linguistic family) เช่น แสก (Seak) ญ้อ (Nyaw) ผู้ไท (Phutai) และกะเลิง (Kaluang) และลาวอีสาน (Lao-Isan) เป็นต้น

ในการศึกษาความสัมพันธ์ระหว่างประชากรสามารถอาศัยหลักฐานจากภาษา วัฒนธรรม และหลักฐานทางโบราณคดี ซึ่งประชากรที่มีภาษาและวัฒนธรรมคล้ายกัน จะแสดงถึงการมีบรรพบุรุษทางชาติพันธุ์ (ethnic ancestor) ร่วมกัน อย่างไรก็ตามทั้งภาษาและวัฒนธรรมอาจเปลี่ยนแปลงไปตามบริบทและการอพยพเคลื่อนย้ายของประชากร เช่น การรับเอาภาษาและวัฒนธรรมจากประชากรอื่น หรือมีการผสมผสานของภาษาและวัฒนธรรม จึงอาจส่งผลให้การศึกษาความสัมพันธ์ระหว่างประชากรผิดพลาดไปได้ โดยประชากรที่มีบรรพบุรุษทางชาติพันธุ์เดียวกันอาจไม่ได้มีความเกี่ยวข้องกันทางเชื้อสาย หรือเรียกว่ามีบรรพบุรุษทางเชื้อสาย (biological ancestor) ต่างกัน ดังนั้นการศึกษาความสัมพันธ์ระหว่างประชากรให้มีความถูกต้องมากยิ่งขึ้นจึงต้องอาศัยดีเอ็นเอซึ่งเป็นสารพันธุกรรมที่ถ่ายทอดข้อมูลที่แท้จริงจากบรรพบุรุษสู่ลูกหลาน

ในศตวรรษใหม่นี้ความรู้ด้านอณูพันธุศาสตร์นับเป็นองค์ความรู้ที่มีบทบาทสำคัญยิ่ง เพราะสามารถนำไปประยุกต์ใช้ศึกษาวิวัฒนาการ และความสัมพันธ์ทางเชื้อสายระหว่างประชากรจึงทำให้ได้ข้อมูลโครงสร้างทางพันธุกรรมของประชากรมีเพิ่มมากขึ้น ดีเอ็นเอไมโทคอนเดรีย (mitochondrial DNA) เป็นเครื่องหมายทางพันธุกรรมที่ได้รับการพิสูจน์แล้วว่ามีประสิทธิภาพในการศึกษาโครงสร้างและความสัมพันธ์ระหว่างประชากร (Cavalli-Sforza and Feldman, 2003; Malyarchuk *et al*., 2008) เนื่องจากมีคุณสมบัติที่เหมาะสมหลายประการ เช่น ถ่ายทอดผ่านทางมารดาสู่ลูก (maternal inheritance) มีจำนวนมาก (high copy number) เมื่อเทียบกับดีเอ็นเอในนิวเคลียส ไม่มีรีคอมบิเนชั่น (recombination) และ มีอัตราการกลายพันธุ์ที่สูงกว่าดีเอ็นเอนิวเคลียส (high mutation rate) เป็นต้น (Pakendorf and Stoneking, 2005)

ดีเอ็นเอไมโทคอนเดรียมีโครงสร้างเป็นวงแหวนเกลียวคู่ ประกอบด้วยสายพอลีนิวคลีโอไทด์ 2 สาย โดยเรียกสายที่มีลำดับเบสพิวรีนมากกว่าว่า heavy strand (H-strand) ส่วนสายที่มีลำดับเบสเป็นไพริมิดีนมากกว่าเรียกว่า light strand (L-strand) ดีเอ็นเอไมโทคอนเดรียมียีนอยู่ทั้งหมด 37 ยีน อยู่บนสาย H-strand จำนวน 28 ยีน และอยู่บน L-strand จำนวน 9 ยีน นอกจากนี้ยังมีบริเวณ control region หรือ D-loop (displacement loop) ซึ่งเป็นบริเวณที่มีความผันแปรของลำดับเบสสูง (hypervariable region, HVR) เนื่องจากเป็นบริเวณที่ไม่ใช่ยีน ดังนั้นเมื่อเกิดการกลายพันธุ์จึงไม่ส่งผลต่อการดำรงชีวิต และสามารถสะสมการกลายพันธุ์ได้ บริเวณ HVR นี้แบ่งเป็น 2 ส่วนคือ HVR-I และ HVR-II ตรงตำแหน่งของลำดับเบสที่ 16024-16383 และ 57-372 ตามลำดับ (Greenberg *et al*., 1983) ซึ่งในการศึกษาพันธุศาสตร์ประชากรจะใช้บริเวณ HVR-I เป็นหลักเนื่องจากมีความผันแปรสูง จึงสามารถใช้แยกแยะความแตกต่างและความใกล้ชิดของประชากรได้

การศึกษาพันธุศาสตร์ประชากรในภาคอีสาน เริ่มต้นในปี ค.ศ. 2001 โดย Fucharoen และคณะ ทำการศึกษาความผันแปรของดีเอ็นเอไมโทคอนเดรียบริเวณ HVR-I จาก 6 กลุ่มชาติพันธุ์ในประเทศไทยได้แก่ ลีซอ มูเซอร์ ผู้ไท ลาวซ่ง จ้วง ซาไก และนำลำดับเบสของทั้ง 6 ชาติพันธุ์มาทำการวิเคราะห์เปรียบเทียบกับชาวไทยจากภาคเหนือและภาคอีสานพบว่าประชากรที่ศึกษาเช่น ชาวไทยจากภาคเหนือและภาคอีสานมีระยะห่างทางพันธุกรรมระหว่างกันสูง แสดงถึงการมีโครงสร้างทางพันธุกรรมที่แตกต่างกัน ต่อมา Lertrit *et al*. (2008) ศึกษาดีเอ็นเอไมโทคอนเดรีย บริเวณ HVR-I จากตัวอย่างโครงกระดูกโบราณ อายุประมาณ 3,000 ปี ที่ถูกขุดค้นพบในแหล่งโบราณคดีที่จังหวัดนครราชสีมา และเปรียบเทียบกับดีเอ็นเอไมโทคอนเดรียของประชากรปัจจุบัน คือชาวเขมร และชาวบน ผลการศึกษาพบว่าโครงกระดูกทั้งสองน่าจะเป็นประชากรดั้งเดิมที่พูดภาษากลุ่มย่อยมอญ-เขมร ที่อาศัยอยู่ในบริเวณพื้นที่ดังกล่าวมากว่า 3,000 ปีแล้ว ในปี ค.ศ. 2013 พิชชาภา บุญโสดา และคณะ ศึกษาความผันแปรของดีเอ็นเอไมโทคอนเดรียในประชากรชาวเขมรในจังหวัดสุรินทร์ เปรียบเทียบกับประชากรอื่นในประเทศไทย ผลระยะห่างทางพันธุกรรมระบุว่าชาวเขมรมีความสัมพันธ์ทางเชื้อสายใกล้ชิดกับชาวถิ่นกลุ่มย่อยปรัยมากที่สุด นอกจากนี้ยังพบว่าชาวเขมร ชาวถิ่นปรัย ชาวขมุ และชาวพล่าง มีภาษาพูดที่ถูกจัดอยู่ในกลุ่มตระกูลภาษามอญ-เขมร แสดงถึงความสัมพันธ์ระหว่างภาษาพูดและพันธุกรรม จนถึงปัจจุบันการศึกษาความผันแปรของดีเอ็นเอไมโทคอนเดรียในภาคตะวันออกเฉียงเหนือมีจำกัด โดยมีรายงานเพียง 5 ประชากรเท่านั้น คือ ผู้ไท ชาวบน ไทยขอนแก่น ไทยโคราช และชาวเขมร (Fucharoen *et al*., 2001; Letrit *et al*., 2008; Boonsoda *et al*., 2013) ดังนั้นงานวิจัยนี้จึงต้องการศึกษาพันธุศาสตร์ประชากรของกลุ่มชาติพันธุ์ที่อาศัยอยู่ในภาคตะวันออกเฉียงเหนือของประเทศไทย จำนวน 10 กลุ่ม คือ มอญ เขมร ส่วย โส้ ชาวบน แสก ญ้อ ผู้ไท กะเลิง และลาวอีสาน โดยอาศัยความผันแปรของดีเอ็นเอไมโทคอนเดรีย บริเวณ HVR-I โดยประชากรที่ศึกษาสามารถแบ่ง 2 กลุ่มตามภาษาพูดคือ ประชากรที่พูดภาษาตระกูลออสโตรเอเชียติค (มอญ เขมร ส่วย โส้ ชาวบน) และประชากรที่พูดภาษาไท-กะได (แสก ญ้อ ผู้ไท กะเลิง และลาวอีสาน) และตามลักษณะภูมิศาสตร์คือ ประชากรที่อาศัยอยู่ในแอ่งที่ราบสกลนคร (โส้ แสก ญ้อ ผู้ไท และกะเลิง) และประชากรที่อาศัยอยู่ในแอ่งโคราช (มอญ เขมร ส่วย ชาวบน และลาวอีสาน)

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) ชาวบน (BON) แสก (SAK) ญ้อ (YOH) ผู้ไท (PUT) กะเลิง (KAL) และลาวอีสาน (LAO) โดยประชากรที่ศึกษาสามารถแบ่ง 2 กลุ่มตามภาษาพูด คือ ประชากรที่พูดภาษาตระกูลออสโตรเอเชียติค (AA) (มอญ เขมร ส่วย โส้ และชาวบน) และประชากรที่พูดภาษาไท-กะได (TK) (แสก ญ้อ ผู้ไท กะเลิง และลาวอีสาน) และแบ่งตามลักษณะภูมิศาสตร์ คือ ประชากรที่อาศัยอยู่ในแอ่งที่ราบสกลนคร (SK) (โส้ แสก ญ้อ ผู้ไท และกะเลิง) และประชากรที่อาศัยอยู่ในแอ่งโคราช (KR) (มอญ เขมร ส่วย ชาวบน และลาวอีสาน) (ตารางที่ 1 และภาพที่ 1)

ก่อนทำการเก็บตัวอย่างจะมีการสัมภาษณ์ประวัติครอบครัว เพื่อคัดกรองอาสาสมัครที่ไม่มีความเกี่ยวข้องกันทางสายเลือด อย่างน้อย 2 ชั่วรุ่น จากนั้นทำการเก็บตัวอย่างทำโดยใช้แปรงเก็บเยื่อบุข้างแก้ม (buccal collection brush) (Qiagen, USA) ถูที่เนื้อเยื่อบริเวณกระพุ้งแก้มประมาณ 40 ครั้ง

2. การสกัดดีเอ็นเอและตรวจสอบคุณภาพและปริมาณของดีเอ็นเอ

นำตัวอย่างเยื่อบุข้างแก้มที่ได้มาสกัดดีเอ็นเอด้วยชุดสกัดดีเอ็นเอสำเร็จรูป Gentra Puregene Buccal Cell Kit ดังนี้

2.1 นำตัวอย่างเซลล์เยื่อบุข้างแก้มซึ่งติดมากับแปรงเก็บตัวอย่าง ใส่ลงในหลอดทดลองขนาด 1.5 มิลลิลิตร จากนั้นเติมสารละลาย (cell lysis solution) เพื่อทำให้เซลล์แตก ปริมาตร 300 ไมโครลิตร

2.2 เติมเอนไซม์ proteinase K ปริมาตร 2 ไมโครลิตร เพื่อย่อยโปรตีน

2.3 บ่มที่อุณหภูมิ 55 องศาเซลเซียส (ข้ามคืน)

2.4 นำแปรงเก็บตัวอย่างออกจากหลอดทดลอง เติมเอนไซม์ RNase ปริมาตร 2 ไมโครลิตรเพื่อย่อยอาร์เอ็นเอ

2.5 บ่มหลอดทดลองที่อุณหภูมิ 37 องศาเซลเซียส 15 นาที พลิกหลอดกลับไปกลับมา 25 ครั้ง

2.6 เติมสารเพื่อตกตะกอนโปรตีน (protein precipitation) ปริมาตร 100 ไมโครลิตร

2.7 เขย่าให้เข้ากันด้วยเครื่อง vortex 20 ครั้ง

2.8 บ่มหลอดทดลองในน้ำแข็ง 5 นาที

2.9 ปั่นเหวี่ยงด้วยความเร็ว 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 พลิกหลอกกลับไปกลับมา 20 ครั้ง

2.13 ปั่นเหวี่ยง ด้วยความเร็ว 15,000 รอบ/นาที 5 นาที และทิ้งไว้ให้แห้งที่อุณหภูมิห้อง ในขั้นตอนนี้ตะกอนที่ได้ (pellet) จะเป็นดีเอ็นเอ

2.14 เติมสารละลายดีเอ็นเอ (DNA hydration) ปริมาตร 100 ไมโครลิตร และเก็บสารละลายดีเอ็นเอที่อุณหภูมิ -20 °C

2.15 ตรวจสอบคุณภาพและปริมาณดีเอ็นเอที่สกัดได้โดยวัดค่าการดูดกลืนแสงอัลตราไวโอเลต ที่ความยาวคลื่อน 260 นาโนเมตร และ 280 นาโนเมตร ด้วยเครื่องสเปคโตรโฟโตมิเตอร์และวัดการเรืองแสงของดีเอ็นเอที่จับตัวกับเอธิเดียมโบไมด์หลังจากแยกขนาดดีเอ็นเอโดยวิธีอิเล็กโทรโฟรีซีส โดยใช้วุ้นอะกาโรสความเข้มข้น ร้อยละ 1

ตารางที่ 1 ข้อมูลทั่วไปของประชากรที่ศึกษาและค่าทางสถิติที่ใช้วิเคราะห์

ประชากร	เขมร	มอญ	ส่วย	ชาวบน	โส้
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	สำโรงทาบ, สุรินทร์	เทพสถิตย์, ชัยภูมิ	กุสุมาลย์, สกลนคร
จำนวนตัวอย่าง	68	44	44	42	47
ขนาดประชากร	1,266,828	1,000	407,724	6,283	71,532
แฮโปลไทป์	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 *Fs*	-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

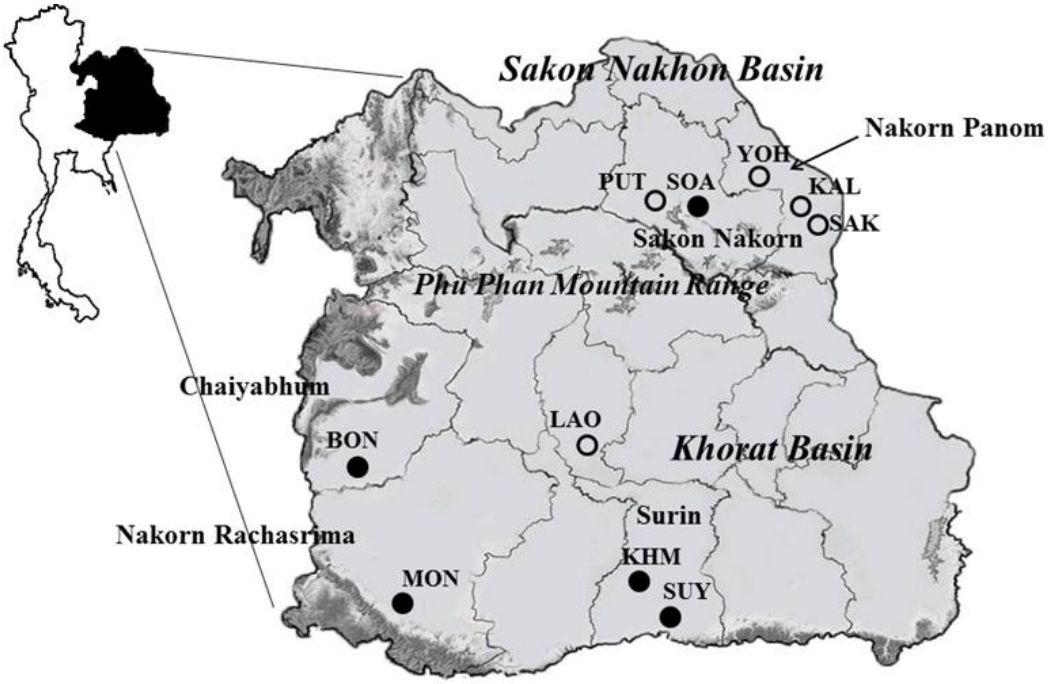
ตารางที่ 1 (ต่อ) ข้อมูลทั่วไปของประชากรที่ศึกษาและค่าทางสถิติที่ใช้วิเคราะห์

ประชากร	ลาวอีสาน	ภูไท	ญ้อ	แสก	กะเลิง
อักษรย่อ	LAO	PUT	YOH	SAK	KAL
ละติจูด	15.62	17.28	17.55	17.45	17.33
ลองติจูด	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
แฮโปลไทป์	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 *Fs*	-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 ● แสดงประชากรที่พูดภาษาตระกูลออสโตรเอเชียติค ○ แสดงประชากรที่พูดภาษาตระกูลไท-กะได

3. การเพิ่มขยายชิ้นส่วนของดีเอ็นเอไมโทคอนเดรียด้วยปฏิกิริยาลูกโซ่พอลีเมอเรส (พีซีอาร์)

เพิ่มปริมาณดีเอ็นเอ ไมโทคอนเดรียบริเวณ D-loop ด้วยปฏิกิริยาลูกโซ่พอลีเมอเรส โดยใช้เอนไซม์ *Pfu* DNA polymerase (Enzynomics, 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 Pfu polymerase	0.50
50 นาโนกรัม/ไมโครลิตร ดีเอ็นเอ	0.50
น้ำกลั่นปราศจากเชื้อ	34.00
รวม	**50.00**

ทำปฏิกิริยาลูกโซ่พอลีเมอเรสในเครื่องควบคุมอุณหภูมิอัตโนมัติ (Thermal cycler) โดยใช้อุณหภูมิต่างๆ ดังนี้

ช่วงที่ 1 (1รอบ)	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 ด้วยวิธีอิเล็กโทรโฟรีซีส โดยใช้วุ้นอะกาโรสความเข้มข้น ร้อยละ 1 โดยเทียบขนาดผลผลิตพีซีอาร์กับดีเอ็นเอมารตฐาน (DNA ladder) (Norgen Biotek Corp, Thorold Ontario, Canada)

4. หาลำดับเบส

ทำการหาลำดับเบสของผลผลิตพีซีอาร์ โดยใช้ไพร์เมอร์ที่จำเพาะอีกคู่ 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 Cambride reference sequence) วิเคราะห์ความหลากหลายของนิวคลีโอไทด์ (nucleotide diversity, π) ด้วยโปรแกรม DNASP v.5 (Librado and Rozas, 2009)

5.3 วิเคราะห์ความหลากหลายของแฮโปลไทป์ (haplotype diversity, *h*) ชนิดของแฮโปลไทป์ ค่า mean number of pairwise difference (MPD) ค่าพารามิเตอร์ของการเพิ่มจำนวนประชากร (demographic expansion) เช่น raggedness index value *(r)* (Harpending, 1994) และ neutrality test คือ Fu's *Fs* (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 Multidimentional 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 โดยการกำหนดประชากรตามภาษาพูดและลักษณะทางภูมิศาสตร์ แล้วทดสอบ ค่าความแคความมีนัยสำคัญทางสถิติด้วยวิธี 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 ซึ่งถูกกำหนดด้วยหมายเลย 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) ระหว่างคู่ของประชากรโดยจะใช้พิกัดของละติจูดและลองติจูดของที่ตั้งหมู่บ้านของแต่ละประชากร

	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	

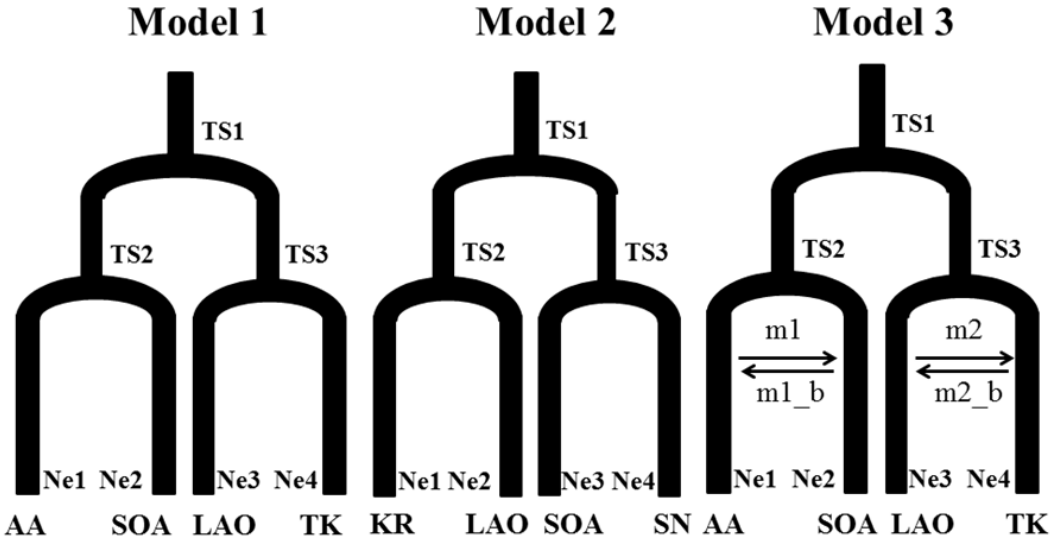
5.8 การเลือกโมเดลด้วย Approximate Bayesian Computation

สร้างโมเดลแสดงความสัมพันธ์ทางวิวัฒนาการระหว่างประชากร ดังภาพที่ 2 โดยพิจารณาจากตำแหน่งที่ตั้งทางภูมิศาสตร์และภาษาพูด โดยประชากรที่เป็นตัวแปรคือโส้ (SOA) และ ลาวอีสาน (LAO) เนื่องจาก SOA เป็นประชากรที่พูดภาษาตระกูลออสโตรเอเชียติค (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 ออกจากกันเมื่อระยะเวลา 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 จะมีอัตราการอพยพคงที่ (m1 และ m1_b) และ SOA และ TK จะมีอัตราการอพยพคงที่ (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 เพื่อให้ได้ค่าทางสถิติที่มีค่าใกล้เคียงกับค่าสังเกตุมากที่สุด (มี Euclidean distance น้อยที่สุด) จากนั้นคำนวณ posterior probabilities ด้วยวิธี Approximate Bayesian Computation (ABC) (Bertorelle *et al*., 2010)

การศึกษาครั้งนี้จะใช้ 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

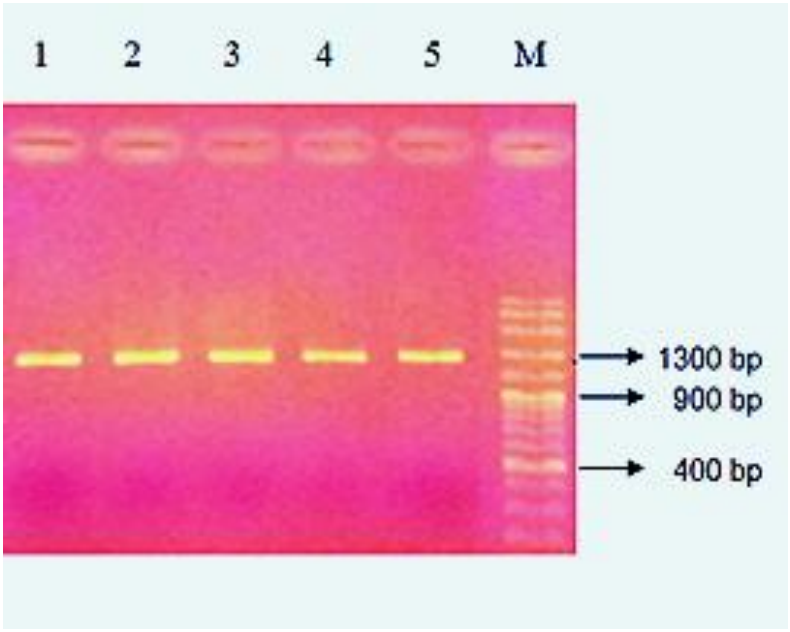
ตารางที่ 3 Prior distributions ของพารามิเตอร์ในแต่ละโมเดล โดย N แสดง effective population sizes T แสดงระยะเวลาในการแยกระหว่างประชากร หน่วยเป็นชั่วรุ่น และ m แทนอัตราการกลายพันธุ์ หน่วยเป็นต่อเบสต่อชั่วรุ่น

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 ได้ขนาดผลผลิตจากการทำปฏิกิริยาลูกโซ่พอลีเมอเรสประมาณ 1,200 คู่เบส เทียบขนาดกับ 100 bp DNA Ladder (Norgen Biotek Corp, Thorold Ontario, Canada) (ภาพที่ 3)



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

2. ความหลากหลายทางพันธุกรรม

เมื่อนำผลผลิตจากการทำปฏิกิริยาลูกโซ่พอลีเมอเรสไปหาลำดับเบส บริเวณ HVR-I ทั้งสาย H และ L และทำการรวมลำดับเบสทั้งสองสายแล้ว จะได้ลำดับเบสที่มีความยาว 596 คู่เบส (ตำแหน่งที่ 16001 ถึง 16596) จากตัวอย่างดีเอ็นเอที่ใช้ศึกษาจำนวน 433 ตัวอย่าง พบแฮโปลไทป์ที่แตกต่างกันทั้งหมด 173 แฮโปลไทป์ จากตำแหน่งที่เกิดความผันแปร (polymorphic site) ทั้งหมด 135 ตำแหน่ง (ภาคผนวก 1) โดยเป็นแฮโปลไทป์ที่พบเพียงตัวอย่างเดียว (single unique) จำนวน 108 แฮโปลไทป์ และพบแฮโปไทป์ที่พบมากกว่า 1 ตัวอย่าง แต่พบเพียงประชากรเดียว (multiple unique) จำนวน 34 ตัวอย่าง ส่วนที่เหลืออีก 34 แฮโปลไทป์ จะเป็นแฮโปลไทป์ที่พบมากกว่า 1 ประชากร (shared haplotypes) โดยประชากรที่มี shared haplotypes จำนวนมากที่สุด คือ ประชากร SOA-PUT จำนวน 6 แฮโปลไทป์ ในขณะที่ไม่พบแฮโปลไทป์ชนิด shared ระหว่างคู่ของประชากรจำนวน 5 คู่ ดังนี้ MON-

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

ตารางที่ 4 ครึ่งล่างซ้ายแสดงระยะห่างทางพันธุกรรมระหว่างประชากร แบบ pairwise F_{st} และครึ่งล่างขวาแสดงจำนวนแฮโปลไทป์ที่เหมือนกันระหว่างประชากร 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

เมื่อคำนวณความหลากหลายของแฮโปลไทป์ (*h*) ซึ่งมีค่าสูงสุดในประชากร LAO **(**0.9899) และต่ำสุดในประชารกร SAK (0.7920**)** ความหลากหลายของนิวคลีโอไทด์ **(*π*)** 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) และสูงสุดในประชากร PUT (8.6956) (ตารางที่ 3) ซึ่งแสดงถึงการมีความเหมือนกันทางพันธุกรรม (genetic homogeneity) หรือการเกิด recent divergence ในตัวอย่างประชากรชาวมอญ ในขณะที่ตัวอย่างในประชากรชาวภูไทมีความผันแปรทางพันธุกรรม (genetic heterogeneity) มากที่สุด จากนั้นเมื่อพิจารณาค่า inter-MPD ซึ่ง อยู่ในช่วงระหว่าง 7.2024 ถึง 10.2660 ช่วงดังกล่าวมีค่าสูง เมื่อเปรียบเทียบกับผลการรายงานจากงานวิจัยก่อนหน้านี้ (Bodner *et al*., 2011) ซึ่งแสดงถึงการมีโครงสร้างทางพันธุกรรมที่แตกต่างกันระหว่างประชากร

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

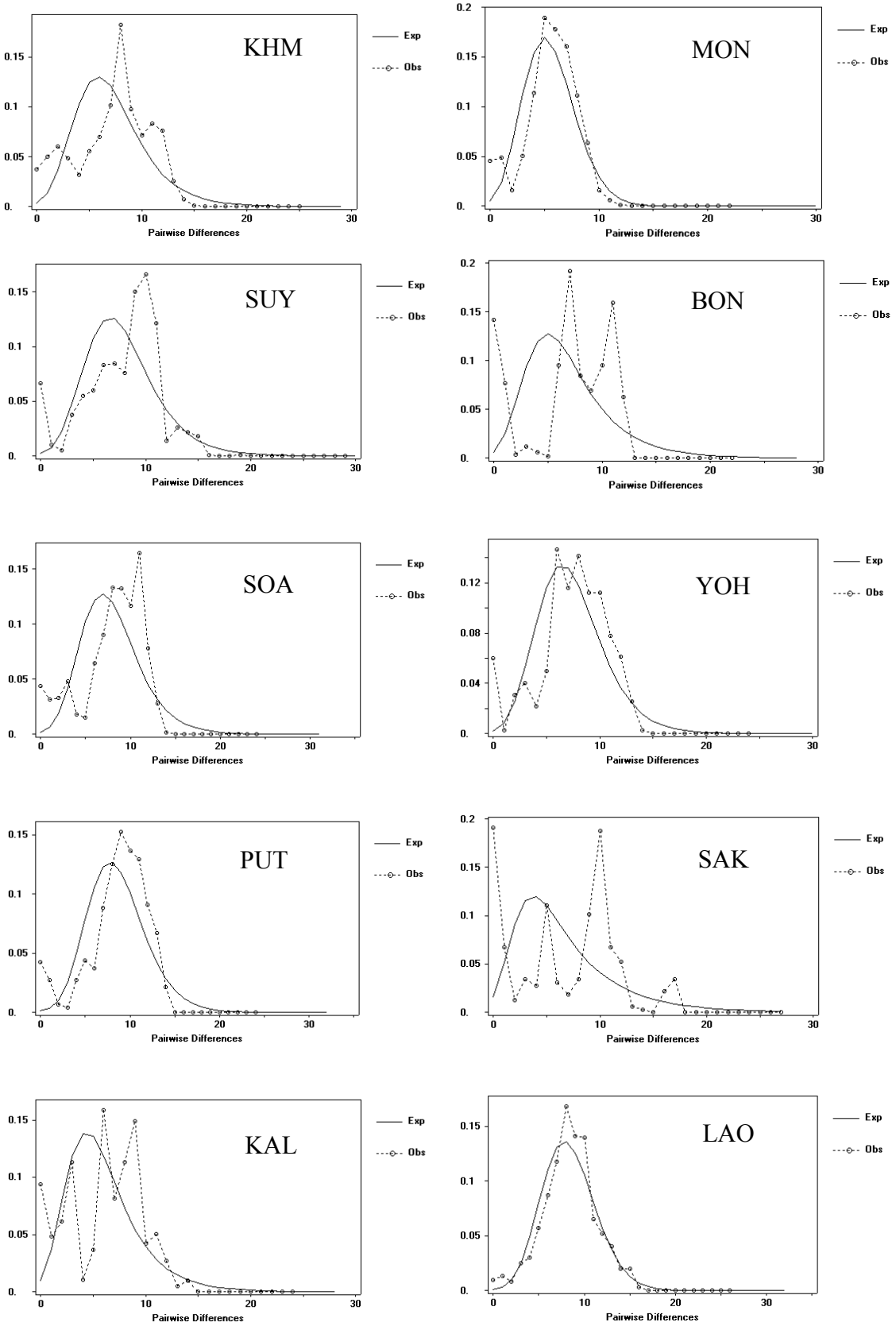
	KHM	MON	SUY	BON	SOA	LAO	PUT	YOH	SAK	KAL
KHM	7.1242	7.4937	7.9806	7.4471	8.4136	8.0042	8.2767	7.6869	8.8677	7.2024
MON	1.1689	5.5254	7.6245	7.2944	7.6267	7.6734	7.8098	7.0183	8.5008	7.5420
SUY	0.3656	0.8089	8.1057	7.7808	9.2810	8.7221	9.0730	8.1452	10.2660	8.3389
BON	0.4715	1.1182	0.3145	6.8270	9.0613	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) เป็นการวิเคราะห์หาการเพิ่มจำนวนของประชากรอย่างรวดเร็วในอดีต (rapid population expansion) โดยแสดงจากกราฟการกระจายของจำนวนเบสที่แตกต่างกันระหว่างแฮโปลไทป์แต่ละคู่ในประชากรนั้น 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 *Fs* และ Tajima' D พบว่าค่า Fu's *Fs* ในประชากร KHM MON SOA LAO และ PUT มีค่าติดลบอย่างมากและมีนัยสำคัญทางสถิติ ($P < 0.05$) (ตารางที่ 1) จากการคำนวน neutrality test พบว่าปฏิเสธ null hypothesis แม้ว่าค่า 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 ซึ่งพบว่าค่า 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} ต่ำและไม่มีนัยสำคัญทางสถิติ ซึ่งแสดงถึงการมีโครงสร้างทางพันธุกรรมที่เหมือนกันระหว่างประชากร KHM กับ LAO และ SUY กับ BON เป็นที่น่าสนใจว่าประชากรที่อาศัยอยู่ในแอ่งที่ราบสกลนคร (SK) มีค่า F_{st} ที่แตกต่างกันอย่างไม่มีนัยสำคัญ SOA PUT และ YOH ซึ่งแสดงถึงการมีโครงสร้างทางพันธุกรรมที่เหมือนกัน ซึ่งอาจเกิดจากยีนโฟล์วระหว่างประชากรในแอ่งที่ราบสกลนคร

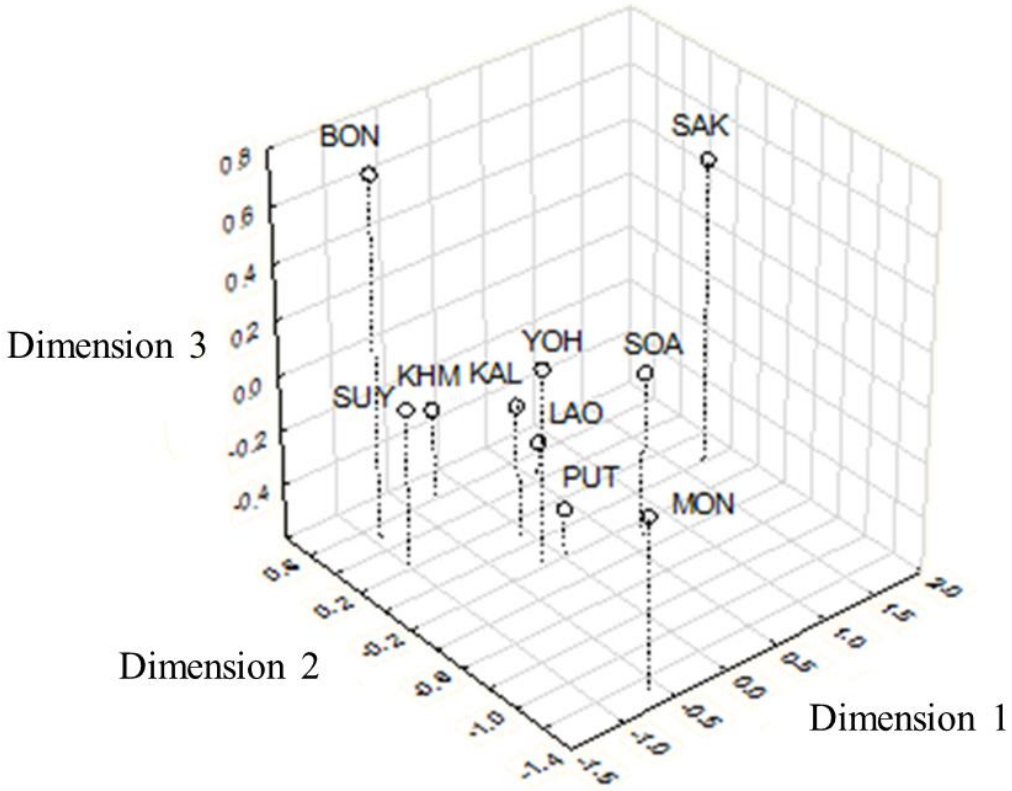
รูปแบบของความแตกต่างทางพันธุกรรมระหว่างประชากร จากการคำนวณค่า F_{st} และ corrected MPD มีความสอดคล้องกัน (ตารางที่ 4) โดยชาว SAK มีโครงสร้างทางพันธุกรรมที่แตกต่างจากประชากรอื่นมากที่สุด รองลองมาคือ BON MON และ SUY ตามลำดับ

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

ตารางที่ 6 การวิเคราะห์ SAMOVA อักษรย่อของประชากรแสดงในตารางที่ 1

จำนวนกลุ่ม	สมาชิกของประชากรในแต่ละกลุ่ม									F_{ct}
2	SAK	KHM,MON,SUY,BON,SOA,LAO,PUT,YOH,KAL								0.1276
3	SAK	MON	KHMSUY,BON,SOA,LAO,PUT,YOH,KAL							0.0849
4	SAK	MON	SUY, BON	KHM,SOA,LAO,PUT,YOH,KAL						**0.0809**
5	SAK	MON	SUY,BON	KHM	SOA,LAO,PUT,YOH,KAL					**0.0713**
6	SAK	MON	SUY	BON	KHM	SOA,LAO,PUT,YOH,KAL				**0.0728**
7	SAK	MON	SUY	BON	KHM	SOA, KAL	LAO,PUT,YOH			**0.0693**
8	SAK	MON	SUY	BON	KHM	SOA	KAL	LAO,PUT, YOH		0.0702
9	SAK	MON	SUY	BON	KHM	SOA	KAL	YOH	LAO, PUT	0.0664

ตัวหนาแสดงระดับนัยสำคัญทางสถิติ ที่ $P< 0.01$

F_{ct} = Fixation index among groups

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) จากผลดังกล่าวแสดงถึงประชากรที่อาศัยอยู่ในแต่ภูมิศาสตร์มีความใกล้ชิดทางพันธุกรรม 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

การทดสอบเมนเทลเพื่อวิเคราะห์ correlation test และ and partial correlation test ระหว่างเมทริกซ์ของระยะห่างทางพันธุกรรม (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)] แต่พบว่าระยะห่างทางพันธุกรรมไม่สอดคล้องกับระยะห่างทางภาษา [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

	จำนวนกลุ่ม	จำนวนประชากร	ระดับความผันแปร (ร้อยละ)					
			a	b	c	F_{st}	F_{sc}	F_{ct}
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 ในแต่ละโมดล 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 เพื่อประเมินว่าข้อมูลและวิธีการที่นำคัดเลือกโมเดลน่าเชื่อถือได้หรือไม่ เนื่องจากโมเดลทั้ง 3 แบบ ค่อนข้างคล้ายกันและข้อมูลพันธุศาสตร์มีเพียงดีเอ็นเอไมโทคอนเดรีย ซึ่งมีการถ่ายทอดในรูปแบบแฮโปลไทป์จึงถูกพิจารษเป็นเพียงเครื่องหมายทางพันธุกรรม 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 โดยฉพาะเมื่อเพิ่มระดับของ 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 1 (true)	Model 2	Model 3	Not Assigned		Model 1 (true)	Model 2	Model 3	Not 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 1	Model 2 (true)	Model 3	Not Assigned		Model 1	Model 2 (true)	Model 3	Not 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 1	Model 2	Model 3 (true)	Not Assigned		Model 1	Model 2	Model 3 (true)	Not 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 ของการศึกษาครั้งนี้ทำให้สามารถศึกษาความผันแปรทางพันธุกรรมในระดับภูมิภาค (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) ซึ่งเป็นบริเวณที่ประชากรที่พูดภาษาตระกูลออสโตรเอเชียติค (AA) อาศัยอยู่เป็นจำนวนมาก ในขณะที่บริเวณแอ่งที่ราบสกลนคร (SN) จะมีกลุ่มประชากรที่พูดภาษาตระกูลไท-กะได (TK) ซึ่งอพยพมาจากประเทศลาวและวียดนาม อาศัยอยู่เป็นจำนวนมาก

ในแอ่งที่ราบสกลนคร ประชากรที่ศึกษาประกอบด้วย ชาวโส้ (SOA) ภูไท (PUT) ไทแสก (SAK) กะเลิง (KAL) และไทญ้อ (YOH) ซึ่ง SOA เป็นประชากรเดียวในแอ่งที่ราบสกลนครที่มีภาษาพูดจัดอยู่ในกลุ่ม AA ในขณะที่ประชากรที่เหลือพูดภาษา TT ชาวโส้มีถิ่นฐานเดิมอยู่ในบริเวณแขวงคำม่วน ประเทศลาว ในปี ค.ศ.1844 ชาวโส้บางส่วนได้อพยพ เข้าสู่ดินแดนของประเทศไทย อาศัยอยู่ในเขต อ. กุสุมาลย์ จากผลการวิเคราะห์ ABC (ภาพที่ 2) โมเดลทางวิวัฒนาการที่เสนอ พบว่าชาวโส้มีโครงสร้างทางพันธุกรรมที่คล้ายกับประชากรอื่นในแอ่งที่ราบสกลนคร ซึ่งสาเหตุดังกล่าวอาจเกิดจากการผสมผสานทางพันธุกกรมระหว่างชาวโส้และประชากรข้างเคียงหลังจากการอพยพเข้ามาอาศัยในประเทศไทย เมื่อประมาณ 200 ปีที่ผ่านมา นอกจากนั้นการที่ SOA กับ PUT และ SOA กับ YOH มีแฮโปลไทป์ที่เหมือนกัน ถึง 6 และ 5 แฮโปลไทป์ 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 (มีค่าความหลากหลายของแฮโปลไทป์และ intra-MPD ต่ำที่สุด) นอกจากนั้นข้อถกเถียงสำหรับการจัดกลุ่มภาษาของ SAK โดยมีนักภาษาศาสตร์กลุ่มเก่าเชื่อว่าภาษาพูดของชาว SAK เป็นภาษา AA ตระกูลย่อยมอญ-เขมร ในขณะที่นักภาษาศาสตร์กลุ่มใหม่จัดภาษาแสกให้อยู่ในตระกูล TK ตระกูลย่อยไทเหนือ (Smalley, 1994; Schliesinger, 2000) ผลจากการวิเคราะห์ดีเอ็นเอไมโทคอนเดรีย พบว่า SAK มีความสัมพันธ์ทางเชื้อสายใกล้กับชาว SOA (พูด AA) มากที่สุด จากการศึกษาถึงความสัมพันธ์ระหว่างภาษาและพันธุกรรมจำนวนมาก ทำให้สามารถอนุมานว่าประชากรที่มีภาษาพูดใกล้เคียงกันจะมีโครงสร้างทางพันธุกรรมที่คล้ายกัน (Barbujani and Sokal, 1990; Cavalli-Sforza *et al*., 1992; Barbujani and Pilastro, 1993; Boattini *et al*., 2011) ดังนั้นจากการศึกษาครั้งนี้พบว่าการจัดกลุ่มภาษาของชาวแสกไม่สอดคล้องกับโครงสร้างทางพันธุกรรม อย่างไรก็ตามภาพรวมของโครงสร้างทางพันธุกรรมของ SAK อาจได้รับอิทธิพลจากปัจจัยอื่นที่มีอิทธิพลมากกว่า เช่นภู มิศาสตร์

ในแอ่งที่ราบโคราช ประชากรที่ศึกษามีทั้งหมด 5 ประชากร คือเขมร (KHM) ชาวบน (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 ฮัญกูร นักภาษาศาสตร์เชื่อว่าชาวบนเป็นกลุ่มชนดั้งเดิมที่อาศัยอยู่ในพื้นที่แห่งนี้ก่อนที่ชาวเขมรและกลุ่มคนที่พูดภาษาตระกูลไท-กะไดจะครอบครองดินแดนแห่งนี้ ชาวบนมีถิ่นที่อยู่อาศัยในประเทศไทยเท่านั้น ในเขต จ.ชัยภูมิ จ.เพชรบูรณ์ และ จ.นครราชสีมา โดยตัวอย่างของประชากรที่ศึกษาเป็นชาวบนจากบ้านหวังอ้ายโพธิ์ อ.เทพสถิตย์ จ.ชัยภูมิ ซึ่งเป็นหมู่บ้านชาวบนที่ยังคงมีภาษาและวัฒนธรรมที่เป็นเอกลักษณ์ (Premsrirat, 2002; Prasert, 2009) ผลการศึกษาของพารามิเตอร์ของความหลากหลายทางพันธุกรรมที่มีค่าต่ำ เช่น *h S* และ intra-MPD (ตารางที่ 1) ซึ่งอาจเกิดจากสาเหตุการผสมเลือดชิดเนื่องจากประเพณีนิยมการแต่งงานภายในกลุ่ม อย่างไรก็ตาม ในปัจจุบันชาวบนมีการผสมผสานกับชาวลาวอีสานมากยิ่งขึ้น ก็อาจทำให้การมีโครงสร้างทางพันธุกรรมที่จำเพาะนี้จางหายไป การศึกษาครั้งนี้จึงเป็นการศึกษาด้านพันธุศาสตร์ครั้งแรกที่รายงานความหลากหลายทางพันธุกรรมที่ต่ำในชาวบนจากบ้านหวังอ้ายโพธิ์ นอกจากนั้นนักภาษาศาสตร์ระบุว่าภาษาพูดของชาวบนเป็นภาษาเดียวกับภาษามอญโบราณ ตั้งแต่สมัยทวาราวดี ดังนั้นถ้าใช้หลักว่าประชากรที่มีภาษาพูดใกล้เคียงกันจะมีโครงสร้างทางพันธุกรรมที่คล้ายกัน ก็คาดว่า BON และ MON จะมีความสัมพันธ์ทางเชื้อสายใกล้ชิดกัน แต่ผลการศึกษาครั้งนี้กลับตรงกันข้าม กล่าวคือ จะพบว่า BON และ MON มีระยะห่างทางพันธุกรรมที่ห่างกัน จึงสามารถสรุปได้ว่าชาวมอญ จาก จ.นครราชสีมา และชาวบนจาก จ.ชัยภูมิไม่มีความสัมพันธ์ทางเชื้อสายกัน

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

มอญ ถูกสร้างขึ้นประมาณศตวรรษที่ 3 และรุ่งเรืองจนถึงศตวรรษที่ 10 ในบริเวณตอนกลางของประเทศไทย (Schliesinger, 2000) อาณาจักรทวาราวดีได้แพร่ขยายไปทั่วทุกสารทิศ ทั้งภาคเหนือและภาคใต้ของประเทศไทยในปัจจุบัน จนกระทั่งในปี ค.ศ. 1775 ชาวมอญกลุ่มหนึ่งได้อพยพมาจากประเทศพม่าเข้ามาอาศัยอยู่ในบริเวณจ.นครราชสีมา จนกระทั่ง ในปี ค.ศ. 1793 จ.นครราชสีมา มีชาวมอญอาศัยอยู่ประมาณ 2,500 คน จากประวัติศาสตร์การอพยพ MON ที่ศึกษาครั้งนี้มาจากประเทศพม่า ดังนั้นการไม่พบความสัมพันธ์ทางเชื้อสายระหว่าง MON และ BON จึงน่าจะมีสาเหตุมาจากชาวมอญในพม่าและชาวมอญในอาณาจักรทวาราวดี มีพันธุกรรมที่แตกต่างกันมาตั้งแต่อดีต สิ่งที่น่าสนใจคือผลการศึกษาทางด้านพันธุศาสตร์ครั้งนี้ยังพบว่าชาว MON มีค่า π intra-MPD และจำนวนของแฮโปลไทป์เฉพาะที่ต่ำ (ตารางที่ 1) และยังแสดงการมีการเพิ่มขยายขนาดของประชากร (ตารางที่ 1 และภาพที่ 4) ดังนั้นประชากร MON จึงน่าจะเกิดปรากฏการณ์คอขวด (bottleneck) และหลังจากนั้นมีการเพิ่มขยายขนาดของประชากรอย่างรวดเร็ว

ชาวส่วย (SUY) เป็นกลุ่มชาติพันธุ์ที่พูดภาษาออสฌตรเอเชียติค กุล่มย่อยอมอญ-เขมรอีกกลุ่มหนึ่งในบริเวณภาคตะวันออกเฉียงเหนือตอนใต้ ผลการศึกษาพบว่า SUY มีความใกล้ชิดทางพันธุกรรมกับ KHM และ BON (ตารางที่ 3 และ ภาพที่ 5) สอดคล้องกับหลักฐานทางประวัติศาสตร์ ภาษา และสังคม ที่ระบุความสัมพันธ์ที่ใกล้ชิดกันระหว่าง SUY และ KHM ชาวส่วยหรือกุย (Kui) หรือ ชาวเขมรเรียกว่าขะแมร์-บอเร็น (แปลว่าเขมรโบราณ) เป็นประชากรดั้งเดิมที่อาศัยอยู่ในพื้นที่ของประเทศไทย ลาว และกัมพูชาในปัจจุบัน ก่อนที่ชาวเขมรและไท-กะไดจะครอบครองดินแดนแห่งนี้ ชาวส่วยที่ศึกษาอาศัยอยู่ในจังหวัดสุรินทร์ซึ่งอพยพมาจากประเทศลาวจำนวน 2 ช่วง คือ ระหว่างปี ค.ศ. 1656-1688 และ การอพยพอีกจำนวนมาก ในปี ค.ศ.1760 ในปัจจุบันชาวส่วยที่อาศัยอยู่ในภาคตะวันออกเฉียงเหนือของประเทศไทย ได้รับเอาภาษาพูดของชาวลาวอีสานและเขมร โดยชาวส่วยที่สามารถพูดภาษาลาวได้ เรียกว่า ลาวส่วย และส่วยที่สามารถพูดภาษาเขมรได้ เรียก เขมรส่วย ดังนั้นการที่ชาวส่วยมีโครงสร้างทางพันธุกรรมใกล้กับชาวเขมรจึงน่าจะเกิดมาจากการผสมผสานระหว่างกลุ่มชาติพันธุ์ อย่างไรก็ตามได้มีนักมานุษยวิทยา เสนอถึงความสัมพันธ์ระหว่างกลุ่มชาติพันธุ์ส่วยและชาวบน (Saard, 1984) ซึ่งผลการศึกษาครั้งนี้ก็สนับสนุนสมมติฐานดังกล่าว

งานวิจัยนี้ทำการศึกษาโครงสร้างและความผันแปรทางพันธุกรรมของหลายกลุ่มชาติพันธุ์ที่อาศัยอยู่ในภาคตะวันออกเฉียงเหนือของประเทศไทย ผลการวิเคราะห์ทางพันธุศาสตร์พบว่าปัจจัยทางด้านภูมิศาสตร์มีอิทธิพลต่อความผันแปรทางพันธุกรรมของประชากรที่ศึกษา โดยประชากรที่อาศัยอยู่ในแอ่งที่ราบเดียวกันจะมีความสัมพันธ์ทางเชื้อสายใกล้ชิดกัน โดยเฉพาะประชากรที่อาศัยอยู่ในแอ่งที่ราบสกลนคร อย่างไรก็ตามภายในแอ่งที่ราบเดียวกัน 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., Castrì, 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. & Wuattro, 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´nsson, 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).

Premsrirat, 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. & Janthed, 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).

Torroni, 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-Chromosal Distribution in Northern European Populations. *Mol. Biol. Evo1*. **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).

ภาคผนวก

ภาคผนวก 1

ลำดับเบสของแต่ละแฮโปลไทป์

	haplotype (Sequence)
1	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACCCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCCACC CACCCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTATAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAG GGGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGT GAACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
2	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCTAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATGAAAACCCAACCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
3	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCTA?CCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATGAAAACCCAACCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAATTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCAC CCTTAACAGCACATAGTACATAAAGCCATTTATCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
4	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCCCCTCAT GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCCTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
5	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCTA?CCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATGAAAACCCAACCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGTCACCCCTCACCCACTAGGATACCAACAAACCTACCCAC CCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAA?ACATCACGATG
6	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCCCCCATCAACAACCGCTATGTATTTCGTA CATTACTGCTAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAACCCACATCAAAACCCCCTCCCCATG CTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGTCACCCCTCACCCACTAGGATACCAACAAACCTACCCA CCCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGG GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA ACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
7	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAAATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAACCCCCCCCCCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCATACATCAACTGCAACTCCAAAGCCACCCCTCAACCACTAGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
8	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCTAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATGAAAACCCAACCCACATCAAAACCCCCCCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGTCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
9	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCTAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATGAAAACCCAACCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATTAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG

	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACCAGGATACCAACAAACCTACC
	CACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAG
1	GGGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGT
0	GAACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCTAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAACCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCATACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTATCC
	ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
1	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
1	AACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAACCCCCCCCCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCATACATCAACTGCAACTCCAAAGCCACCCCTCAACCACTAGGATACCAACAAACCTACCC
	ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
1	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
2	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCTAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATGAAAACCCAACCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC
	ACCCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
1	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
3	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATCTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAACCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC
	ACCCTCAACAGTACATAGCACATAAAGCCATTTATCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
1	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
4	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCCCCTCATG
	CTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCA
	CCCCTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGG
1	GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA
5	ACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAACCCCCCCCCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCTCTCAACCACTAGGATACCAACAAACCTACCC
	ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
1	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
6	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCTCATG
	CTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCA
	CCCCTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGG
1	GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA
7	ACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTACTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAACCCCCCCCCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCAACCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
8	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG

	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
19	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGATCACCTGCAGTACATAAAAACCCAATCCACATCAAACACCCCCCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
20	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATG CTTACAAGCAAGCACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTGGGATACCAACAAACCTACCCA CCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGG GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA ACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
21	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCCTCAACTATCATACATCAACTGCAACCCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCCACCCAC CCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
22	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCATACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ATCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCGTCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGT GAACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
23	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAATCCCCTCCCCATG CTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCA CCCTCAACAGTACATAGTACATAAAACCATTTACCGTACATAGCACATTACAGTCAAATCCCCTCTCGTCCCCATGGATGACCCCCCTCAGATAGGG GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA ACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
24	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
25	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAACCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
26	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAACCCCCCCCCCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCAACCACTAGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
27	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAACCCCCCCCCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCAATCACTAGGATACCAACAAACCTACCC GCCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG

	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAACCCCCCCCCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC
	ACCCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
2	AGTCCCTTGGCCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
8	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCCATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC
	ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
2	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
9	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCTAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAACCCACATCAAAACCCCCTCCCCATG
	CTTACAAGCAAGTACAGCAATCAACCTTCAACTATCATACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTATCCA
	CCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGG
3	GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA
0	ACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATG
	CTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCA
	CCCCTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGG
3	GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA
1	ACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCTAGCCACCATGAAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATGAAAACCCAACCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGTCACCCCTCACCCACTAGGATACCAACAAACCTACCC
	ACCCTTAACGGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
3	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
2	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAACCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACCCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCCACC
	CACCCTTAATAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAG
3	GAGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGT
3	GAACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAGCCCAATCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAACCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC
	ACCCTTAACAGTACATAGTACATAAAACCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCTCATGGATGACCCCCCCCAGATAGG
3	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
4	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC
	ACCCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
3	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
5	AACTGTATCCGACATCTGGTTCCTACTTCAGGGGCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAACCCCCCCCCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC
	ACCCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
3	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
6	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
3	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA

7	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAACCCCCCCCCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCAACCACTAGAATACCAACAAACCTACCC ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
38	ATTCTAATTTAAACTATTCTCTGTTCTTTCAGGGGGAAGCAAATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTACTTCGTA CATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATG CTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCA CCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGGG GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA ACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
39	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCCCCCCCATG CTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCA CCCTCAACAGTACATAGTACATAAAGCCATTTATCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGG GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA ACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
40	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCAC CCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
41	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCATCCACTAGGATACCAACAAACCTACCCAC CCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
42	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTAACCACCTGTAGTACATAAAAACCCAATCCATATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCTCTCACCCACTAGGATACCAACAAACCTACCCAC CCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCA?ATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAA?AGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
43	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCACCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAACCCCCCCCCCCATGC TTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACTCAC CCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
44	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCTAC CCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
45	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
46	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATGC

	TTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAAGATACCAACAAACCTACCCAC CCTTAACAGCACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
47	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC TTTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCAC CCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
48	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCCCCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCTAC CCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
49	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCTAC CCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
50	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCCATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGCACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTGGGATACCAACAAACCTACCCAC CCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
51	ATTCTAATTTAAACTATTCTCTGTTCTTTCAGGGGGAAGCAAATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCA?CCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAGCCCCCTCCCCATG CTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATATCAACAAACCTACCCA CCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGG GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA ACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
52	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCAC CCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
53	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATG CTTACAAGCAAGTACAGCAATTAACCTTCAACTACCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCA CCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCCTCTCGCCCCCATGGATGACCCCCCTCAGATAGGG GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA ACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
54	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCCATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATG CTTACAAGCAAGCACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTGGGATACCAACAAACCTACCCA CCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGG GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA ACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
55	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTATCCAC

	CCTTAACAGTACATAGTACATAAAACCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
5 6	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATCTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTAACCACCTGTAGTACATAAAAACCCAATCCATATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCTCTCACCCACTAGGATACCAACAAACCTACCCAC CCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
5 7	ATTCTAATTTAAACTATTCTCTGTTCTTTCAGGGGGAAGCAAATTTGGGTACCACCCAAGTATTGACTCCCCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATG CTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCATCCACTAGGATACCAACAAACCTACCCA CCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGG GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA ACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
5 8	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCTAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATGAAAACCCAACCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCTTAACAGCACATAGTACATAAAGCCATTTATCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
5 9	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAACCCCCCCCCCCAT GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATTAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCTCAACAGTACATAGTACATAAAGCCATTTATCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
6 0	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTTTTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCAC CCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
6 1	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
6 2	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCTA?CCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAACCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCTTCAACTATCATACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTATCCAC CCTTAGCAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
6 3	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGTAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCAC CCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
6 4	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACCCCAAAGCCATCCCTCACCCACTAGGATACCAACAAACCCACC CACCCTTAATAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAG

	GAGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGT GAACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
65	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATTAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTTACCCACTAGGATACCAACAAACCTACCCAC CCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
66	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAA- CCCAATCCACATCAAAACCCCCTCCCCATGCTTACAAGCAAGTACAGTAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCT CACCCACTAGGATACCAACAAACCTACCCACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCG TCCCCATGGATGACCCCCCTCAGATAGGGGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGG GCCCATAACACTTGGGGGTAGCTAAAGTGAACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAA ATAAGACATCACGATG
67	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCCTCAACTATCATACATCAACTGCAACCCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCCACCCAC CCTTAACAGCACATAGTACATAAAGCCTTTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCAGGGATGACCCCCCTCAAATGGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCAACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCTTTAAATAAGACATCACGATG
68	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAACCCCCCCCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCAACCACTAGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
69	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCCCCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAA- CCCAATCCACATCAAAACCCCCTCCCCATGCTTACAAGCAAGTACAGTAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCT CACCCACTAGGATACCAACAAACCTACCCACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCG TCCCCATGGATGACCCCCCTCAGATAGGGGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGG GCCCATAACACTTGGGGGTAGCTAAAGTGAACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAA ATAAGACATCACGATG
70	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCCTCAACTATCATACATCAACTGCAACCCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCCACCCAC CCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACACCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCAAACGTTCCCCTTAAATAAGACATCACGATG
71	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAACCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCAC CCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGACG
72	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCCTCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAACCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCAC CCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAACCCAAAAAGCCCCCACGTTCCCCTTAAATAAGACATCACGATG
73	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAACCCACATCAAAACCTCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCAC

	CCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
74	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
75	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCCACCCAC CCTTAACAGCACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGAG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
76	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTGCCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAGGCCACCCCTTACCCACTAGGATACCAACAAACCTACCCAC CCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
77	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCATACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATATCAACAAACCTACCC ATCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
78	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATGC TTACAAGCAGGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCAC CCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGGCCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
79	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCTAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATGAAAACCCAACCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCTTAACAGCACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
80	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAATCCTCAACTATCACACATCAACTGCAACCCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCCACCCAC CCTTAATAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGAG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
81	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCCACCCAC CCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGAG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
82	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATGAAAACCCAACCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCAC CCTTAACAGCACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGG

	TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
83	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCTCATGC TTACAAGCAAGTACAGCAATCAACCTTCAACTATCATACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATATCAACAAACCTACCCAT CCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
84	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCTAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATGAAAACCCAACCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCTTAACAGCACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGGCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
85	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCCCCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCAC CCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
86	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCCA?CCACCATGAATATTGTACAGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCATTCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCCACCCAC CCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGAG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
87	ATTCTAATTTAAACTATTCTCTGTTCTTTCA?GGGGAAGCA?ATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTAC ATTACTGCTA?CCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATGAAAACCCAATCCACATCAAAACCCCCTCCCCATGC TTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCAC CCTTAACAGCACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGG TCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAA CTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
88	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAACCCCCCCCCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCAACCACTAGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
89	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAACCCACATCAAACCCCCCCCCCCAT GCT?ACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
90	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCATATCAAAACCCCCTCCCCATG CTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACCAGGATACCAACAAACCTACCCA CCCCTAACAGTACATGGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGG GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA ACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
91	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAACCCCCCCCCCCCAT GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCTCTCAACCACTAGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG

	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
92	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCCATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATG CTTACAAGCAAGCACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCCCACCCACTGGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
93	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
94	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCTAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATGAAAACCCAACCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAATTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
95	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCATCCCTCACCCACTAGGATATCAACAAACCTACTC ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
96	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAACCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCAACCTCTCACCCACTAGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
97	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAACCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGCACAGCAATCAACCTTCAACTATCATACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTATCC ACCCTTAACAGTACATAGTACATAAAACCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
98	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATCTCGTA CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATG CTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGTCACCCCTCACCCACTAGGATACCAACAAACCTACCCA CCCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTCCTCGTCCCCATGGATGACCCCCCTCAGATAGGG GTCCCTTGGCCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA ACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
99	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATCTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAA- CCCAATCCACATCAAAACCCCCTCCCCATGCTTACAAGCAAGCACAGTAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCT CACCCACTAGGATACCAACAAACCTACCCACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCG TCCCCATGGATGACCCCCCTCAGATAGGGGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGG GCCCATAACACTTGGGGGTAGCTAAAGTGAACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAA ATAAGACATCACGATG
100	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAACCCCCCCCCCCAT GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCAACCACTAGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG

	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
10 1	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAACCCCCCCCCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCAACCACTAGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGTACATAAAGCCATTCACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GATCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
10 2	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGATCACCTGTAGTACATAAAAACCCAATCCACATCAAACACCCCCCCCCCAT GCTTACAAGCAAGTACAGCAACCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACC CATCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAG GGGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGT GAACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
10 3	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCCCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATATCAACAAACCTACCC CCCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
10 4	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATG CTTACAAGCAAGCACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCCCACCCACTGGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
10 5	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAACCCCCCCCCCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCAACCACTAGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGTCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
10 6	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATCTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAACCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCTCAACAGTACATAGTACATAAAGCCATTTATCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
10 7	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATCGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCCCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATATCAACAAACCTACCC CCCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
10 8	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGATCACCTGTAGTACATAAAAACCCAATCCACATCAAACACCCCCCCCCCAT GCTTACAAGCAAGTACAGCAACCAACCCTCAACTATCACGCATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACC CACCCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAG GGGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGT GAACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
10 9	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCTAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATGAAAACCCAACCCACATCAAAATCCCCTCCCCATG CTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCA CCCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGG GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA ACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG

	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
1	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
0	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACTACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATG
	CTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCA
1	CCCTCAACAGTACATAGTACATAAAGCCATTTATCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGA
1	GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA
1	ACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATCTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAACCCCCCCCCCCCCAT
	GCTTACAAGCAAGTACAGCAACCAACCCTCAACTATCACACATCAACTGCAATTCCAAAGCCACCTCTCACCCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTCCTCGTCCCCATGGATGACCCCCCTCAGATAGG
1	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
2	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAACCCACATCAAACCCCCCCCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATATCAACAAACCTACCC
1	ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
1	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
3	AACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCGACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTATTC
1	ACTCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
1	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
4	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCTAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAACCCACATCAAAATCCCCTCCCCATG
	CTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCA
1	CCCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGG
1	GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA
5	ACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATG
	CTTACAAGCAAGCACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTGGGATACCAACAAACCTACCCA
1	CCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGG
1	GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA
6	ACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACCCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGCACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAACCCCCCCCCCCCAT
	GCTGACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGTACATAGTACATAAAACCATTCACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
1	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
7	AACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATGAAAACCCAACCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
1	GGTCCCTTGGCCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGT
8	GAACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
1	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA

19	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGAATACCAACAAACCTACCC ACCCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
1 2 0	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCCCCCCAT GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCC- AACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGGTCCC TTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAACTGT ATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
1 2 1	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCCCCCCAT GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCAACCTCTCACCCACTAGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
1 2 2	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCTCATG CTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGAATACCAACAAACCCACCCA CCCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGA GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA ACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
1 2 3	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCTCATG CTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCA CCCCTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGG GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA ACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
1 2 4	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCCATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAACCCCCCCCCCCAT GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCTTTCTCGCCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
1 2 5	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTTTTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACTC ACCCTTAACAGCACATAGTACATAAGGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
1 2 6	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACAAAAAAACCCAATCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATTAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGGCATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
1 2 7	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCTAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAACCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCATACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTATCC ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCACACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
1	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA

2	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAACCCCCCCCCCCCCAT
8	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCAACCACTAGGATACCAACAAACCTACCC
	ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCTCCCCCCCAT
	GCTTACAAGCAAGCACAGCAATCAACCCTCAACTATCACACATCAATTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGCACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
2	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
9	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCCCCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCAACCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGTACATAGTACATAAAGCCATTCACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
3	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
0	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCACCCCCCCCCCCCCAT
	GCTTACAAGCAAGTACAGCAACCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCTCTCACCCACTAGGATACCAACAAACCTACCT
1	ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
3	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
1	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCACAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAACCCCCCCCCCCAT
	GCTTACAAGCAAGTACAGCAACCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACC
1	CACCCTTAACAGTACATGGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCTCTTCTTGTCCCCATGGATGACCCCCCTCAGATAG
3	GGGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGT
2	GAACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCCCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC
1	ACCCCTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
3	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
3	AACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCTATG
	CTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCGACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCA
1	CCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGG
3	GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA
4	ACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCACCCCCCCCCCCCCAT
	GCTTACAAGCAAGTACAACAACCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCTCTCACCCACTAGGATACCAACAAACCTACCT
1	ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
3	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
5	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCCCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC
1	ACCCTCAACAGTACATAGTACATAAAACCATTTATCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
3	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
6	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
1	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
3	CATTACTGCCAGCCACCATGAATATTGTACGGTACCACAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAACCCCCCCCCCCAT

7	GCTTACAAGCAAGTACAGCAACCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACC
	CACCCTTAACAGTACATGGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCTCTTCTCGTCCCCATGGATGACCCCCCTCAGATAG
	GGGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGT
	GAACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCTACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGG
3	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
8	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACTCCTCACCCACTAGGATACCAACAAACCTACCC
1	ACCCTCAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCTTTCTCGCCCCCATGGATGACCCCCCTCAGATAGG
3	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
9	AACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCTAGCCACCATGAATATTGCACAGTACCATAAATACTTGACCACCTGTAGTACATGAAAACCCAACCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
4	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
0	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAACCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
4	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAATACTTGGGGGTAGCTAAAGTG
1	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCAACCCCTCACCCACTAGGATACCAACAAACCTACCT
1	ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCATCCCCATGGATGACCCCCCTCAGATAGG
4	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
2	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAACCCCCCCCCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCATACATCAACTGCAACTCCAAAGCAACCCCTCAACCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
4	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
3	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCAACCTCTCACCCACTAGGATACCAACAAACCTACCT
1	ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGG
4	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
4	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAACCCCCCCCCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCTCTCAACCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
4	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
5	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAAAAAGACATCACGATG
1	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATCTCGTA
4	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAACCCCCCTCCCCAT
6	GCTTACAAGCAAGTACAGTAACCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCTCTCACCCACTAGGATACCAACAAACCTACCC

	ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGTACATAGTACATAAAACCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGG
4	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
7	AACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACCCCAAAGCTACCCCTCACCCACTAGGATACCAACAAACCCACC
1	CACCCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAG
4	GAGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGT
8	GAACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGCACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAACCCCCCCCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCAACCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
4	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAGGT
9	GAACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATATCAACAAACTTACCC
1	ACCCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
5	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
0	AACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATCTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCATACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGG
5	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
1	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAACCCCCCCCCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCAACCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGG
5	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
2	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGTCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGCACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAACCCCCCCCCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCAACCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
5	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAAGGT
3	GAACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATCTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAACCCCCCCCCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCTCTCAACCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
5	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
4	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
1	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTAACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATG
5	CTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCA
5	CCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGGG

	GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA
	ACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAACCCCCCCCCCCCCAT
	GCTT??AAACCTACCCACCCTTAACAGTACAT
1	AGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGGTCCCTTGACCACCA
5	TCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAACTGTATCCGACATC
6	TGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAACCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
5	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
7	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAACCCCCCCCCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCTCTCAACCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
5	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
8	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGATCACCTGTAGTACATAAAAACCCAATCCACATCAAACACCCCCCCCCATG
	CTTACAAGCAAGTACAGCAACCAACCCTCAACTATCACGCATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
5	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
9	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCCCAGATAGG
6	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
0	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAATCCCCCCCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC
1	ACCCTCAACAGTACATAGCACATAAAACCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
6	AGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACCTAAGAGTAGCTAAAGTG
1	AACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACTCCCTCCCCATG
	CTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGTCACCCCTCACCCACTAGGATATCAACAAACCTACCCA
1	CCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGG
6	GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA
2	ACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAACCCCCCCCCCCCAT
	GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCGACTGCAACTCCAAAGCCACCTCTCAACCACTAGGATACCAACAAACCTACCC
1	ACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG
6	GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG
3	AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA
	CATTACTGCCAGTCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAATTCCCCCCCCATG
1	CTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCA
6	CTCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGG
4	GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA

	ACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
1 6 5	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAACCTCCC- CCCCATGCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAAC CTACCCCCCCCTTAACAGTACATAGTACATAAAGCCATTCACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCA GATAGGGGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGC TAAAGTGAACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
1 6 6	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAACCCCCCCCCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCAACCACTAGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGTACATAAAGCCATTCACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
1 6 7	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAACACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAACCCCCCCCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCAACCACTAGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGTACATAAAGCCATTCACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
1 6 8	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCTAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATGAAAACCCAACCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGAAGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
1 6 9	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCTCAACAGTACATAGTACATAAAGCCATTTATCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
1 7 0	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCCATGTACTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCTTCAACTATCACACATCAACCGCAACTCCAAAGCCACCCCTCACCCACTAGGATATCAACAAACCTACCC ACCCTTAACAGTACATAGCACATAAAACCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
1 7 1	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCTAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAACCCACATCAAAACCCCCTCCCCATG CTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCTCTCACCCACTAGGATACCAACAAACCTACCCA CCCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGG GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA ACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
1 7 2	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCAT GCTTACAAGCAAGTACAGCAATCAACCTTCAACTAGCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCC ACCCTTAACAGTACATAGCACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGCCCCCATGGATGACCCCCCTCAGATAGG GGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTG AACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
1 7 3	ATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTA CATTACTGCCAGTCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCTCCCCATG CTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCATCCACTAGGATACCAACAAACCTACCCA CCCTTAACAGCACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGG GTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGA ACTGTATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG

ภาคผนวก 2

จำนวนของแฮโปลไทป์ที่พบในแต่ละประชากร

	KHM	MON	SUY	BON	SOA	LAO	PUT	YOH	SAK	KAL	Total
Sample size	68	44	44	42	47	35	38	41	27	46	433
Haplotype1	1	0	0	0	0	0	0	0	0	0	1
Haplotype2	5	0	0	0	3	0	6	1	0	3	18
Haplotype3	1	0	7	0	0	0	0	0	0	0	8
Haplotype4	1	0	0	0	0	0	0	0	0	0	1
Haplotype5	10	0	2	0	0	0	0	0	0	0	12
Haplotype6	1	0	0	0	0	0	0	0	0	0	1
Haplotype7	1	0	0	0	0	0	0	0	0	0	1
Haplotype8	2	0	0	0	0	0	0	0	0	0	2
Haplotype9	3	0	0	0	0	1	0	0	0	0	4
Haplotype10	1	0	0	0	0	0	0	0	0	0	1
Haplotype11	1	0	0	0	0	0	0	0	0	0	1
Haplotype12	1	0	0	0	0	0	0	0	0	1	2
Haplotype13	1	0	0	0	0	0	0	0	0	0	1
Haplotype14	1	0	0	0	0	0	0	0	0	0	1
Haplotype15	1	0	0	0	0	0	0	0	0	1	2
Haplotype16	1	0	0	0	8	0	0	0	0	0	9
Haplotype17	1	0	0	0	1	1	1	1	0	0	5
Haplotype18	1	0	0	0	0	0	0	0	0	0	1
Haplotype19	1	0	0	0	0	0	0	0	0	0	1
Haplotype20	1	0	0	0	0	0	1	0	0	0	2
Haplotype21	1	1	3	0	0	0	0	0	0	0	5
Haplotype22	1	0	0	0	0	0	0	0	0	0	1
Haplotype23	1	0	0	0	0	0	0	0	0	0	1
Haplotype24	7	0	0	0	0	0	0	0	0	0	7
Haplotype25	4	0	0	0	0	0	0	7	0	9	20
Haplotype26	1	0	0	0	0	0	0	0	0	0	1
Haplotype27	1	0	0	0	0	0	0	0	0	0	1
Haplotype28	2	0	0	0	0	0	0	0	0	0	2
Haplotype29	4	0	1	0	0	0	0	2	0	1	8
Haplotype30	2	0	2	0	0	1	0	0	0	0	5
Haplotype31	1	0	0	0	0	0	0	0	0	0	1
Haplotype32	1	0	0	0	0	0	0	0	0	0	1
Haplotype33	2	0	0	0	0	0	0	0	0	0	2
Haplotype34	1	0	0	0	0	0	0	0	0	0	1
Haplotype35	1	0	0	0	0	0	0	0	0	0	1
Haplotype36	1	0	0	0	0	0	0	0	0	0	1

Haplotype37	1	0	0	0	0	0	0	0	0	0	1
Haplotype38	0	2	0	0	0	0	0	0	0	0	2
Haplotype39	0	1	0	0	0	0	0	0	0	0	1
Haplotype40	0	1	0	0	0	0	0	0	0	0	1
Haplotype41	0	2	0	0	0	0	0	0	0	0	2
Haplotype42	0	5	0	0	0	0	0	0	0	0	5
Haplotype43	0	2	0	0	0	0	0	0	0	0	2
Haplotype44	0	5	0	0	0	0	0	0	0	0	5
Haplotype45	0	5	0	0	0	0	0	0	1	0	6
Haplotype46	0	2	0	0	0	0	0	0	0	0	2
Haplotype47	0	1	0	0	0	0	0	0	0	0	1
Haplotype48	0	1	0	0	0	0	0	0	0	0	1
Haplotype49	0	1	0	0	0	0	0	0	0	0	1
Haplotype50	0	1	0	0	0	0	0	0	0	0	1
Haplotype51	0	4	0	0	0	0	0	0	0	0	4
Haplotype52	0	1	0	0	0	0	0	0	0	0	1
Haplotype53	0	1	0	0	0	0	0	0	0	0	1
Haplotype54	0	3	0	0	2	0	0	4	0	0	9
Haplotype55	0	1	0	0	0	0	0	0	0	0	1
Haplotype56	0	1	0	0	0	0	0	0	0	0	1
Haplotype57	0	1	0	0	0	0	0	0	0	0	1
Haplotype58	0	1	0	0	0	0	0	0	0	1	2
Haplotype59	0	1	0	0	0	0	0	0	0	0	1
Haplotype60	0	0	1	0	0	0	0	0	0	0	1
Haplotype61	0	0	1	7	0	0	0	1	0	1	10
Haplotype62	0	0	1	0	0	0	0	0	0	0	1
Haplotype63	0	0	1	0	0	0	0	0	0	0	1
Haplotype64	0	0	5	0	0	0	0	2	0	0	7
Haplotype65	0	0	4	0	0	0	0	0	0	0	4
Haplotype66	0	0	6	0	1	0	1	0	0	0	8
Haplotype67	0	0	1	0	0	0	0	0	0	0	1
Haplotype68	0	0	1	0	3	1	0	3	1	10	19
Haplotype69	0	0	1	0	0	0	0	0	0	0	1
Haplotype70	0	0	1	0	0	0	0	0	0	0	1
Haplotype71	0	0	1	0	0	0	0	0	0	0	1
Haplotype72	0	0	1	0	0	0	0	0	0	0	1
Haplotype73	0	0	1	0	0	0	0	0	0	0	1
Haplotype74	0	0	1	0	0	1	0	0	0	0	2

Haplotype75	0	0	1	0	0	0	0	0	0	0	1
Haplotype76	0	0	1	0	0	0	0	0	0	0	1
Haplotype77	0	0	0	7	0	0	0	0	0	0	7
Haplotype78	0	0	0	1	0	0	0	0	0	0	1
Haplotype79	0	0	0	12	0	1	1	0	1	0	15
Haplotype80	0	0	0	5	0	0	0	0	0	0	5
Haplotype81	0	0	0	2	0	0	0	0	0	0	2
Haplotype82	0	0	0	1	0	0	0	0	0	0	1
Haplotype83	0	0	0	3	0	0	0	0	0	0	3
Haplotype84	0	0	0	1	0	0	0	0	0	0	1
Haplotype85	0	0	0	1	0	0	0	0	0	0	1
Haplotype86	0	0	0	1	0	0	0	0	0	0	1
Haplotype87	0	0	0	1	0	0	0	0	0	0	1
Haplotype88	0	0	0	0	3	1	0	0	0	0	4
Haplotype89	0	0	0	0	2	0	0	0	0	0	2
Haplotype90	0	0	0	0	2	0	0	0	0	0	2
Haplotype91	0	0	0	0	1	0	0	0	0	0	1
Haplotype92	0	0	0	0	3	0	2	0	0	0	5
Haplotype93	0	0	0	0	1	1	1	0	0	0	3
Haplotype94	0	0	0	0	1	0	0	0	0	0	1
Haplotype95	0	0	0	0	1	0	0	0	0	0	1
Haplotype96	0	0	0	0	2	0	0	0	0	0	2
Haplotype97	0	0	0	0	2	0	0	0	0	0	2
Haplotype98	0	0	0	0	1	0	0	0	0	0	1
Haplotype99	0	0	0	0	1	0	0	0	0	0	1
Haplotype100	0	0	0	0	1	0	0	0	0	0	1
Haplotype101	0	0	0	0	1	0	0	0	0	0	1
Haplotype102	0	0	0	0	1	0	0	0	0	0	1
Haplotype103	0	0	0	0	1	0	0	0	0	0	1
Haplotype104	0	0	0	0	1	0	0	0	0	0	1
Haplotype105	0	0	0	0	1	0	0	1	0	0	2
Haplotype106	0	0	0	0	1	0	0	0	0	0	1
Haplotype107	0	0	0	0	1	0	0	0	0	0	1
Haplotype108	0	0	0	0	1	0	1	0	0	0	2
Haplotype109	0	0	0	0	0	3	0	0	0	0	3
Haplotype110	0	0	0	0	0	2	0	0	0	0	2
Haplotype111	0	0	0	0	0	1	0	0	0	0	1
Haplotype112	0	0	0	0	0	1	0	0	0	0	1

Haplotype113	0	0	0	0	0	2	0	0	0	0	2
Haplotype114	0	0	0	0	0	1	0	0	0	0	1
Haplotype115	0	0	0	0	0	1	0	0	0	0	1
Haplotype116	0	0	0	0	0	1	0	0	0	0	1
Haplotype117	0	0	0	0	0	1	0	0	0	0	1
Haplotype118	0	0	0	0	0	1	0	0	0	0	1
Haplotype119	0	0	0	0	0	1	0	0	0	0	1
Haplotype120	0	0	0	0	0	1	0	0	0	0	1
Haplotype121	0	0	0	0	0	1	0	0	0	0	1
Haplotype122	0	0	0	0	0	2	0	0	0	0	2
Haplotype123	0	0	0	0	0	1	0	0	0	0	1
Haplotype124	0	0	0	0	0	1	0	0	0	0	1
Haplotype125	0	0	0	0	0	1	0	0	0	0	1
Haplotype126	0	0	0	0	0	1	0	0	0	0	1
Haplotype127	0	0	0	0	0	1	0	0	0	0	1
Haplotype128	0	0	0	0	0	1	0	0	0	1	2
Haplotype129	0	0	0	0	0	1	0	0	0	0	1
Haplotype130	0	0	0	0	0	1	0	0	0	0	1
Haplotype131	0	0	0	0	0	0	1	0	0	0	1
Haplotype132	0	0	0	0	0	0	1	0	0	0	1
Haplotype133	0	0	0	0	0	0	2	0	0	0	2
Haplotype134	0	0	0	0	0	0	1	0	0	0	1
Haplotype135	0	0	0	0	0	0	3	0	0	0	3
Haplotype136	0	0	0	0	0	0	1	0	0	0	1
Haplotype137	0	0	0	0	0	0	4	0	0	0	4
Haplotype138	0	0	0	0	0	0	3	0	0	0	3
Haplotype139	0	0	0	0	0	0	1	0	0	0	1
Haplotype140	0	0	0	0	0	0	1	0	0	0	1
Haplotype141	0	0	0	0	0	0	2	2	0	0	4
Haplotype142	0	0	0	0	0	0	1	0	0	0	1
Haplotype143	0	0	0	0	0	0	1	0	0	0	1
Haplotype144	0	0	0	0	0	0	1	0	0	0	1
Haplotype145	0	0	0	0	0	0	1	0	0	0	1
Haplotype146	0	0	0	0	0	0	0	1	0	0	1
Haplotype147	0	0	0	0	0	0	0	1	0	0	1
Haplotype148	0	0	0	0	0	0	0	4	0	0	4
Haplotype149	0	0	0	0	0	0	0	1	3	0	4
Haplotype150	0	0	0	0	0	0	0	5	0	0	5

Haplotype151	0	0	0	0	0	0	0	1	0	0	1
Haplotype152	0	0	0	0	0	0	0	1	0	0	1
Haplotype153	0	0	0	0	0	0	0	1	0	0	1
Haplotype154	0	0	0	0	0	0	0	1	0	0	1
Haplotype155	0	0	0	0	0	0	0	1	0	0	1
Haplotype156	0	0	0	0	0	0	0	0	3	0	3
Haplotype157	0	0	0	0	0	0	0	0	3	0	3
Haplotype158	0	0	0	0	0	0	0	0	12	1	13
Haplotype159	0	0	0	0	0	0	0	0	1	0	1
Haplotype160	0	0	0	0	0	0	0	0	1	0	1
Haplotype161	0	0	0	0	0	0	0	0	1	0	1
Haplotype162	0	0	0	0	0	0	0	0	1	0	1
Haplotype163	0	0	0	0	0	0	0	0	0	3	3
Haplotype164	0	0	0	0	0	0	0	0	0	1	1
Haplotype165	0	0	0	0	0	0	0	0	0	1	1
Haplotype166	0	0	0	0	0	0	0	0	0	1	1
Haplotype167	0	0	0	0	0	0	0	0	0	1	1
Haplotype168	0	0	0	0	0	0	0	0	0	5	5
Haplotype169	0	0	0	0	0	0	0	0	0	1	1
Haplotype170	0	0	0	0	0	0	0	0	0	1	1
Haplotype171	0	0	0	0	0	0	0	0	0	1	1
Haplotype172	0	0	0	0	0	0	0	0	0	1	1
Haplotype173	0	0	0	0	0	0	0	0	0	1	1

ภาคผนวก 3

manuscript

Title: Geography has more influence than language on maternal genetic structure of various Northeastern Thai ethnicities

Contributor's names: Wibhu Kutanan[1*], Silvia Ghirotto[2], Giorgio Bertorelle[2],Suparat Srithawong[1], Kanokpohn Srithongdaeng[1], Nattapon Pontham[1], and Daoroong Kangwanpong[3]

[1]Department of Biology, Faculty of Science, Khon Kaen University, Khon Kaen 40002, Thailand

[2]Department of Life Science and Biotechnology, University of Ferrara, Ferrara 44100, Italy.

[3]Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand

*__Correspondence__: Dr.WibhuKutanan, Department of Biology, Faculty of Science, Khon Kaen University, 123 Moo16, Mittapap Road, Mueang District, Khon Kaen 40002, Thailand. Email: wibhu@kku.ac.th

Running Title: Maternal lineage in the Northeast of Thailand

Keywords: mtDNA-HVRI/ genetic structure/ Approximate Bayesian Computation/Austro-Asiatic/ Tai-Kadai/ Sakon Nakhon Basin/Korat Basin/Northeastern Thailand

Abstract

Several literatures have shown the influence of geographic and linguistic factors in shaping genetic variation patterns, but their relative impact, if any, in the very heterogeneous Northeastern region of Thailand has not yet been studied. This area, called Isan, is geographically structured in two wide basins, the Sakon Nakorn Basin and the Korat Basin, serving today as home to diverse ethnicities encompassing two different linguistic families, i.e., the Austro-Asiatic; Suay (Kui), Mon, Chaobon (Nyahkur), So and Khmer, and the Tai-Kadai; Saek, Nyaw, Phu Tai, Kaleung and Lao Isan. In this study, we evaluated the relative role of geographic distance and barriers as well as linguistic differences as possible causes affecting the maternal genetic distances among Northeastern Thai ethnicities. A 596-bp segment of the hypervariable region I (HVRI) mtDNA was utilized to elucidate the genetic structure and biological affinity from 433 individuals. Different statistical analyses agreed in suggesting that most ethnic groups in the Sakon Nakorn Basin are closely related. Mantel test revealed that genetic distances were highly associated to geographic ($r = 0.445$, $P < 0.01$) but not to linguistic ($r = 0.001$, $P > 0.01$) distances. Three evolutionary models were compared by Approximate Bayesian Computation. The posterior probability of the scenario which assumed an initial population divergence possibly related to reduced gene flow among basins was equal or higher than 0.87. All analyses exhibited concordant results supporting that geography was the most relevant factor in determining the maternal genetic structure of Northeastern Thai populations.

Introduction

Northeastern Thailand or Isan is geographically located on the Khorat Plateau. Nearly exactly as wide as one third of Thailand, with almost the same population size, Isan shares borders with Laos to the north and the east and with Cambodia to the southeast. The Phu Phan Mountain Range straddles its northeastern interior, separating it into two wide basins, the Khorat Basin in the southwest and the Sakon Nakhon Basin in the northeast (Figure 1). Numerous archaeological excavations, e.g. in Ban Chiang sites, indicate that Isan was primarily inhabited by prehistoric people. The Chaobon and the Suay, who speak languages classified in the Austro-Asiatic family, sub-family Mon-Khmer, are regarded as the original inhabitants in this region before any prosperous civilizations.[1] During the historical period multiple evidence show that, prior to Angkor influence, the Isan region was dominated by two competing kingdoms: Dvaravati, a Mon-Buddhist culture from central Thailand, and Chenla, a Khmer-Hindu culture from Cambodia. During the early 9th century A.D., the Angkorian Khmer state was established. Isan was then integrated into the state as reflected by several remarkable archaeological records, particularly abundant in the Khorat Basin.[1-2] In the 14th century A.D., the Khmer civilization declined and was unable to resist to the Tai-Kadai speaking people who established the Kingdom of Lan Xang centered in Luang Prabang, in the present-day Laos. Between the late 18th and the early 19th century, during the war within the Lao kingdom, the dissidents began taking refuge into the area of Isan; this represents the first documented evidence of migration from Lao to Isan region. Again in 1827 A.D., an enormous number of Lao people were forcibly migrated to Isan[3], thus increasing the dominance of Lao culture in the Isan region but, at the same time, weakening the populations of modern-day Laos.[4] At that time, besides the Lao people, other ethnic groups from Laos and Vietnam migrated to the area of Northeastern Thailand, including several Tai-Kadai speaking groups, e.g., Phu Tai, Saek, Nyaw and Kaleung as well as the So,

one of the Mon-Khmer speaking populations. Most of them lived in villages along the Mekong River and its tributaries in the Sakon Nakhon Basin. In 1893 A.D., the Isan region became part of the Kingdom of Siam (Thailand) as a result of the Franco-Siamese War.[1, 3]

Isan's long history as well as variety of ethnicities (approximately eighteen groups populated in two distinct geographic locations) make this region an excellent area to elucidate genetic variation and its tentative influencing factors such as geography, language, and culture. A general and simplifying assumption when studying linguistic variation among populations is that a common language frequently signifies a common origin and a related language indicates a common origin further back in time.[5] Such linguistic relationships should be reflected by genetic variation and might be correlated with geographic distances according to a model of Isolation by Distance (IBD hereafter). Under IBD, current patterns of genetic variation would basically result from the interaction between genetic drift (i.e. random fluctuation of allele frequencies in time) and dispersal of individuals between populations, neglecting all gene flow processes other than those in which movements of individuals from their birthplaces are local and random.[6-8] The correspondence between geographic, genetic and linguistic distances would hence be explained by this simple model, except in those cases in which complicating dynamics would affect the expected relationship between geographical distances and genetic and linguistic diversity. These complicating dynamics can be represented by processes of linguistic assimilation during migrations or by the presence of migration resistance factors i.e. geographical barriers to gene flow. In some cases, even language differences themselves can somehow act as a barrier to free gene flow, enhancing the genetic differentiation.[9-12] In Thailand, where both geographic and ethno-linguistic diversities exist, our previous researches showed the influence of both linguistics and geography on genetic diversity of peoples residing exclusively in the North of

Thailand.[13-15] However, it is still not clear how, and to what extent, these two factors are related with the genetic variation of Northeastern Thai populations

Maternal inherited mitochondrial DNA (mtDNA) has been proven to be a powerful genetic marker to infer population history in regional and continental frameworks[16-18], however, until now, only four studies on genetic variation of five Northeastern Thai populations (i.e., Phutai, Chaobon, Thai Khon Kaen, Thai Khorat, Thai Isan) have been published.[19-21]

In the present study, we analyzed new mtDNA data of ten Isan ethnicities speaking languages belonging to two major families, namely the Tai-Kadai (Saek, Nyaw, Phu Tai, Kaleung, and Lao Isan) and the Austro-Asiatic (So, Suay, Mon, Chaobon and Khmer), and inhabiting two geographically separated wide basins, namely the Sakon Nakhon Basin (Saek, Nyaw, Phu Tai, Kaleung and So) and the Khorat Basin (Lao Isan, Suay, Mon, Chaobon and Khmer) to evaluate the relative role of geographic distance and barriers and linguistic differences as possible causes affecting the maternal genetic distances among Northeastern Thai ethnicities.

Materials and methods

Samples and DNA extraction

We studied 433 maternally unrelated individuals (for at least three generations) from ten ethnic groups, namely Khmer (KHM), Mon (MON), Suay (SUY), Chaobon (BON), So (SOA), Lao Isan (LAO), Phu Tai (PUT), Nyaw (YOH), Saek (SAK) and Kaleung (KAL), of the Northeast of Thailand. The studied populations were linguistically classified into 2 groups, Austro-Asiatic (AA) and Tai-Kadai (TK), and geographically separated into two groups, Sakon Nakhon (SN) Basin and Khorat (KR) Basin (Table 1 and Figure 1). General information about the studied populations are listed in Table 1. Prior to sample collection,

information on linguistic, cultural aspects, village and individual history was obtained by interview and the informed consent was signed. Buccal swabs were collected from each subject by using a brush embedded in Gentra Puregene Buccal Cell Kit (Qiagen, Hilden, Germany). Genomic DNA was extracted from the collected buccal cells using Gentra Puregene Buccal Cell Kits according to the manufacturer's protocols. The use of human subjects for this study was approved by Ethics Committee for Human Research of Khon Kaen University, Thailand.

Inserted Figure 1 here

Inserted Table 1 here

mtDNAamplification and sequencing

The mtDNA control region (np15704-430) of the ten ethnic groups was amplified using published primer pairs (LLmt-A, 15704-CATAGCCAATCACTTTATTG-15723; LHmt-E, 430-CTGTTAAAAGTGCATACCGCC-410).[22] PCR reactions were performed by using *n*Pfu-Forte DNA polymerase (Enzynomics, Daejeon, Korea). Each PCR reaction mix had a final volume of 50 µl consisting of 5 µl of 10X nPfu-Forte buffer, 5 µl of 200 µM dNTP mixture, 2.5 µl of each 5 µM PCR primer, 0.5 µl of 2.5U/µl *Pfu* polymerase, 0.5 µl of 50 ng genomic DNA and 34 µl of distilled water. PCR reactions were performed under the following conditions: 2 min at 95^{o}C for an activation step, followed by 35 cycles of 30 second denaturation at 95^{o}C, 1 min primer annealing at 56^{o}C and 1 min extension at 72^{o}C, and 5 min at 72^{o}C for a final extension step. After visualization on a 1% agarose gel with a 100 bp DNA ladder (Norgen Biotek Corp, Thorold Ontario, Canada), amplicons (approximately 1,200 bp) were sent for purification and sequencing of hypervariable region I (HVRI) (np 15897-100) with a published set of primers[19] (SeqLmt-A, 15897-GTATAAACTAATACACCAGTCTTGT-15921; SeqHmt-E, 100-CAGCGTCTCGCAATGCTATCGCGTG-76) at Macrogen Inc., Seoul, Korea. The

sequencing results were edited, assembled and aligned with the revised Cambridge Reference Sequence[23] using SeqScape software v2.7 (Applied Biosystem, Foster City, CA). The HVRI sequences of all samples were submitted to GenBank (accession numbers KJ205639-KJ206068).

Statistical analyses

Genetic variation within population and demographic parameters

We identified the polymorphic sites of the mtDNA sequences of 596 nucleotides (np 16001-16569) using DnaSP v.5 software.[24] Parameters of genetic diversity within populations, i.e. mean pairwise differences (MPD) or intra-MPD, number of segregating sites (S), nucleotide diversity (π), number of observed haplotypes, and the haplotype diversity (h)[25] were calculated by Arlequin v.3.5.[26] The demographic expansion parameters, i.e., a raggedness index value (r)[27]as well as neutrality estimators such as Fu's Fs[28] and Tajima's D[29], were computed by employing the same software. The number of shared haplotypes was determined for each of the 45 possible population pairs by a simple gene-count method.

To compare the genetic variation among populations, we calculated the mean pairwise differences among populations (inter-MPD) and a measure of genetic distance between pairs of populations based on pairwise difference (F_{st}, significance tested by permutation). To characterize population affinity, we plotted in two dimensions the so calculated genetic distance matrix by means of a multidimensional scaling (MDS) using the available Statistica v.10 demo (StatSoft Ltd.). Spatial analysis of molecular variance (SAMOVA) in SAMOVA v.1.0 program was used to infer the most supported genetic structure of the sample, defining groups of populations that are geographically and genetically very similar.[30]

Genetic variance at three hierarchical subdivisions (within individuals of population, among populations within a group, and among groups of populations), was assessed by the analysis of molecular variance (AMOVA) procedure[31] as implemented in Arlequin v. 3.5. In

this analysis, studied populations were grouped by both geography and language (See Table 1)

Mantel test

The correlations and partial correlations between distance matrices of genetics-geography, genetics-language, and geography-language were performed by the Mantel test.[32] Table 2 shows the matrices of geographic and linguistic distance we used for the Mantel test, whereas genetic distance (F_{st}) matrix is shown in Table 3. Geographic distances in Km between the approximate locations of each population were computed as great-circle distances calculated from their latitudinal and longitudinal coordinates. Linguistic distances between pairs of populations were defined as simple dissimilarity indices on the basis of the hierarchical classification of languages reported in Ethnologue.[33] Populations speaking languages belonging to different subfamilies, i.e., AA and TK, were assigned dLAN of 4 while different branches within subfamilies were assigned dLAN of 3. Different sub-branches within branch were assigned dLAN of 2 and then dLAN of similar sub-branches was 1.

Inserted Table 2 here

Approximate Bayesian Computation

To deeply investigate the evolutionary relationship among populations, an Approximate Bayesian Computation (ABC) procedure was applied.[34] An ABC approach, which combines the analysis of large genetic data sets and realistic models, can be briefly summarized as follows: millions of genetic datasets with the same features as the observed one, i.e., number of individuals, type of genetic markers, length of sequences, are generated according to the coalescent theory for each demographic model, taking into account the associated prior distributions. The pattern of genetic variation in the observed and simulated data, summarized by a certain number of statistics, is then compared by Euclidean distance.

The coalescent-based simulations were performed by combinations of parameters for a specific demographic model. Those coalescent-based simulations which generated summary statistics closest to the observed ones, as shown by smallest Euclidean distances, were then considered for calculating the posterior probabilities of each model using two different approaches, acceptance-rejection procedure (AR) and weighted multinomial logistic regression (LR).[35-36] Under the AR, the posterior probability of a model is obtained by considering only a certain number of "best" simulations, and then simply counting the proportion of these retained simulations that have been generated by each model under investigation. This method can be considered reliable only when applied to a few simulations showing an excellent fit with the observed data, i.e. few hundreds.[36] Under LR procedure, a logistic regression is fitted where the model is the categorical dependent variable in the ABC simulations and the summary statistics are the predictive variables. The regression is local around the vector of observed summary statistics, and the probability of each model is finally evaluated at the point corresponding to the observed vector of summary statistics. The β coefficients of the regression model are estimated by maximum likelihood and the standard errors of the estimates might be taken as a measure of the accuracy of the method. To evaluate the stability of the models' posterior probabilities, we considered different thresholds by considering different number of retained simulations for both the model selection procedures (100, 200, 300, 500 best simulations for AR and 25000, 50000, 75000, 100000 best simulations for LR). To generate the simulated datasets we used the software package ABCtoolbox[37], running 500,000 simulations for each model. To calculate the models' posterior probabilities we used R scripts from http://code.google.com/p/popabc/source/browse/#svn%2Ftrunk%2Fscripts, modified by SG. To summarize the genetic information contained in the data we calculated the following

statistics within and between populations: the number of haplotypes (*h*), the number of private polymorphic sites (S), Tajima's D, intra- and inter-MPD, and pairwise F_{st}.

.

Testing the best-fit demographic models and type I error

Based on geographic locations and linguistic affiliations of the studied populations, SOA and LAO were variable populations whose languages differ from their geographically grouped neighbors. Therefore, three demographic models were proposed to describe different aspects of the evolutionary relationships among studied populations, in which geography or language was fitted to describe mtDNA data (Figure 2). In Model 1, the separation of the lineages follows the linguistic affiliation, with a first split (Ts1) involving the AA and the TK groups, and a subsequent separation by geographic location at Ts2 (AA and SOA), and Ts3 (TK and LAO). In Model 2, the "driving force" of the genetic variation is represented by geography. A first separation (Ts1) is started between populations from KR Basin and from SN Basin. Within each geographical group, the LAO and SOA, who speak different languages from their neighbors, were subsequently separated at Ts2 and Ts3, respectively. Model 3 extends Model 1, in which after Ts2 and Ts3, geographically closer populations (LAO-AA; SOA-TK) start to exchange migrants at a certain rate [m1 (m1_b) and m2 (m2_b)]. The effective population sizes were assumed to be constant in time; the prior distributions were all uniform (log-uniform for the effective population sizes), and, where possible, based on historical records (see supplementary Table 1).

We estimated the probability that the true null hypothesis would be rejected by evaluating the type I error. The proportion of cases in which 1,000 pseudo-observed datasets, generated under each model, is not correctly identified by the ABC analysis (both AR and LR procedures, 100 and 50000 retained simulations in turn). The power of the model choice

procedure has been evaluated using a wide range of decision probability thresholds to identify the support for a specific model, i.e. 0.5, 0.6, 0.7, 0.8, 0.9.

Inserted Figure 2 here

Results

Genetic diversity and demographic expansion

A total of 173 distinct mtDNA haplotypes were observed in 433 individuals. Among the observed haplotypes, 142 types were unique within populations, whereas the other 31 types were shared between two or more populations. Out of the 142 unique haplotypes, 34 were shared by two or more individuals within one group (multiple unique), whereas the remaining 108 haplotypes belonged to each individual (single unique). The highest number of shared haplotypes (6 haplotypes) was found between SOA-PUT, but none were shared among five pairs of populations: MON-BON, MON-LAO, MON-PUT, BON-KHM, and BON-SOA.

Genetic diversity within population and population expansion results are reported in Table 1. Haplotype diversity (*h*) varied from 0.9899 (LAO) to 0.7920 (SAK) which was in the same range as previous published populations in Thailand[13-15,19-21], albeit rather a low *h* value was found in SAK, indicating possible drift effect. The lowest nucleotide diversity (π) was observed in MON (0.0098), while PUT had the highest value (0.0153).The intra-MPD ranged from 5.5254 (MON) to 8.6956 (PUT), reflecting genetic homogeneity or recent diverged mtDNA within the MON and genetic heterogeneity in the PUT.

The highly significant negative values of the Fu's *Fs* ($P < 0.05$) were predictions of demographic expansion in KHM, MON, SOA, LAO, and PUT. The lower raggedness index (less than 0.03) as well as the unimodal mismatch distribution graph for these populations (data not shown) also provide congruent evidence for population growth and expansion.[38]

Genetic relationships

Among 45 pairwise F_{st} comparisons, 36 (80%) were statistically significant ($P< 0.01$) (Table 3). The MON, BON, and SAK showed significant F_{st} values for all comparisons, indicated high genetic differentiation. The KHM had genetically differentiated from almost all other populations, except the LAO. It is interesting that SUY and BON has genetic similarity. Most F_{st} comparisons between populations in SN Basin were not statistically significant, particularly among the SOA, PUT, and YOH, reflecting genetic homogeneity. The corrected MPD among populations showed a similar pattern to F_{st} result (see supplementary Table 2), which indicate that the SAK were most differentiated while the next most respectively differentiated populations were the BON, MON and SUY.

Inserted Table 3 here

To visualize the genetic relationship among populations, we plotted a pairwise F_{st} matrix through MDS analysis and performed SAMOVA analysis. In the MDS as shown in Figure 3, most populations residing in the SN basin (YOH, PUT, SOA and KAL) were clustered in the center of the plot with the exception of the SAK which appear to be the most genetically differentiated population, even if still genetically more closely related to neighbors in the SN Basin than to populations from the KR basin. Surprisingly, although the LAO resided in the KR Basin, they clustered together with other SN dwelling populations. For the ethnicities located in the KR basin, the KHM were quite genetically proximate to the SUY and BON, while the MON was considerably distanced from other studied populations indicating their genetic distinction. In SAMOVA analysis, when number of group was increasing from 2-groups until 6-groups category, the SAK, MON, SUY, BON, and KHM respectively were partitioned from the other populations (Table 4). The maximal percent of variation with significant value was observed at 6-groups category (7.287%, $P< 0.01$): SAK, MON, SUY, BON, KHM, YOH-PUT-SOA-KAL-LAO. Interestingly, population grouping by SAMOVA was concordant to MDS plot.

Inserted Figure 3 here

Inserted Table 4 here

Correlation among genetics, geography and language

The AMOVA was used to infer the proportion of total genetic variation accounted by groups. Groupings were defined on the basis of geographic and linguistic classification (Table 5). When populations were grouped according to geography, the results revealed that it can be used to describe the genetic structure of studied populations, since the amount of observed variation among groups was 4.68% with statistical difference (F_{ct}= 0.0468, P< 0.01), whereas the proprtion of variance among population within groups explain 7.429 % (F_{sc}= 0.07429, P< 0.01) and within populations explain 11.765% (F_{st}= 0.11765, P< 0.01). The average F_{st} of populations in the SN basin (F_{st}= 0.06902, P< 0.01) and in the KR basin (F_{st}= 0.07900, P< 0.01) were much lower than the overall F_{st} (0.09889, P< 0.01). It seems evident that there is a certain level of genetic homogeneity among populations within each geographic region, with an higher homogeneity in populations from the SN basin than in populations from the KR basin.

Based on linguistic classification, the proportion of genetic variation among groups was considerably low (0.913 %) with no statistical significance (F_{ct}= 0.00913, P> 0.01), reflecting no relationship between genetic distance and linguistic affiliation. Most of the genetic variance (89.74%) was found within populations (F_{sc}= 0.09434, P< 0.01), while variance among populations within the linguistic groups was 9.35% (F_{st}= 0.10260, P< 0.01). We observed a slight higher value of average F_{st} of AA (F_{st}= 0.10681, P< 0.01) respect to Tai speaking group (F_{st}= 0.07820, P< 0.01), possibly indicating more genetic heterogeneity among Austro-Asiatic than among Tai-Kadai groups. A notable amount of genetic variance was found among geographic groups, which is higher than variance among linguistic groups.

Mantel testing showed that genetic distances strongly correlated to geographic distances by means of correlation test (r = 0.4713, P< 0.01) and partial correlation test (r = 0.4449, P< 0.01), whereas, we detected no correlation and partial correlation between genetic and linguistic distances (r = 0.1735, P> 0.01 and r = 0.0008, P> 0.01, respectively). Among geographic and linguistic matrices, no correlation (r = 0.3667, P> 0.01) and partial correlation (r = 0.3281, P> 0.01) was observed.

Inserted Table 5 here

Model Selection

Table 6 shows the posterior probabilities of the three considered evolutionary scenarios. Model 2, in which the geography has a major role in shaping the genetic variation, received the strongest support. The posterior probability of Model 2 was never lower than 87%, considering both AR and LR and remained stable over different number of retained simulations. To assess the reliability of the probabilities estimated, we also evaluated the models' posterior probabilities within two times the range of the standard error associated to the β coefficients of the regression model (in both directions). The support remained in favor of Model 2.

Inserted Table 6 here

To evaluate whether there is enough power in the data for these models to be discriminated, we calculated a type I error, i.e., the incorrect rejection of a true null hypothesis. This analysis has fundamentally verified the reliability of the estimated probabilities because the compared models were quite similar to each other, and only a single genetic locus was analyzed. The type I error analysis considered both AR and LR as criterion of model selection, and several probability thresholds to identify the support for a specific model (Table 7). The results of the logistic regression (50,000 best simulations) were in general better than those obtained with the acceptance-rejection, especially for higher

probability thresholds. The models appeared to be well recognized even when the decision probability threshold was 0.5, since the probability of recognize the right model was never lower than 0.6. Moreover, when the right model was not selected as the "true" one, the alternative models almost never reached a probability high enough to be supported by the ABC model selection procedure. This was particularly true when the decision probability threshold was very high (0.9), i.e., similar to the value we obtained from the real data for Model 2. All together these results can be considered highly significant, and the model that has been selected here (Model 2) can be confidently regarded as the best one.

Inserted Table 7 here

Discussion

MtDNA sequences data analyzed in the current study provide us a better understand about the level of genetic variation in a micro-geographic scale and about past population dynamics in several ethnicities of the Isan region or Northeastern Thailand. Whether geography or language most influenced genetic variation of populations within regional, continental, and worldwide scales have been long-standing questions for molecular anthropologists and human geneticists.[39-43] To date, there has been no report aimed to answer the above question for populations residing in the Northeast of Thailand, addressed here for the first time. Results obtained from Mantel test, AMOVA, SAMOVA, and ABC procedures indicate that geography plays an important role to determine Northeastern Thai genetic structure, according to IBD model. Under IBD, current patterns of genetic variation would simply result from the interaction between genetic drift and dispersal of individuals between populations, thus resulting in a decrease of genetic similarities between populations when geographic distance increases [44-45]. Based on linguistic and archaeological data, each of the two different geographic regions in Isan was occupied by linguistically distinct groups of

people. The native AA populations were mainly resided in KR basin[46], while the TK people who migrated from Laos and Vietnam were centered in the SN Basin.[3]

Living in close geographical proximity, the SOA and their neighbors (PUT, SAK, KAL, and YOH), have languages of unrelated ancestry. Our genetic findings reveal the parallelism between genetic variation and geographic factors. The SOA's historical homeland is in the forested covered hills of Khammuan Province in Laos. Some of the tribe members migrated to Thai soil, in the area of Kusumal District, in 1844 A.D.[1] Based on ABC procedure (Figure2), the three demographic scenarios might be suggested to explain the degree of genetic resemblance between SOA and other SN populations, possibly linked to a recent common origin. The SOA and other neighbors within SN Basin might have shared genetic similarity from the time they resided in their historical homeland in Laos and Vietnam. At that time, they might have come into contact and after the migratory time with spatial and temporal different settlement in Thailand, their genetic homogeneity continued. The two greatest numbers of shared haplotypes between SOA-PUT and SOA-YOH, respectively, could be additionally explained by the same genetic source between the SOA and their neighbors.

In accordance with pairwise F_{st}, MPD and SAMOVA, the peculiar genetic divergence of the SAK made this population particularly interesting. Historically, the SAK originated in Vietnam and then with the influence of the Kinh (the vast majority of Vietnamese) they moved westward to Laos around 380 years ago. The majority of SAK are centered in Khammuan Province of Laos and they migrated across the Mekong river into Nakorn Panom Province of Thailand about 200 years ago.[1] The greatest differentiation as seen in the SAK is likely a consequence of genetic drift associated with female immigrants during the settlement period. The limited genetic diversity, as reflected by the lowest haplotype diversity and second lowest intra-MPD (Table 1), were regarded as reliable indicators of a genetic

bottleneck.[47] The debates on the origin of the SAK have arisen in linguistic classification. At first, the language of the SAK was classified as belonging to the AA family in the Mon-Khmer sub-family, but later most linguists classified the SAK language to the TK family in the Northern Tai branch, spoken mainly by the Tai in Gwangsi Province of China.[1,48] The SAK exhibited closest genetic relationship to the SOA. It might be indicated that the SAK are genetically more closely related to AA than to TK groups. Thus, based on several articles reporting the strong association between linguistic and genetic classifications[49-52], to our knowledge, the present-day SAK language classification is not in agreement with genetic affinity. However, it should be cautioned that the genetic ancestry of the SAK might be blurred by strong influences of the geographic factor.

Almost all AA groups, KHM, BON, SUY, and MON, as well as the only TK village of the LAO, were dispersedly situated in the KR basin. LAO or Lao Isan refers to peoples who are ethnically Lao but are Thai citizens.[4] They comprise the majority of inhabitants and are widely distributed in all provinces of Northeastern Thailand. Most of Lao Isan people were forcibly migrated from their historical homeland in the present-day Laos during 1827-1870 A.D.[3] Although the LAO village in this study was located within the area of KR Basin, close genetic affinity between the LAO and populations in the SN Basin was detected. Through previous massive migration, the LAO in the SN and KR Basins might have still preserved genetic similarity, thus, close genetic relationship might have resulted in low levels of differentiation between LAO populations in the SN and KR Basins. Future study with more broadly samples of LAO from the SN Basin will be helpful to evaluate this assumption.

Interestingly, non-significant pairwise F_{st} between LAO and KHM could be plausibly explained by extensive gene flow, concordant with an earlier genetic study[20], and socio-linguistic research.[53-55] Although current study's results support that geography explains genetic variation and relationship among populations, we somehow detect significant genetic

differentiations among populations within the KR Basin. It might be suggested that geographic proximity determined the genetic homogeneity among AA populations in the past, but later on the factors of cultural and linguistic differences as well as evolutionary factors, like drift effect, inbreeding, and genetic exchange, overcame the influence of spatial isolation, as reflected in KHM, BON, SUY, and MON.

A certain degree of inbreeding is evident particularly in the Chaobon (BON), alternately called Nyahkur. Chaobon inhabited the area that is now Thailand preceding the coming of the Khmer and the Tai groups. They now lived in Thailand only in Chaiyabhum, Petchaboon, and Nakorn Rachasima provinces. The bulk of these people live in Chaiyabhum Province, scattered among different deep jungle and mountainous villages.[1,56-57] The most original Chaobon tribe in Wang Ai Pho village in Chaiyabhum Province, who still preserved their language and culture, was sampled in this study. Loss of genetic diversity, as indicated by low values of *h*, *S* and intra-MPD, might reflect consanguineous marriage due to cultural isolation. This study has documented the sequential genetic effects from preserved cultural practice within this population before they may be possibly erased by the opportunity for admixture with Lao Isan people. Based on linguistic research, Chaobon are believed to be the remaining descendants of the ancient Mon from the historic Dvaravati period. Contrary to our expectation, the present results do not support the genetic bond between the extant BON and MON.

The Mon are one of the oldest settlers in Southeast Asia. Their origin is uncertain. It is known that they once lived in Southwest China, and moved down to upper Myanmar early in the Christian era. They were politically driven southward to settle in Pegu and Thaton, in Myanmar and eastward to the present-day Central and Southern Thailand, respectively. The great Mon Dvaravati Kingdom with an advanced civilization was founded between the 3rd and 10th century A.D. in the area of Central Thailand[1]. The prosperous Mon Kingdom

expanded to present-day Southern, Northern and Northeastern Thailand. In 1775 A.D., the first group of studied Mon migrated from Myanmar to settle down in Nakorn Rachasima, further increasing in population size to approximately 2,500 around 1793 A.D. The studied MON who historically migrated from Myanmar was indeed different from Dvaravati Mon in Central Thailand, therefore a genetic link between BON and MON was not apparent. Another important finding emerged from the results of genetic diversity and demographic expansion parameters which exhibit the lowest π, intra-MPD, and number of multiple unique haplotypes. These, as well as positive signals of population growth in the MON (Table 2), provide congruent evidence for a recent bottleneck followed by an expansion in the population, which have not yet been recognized in socio-linguistic and historic literatures.

Worthy of attention is the genetic ancestry of the Suay (SUY). MDS result reveals the close genetic relatedness between SUY and KHM, while pairwise F_{st} indicates non-significant genetic difference between SUY and BON. These results seem to be congruent with previous historic research documenting connections between SUY and KHM in language, history, society, and ancestry. The Suay or Kui, called Kamen-boran (meaning ancient Khmer) by Khmer people, are the original inhabitants of part of Thailand, Laos, and Cambodia, predating the invasion of the Khmer and the Tai group. Nowadays Suay in Thailand have been adopted a Thai-Lao language referred to as Lao-Suay or a Khmer language referred to as Khmer-Suay.[1] The current studied Suay from Surin Province migrated at first from Southern Laos during 1656-1688 A.D. and then sporadically moved until around 1760 A.D. when the mass migration period occurred.[58] However, it has been proposed by some scholars[14] that SUY share ancestry with BON, now strengthened by our investigation.

To summarize, this study highlighted some main aspects of maternal genetic structure of various populations in Northeastern Thailand. Genetic findings obtained through this

study made it possible to infer the influence of geographic factors in shaping patterns of genetic variations and affinity among linguistically diverse populations. Genetic divergence between populations was primarily influenced by geography. Then, within the same geographic location different driving forces, including language and culture as well as evolutionary driven factors, like genetic drift from founder effect, inbreeding, and admixture are considered to be the plausible additional factors. Our results remain open to future investigations with further mtDNA sequences from other populations and genetic data from different genetic markers to gain more insight into genetic history of Northeastern Thai people.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

The authors would like to thank village chiefs and all voluntary donors. We also thank Dr. Alvin Yoshinaga for English approval on this manuscript. This study was supported by Thailand Research Fund (TRF) (Grant No.MRG5580058).

References

1 Schliesinger, J. *Ethnic groups of Thailand: Non-Tai-speaking peoples* (White Lotus Press, Bangkok, Thailand, 2000)

2 Wongtaed, S. *Explore Isan Society and Cultures* (Art and culture Press, Bangkok, Thailand, 1999) (in Thai)

3 Schliesinger, J. *Tai Group of Thailand, Volume 1: Introduction and overview* (White Lotus Press, Bangkok, Thailand, 2001)

4 Bonnie Pacala, B., & Somroay, Y. *Buddhist murals of northeast Thailand: Reflection of the Isan heartland* (Silkworm Books, Chiangmai, 2010)

5 Ruhlen, M. *A Guide to the World's Languages,* Vol. 1: Classification (Stanford University Press, Stanford, CA, (1987)

6 Wright, S. Isolation by distance.*Genetics*. **28**, 114-138 (1943).

7 Slatkin, M. Isolation by distance in equilibrium and nonequilibrium populations.*Evolution*. **47**, 264-279 (1993).

8 Barbujani, G. Geographic patterns: how to identify them and why. *Hum. Biol*. **72**, 133-153 (2000).

9 Sokal, R.R. Genetic, geographic and linguistic distances in Europe. *Proc. Natl. Acad. Sci. USA*. **85 (5)**, 1722-1726 (1988).

10 Zerjal, T., Beckman, L., Beckman, G., Mikelsaar, A.V., Krumina, A., Kucinskas, V. *et al.* Geographical, Linguistic and Cultural Influences on Genetic Diversity: Y-Chromosal Distribution in Northern European Populations. *Mol. Biol. Evo1*. **8(6)**, 1077-1087 (2001).

11 Cavalli-Sforza, L.L., Minch, E. & Mountain, J.L. Coevolution of genes and languages revisited. *Proc. Natl. Acad. Sci. USA*. **89**, 5620-5624 (1992).

12 Chaubey, G., Metspalu, M, Karmin, M., Thangaraj, K., Rootsi, S., Parik, J. *et al*. Language shift by indigenous population: A model genetic study in South Asia. *Int J Hum Genet.* **8**, 41-50(2008)

13 Besaggio, D., Fuselli, S., Srikummool, M., Kampuansai, J., Castrì, 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).

14 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).

15 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).

16 Cavalli-Sforza, L.L. & Feldman M.W. The application of molecular genetic approaches to the study of human evolution. *Nat. Genet*. **33(Suppl.)**, 266-275 (2003).

17 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).

18 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).

19 Fucharoen,G., Fucharoen, S. & Horai, S. Mitochondrial DNA polymorphisms in Thailand.*J. Hum. Genet*. **46**, 115-125 (2001).

20 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).

21 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).

22 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).

23 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).

24 Librado, P. &Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. **25**, 1451-1452 (2009).

25 Nei, M. *Molecular Evolutionary Genetics* (Columbia University Press, New York, USA,1987)

26 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).

27 Harpending, H.C. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum. Biol*. **66**, 591-600 (1994).

28 Fu, Y.X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*. **147**, 915-925 (1997).

29 Tajima, F. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics*. **123**, 585-595 (1989).

30 Dupanloup, I., Schneider, S. & Excoffier, L.A simulated annealing approach to define the genetic structure of populations. *Mol. Eco*. **11**, 2571-2581 (2002).

31 Excoffier, L., Smouse, P. & Wuattro, J. Analysis of molecular variance inferred from metric distance among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*. **131**, 479-491 (1992).

32 Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**: 209-220.

33 Lewis, M.P. *Ethnologue: Languages of the World* 16th edn. (SIL International, Dallas, Texas, USA, 2009) Online version: http://www.ethnologue.com/.

34 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).

35 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).

36 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)

37 Wegmann, D., Leuenberger, C., Neuenschwander, S. & Excoffier, L. ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC bioinformatics*. **11**, 116 (2010).

38 Rogers, A.R.& Harpending, H. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol*. **9(3)**, 552-569 (1992).

39 Helgason, A., Yngvado´ttir, B., Hrafnkelsson, B., Gulcher, J. & Stefa´nsson, K. An Icelandic example of the impact of population structure on association studies. *Nat. Genet*. **37**, 90-95 (2004).

40 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).

41 Cavalli-Sforza, L.L., Menozzi, P. & Piazza A. *The history and geography of human genes* (Princeton University Press, Princeton, USA, 1994).

42 Eller, E. Population substructure and isolation by distance in three continent regions. *Am. J. Phys. Anthropol*. **108**, 147-159 (1999).

43 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).

44 Wright, S. Isolation by distance. *Genetics*. **28**, 114-138 (1943).

45 Slatkin, M. Isolation by distance in equilibrium and non equilibrium populations. *Evolution*. **47**, 264-279 (1993).

46 Premsrirat, S. Linguistic contributions to the study of the Northern Khmer language of Thailand in the last two decades. *Mon-Khmer Studies*. **27**, 129-136 (1997).

47 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).

48 Smalley, W.A. *Linguistic Diversity and National Unity: Language Ecology in Thailand* (University of Chicago Press, Chicago,USA, 1994).

49 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).

50 Cavalli-Sforza, L.L., Minch, E. & Mountain, J.L. Coevolution of genes and languagesrevisited.*Proc. Natl. Acad. Sci. USA*.**89**, 5620-5624 (1992).

51 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).

52 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).

53 Smalley, W.A. Multilingualism in the Northern Khmer population of Thailand. *Language Sci.* **10(2)**, 395-408 (1988).

54 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).

55 Talbot, S. & Janthed, C. Northeast Thailand before Angkor: Evidence from an Archaeological Excavation at the Prasat Hin Phimai. *Asia Perspectives*. **40 (2)**, 179-194 (2002).

56 Premsrirat, S.The Future of NyahKur.in*Collected papers on Southeast Asian and Pacific languages* (eds Bauer, R.S.) 155-165 (The Australian University, Canberra, Australia, 2002).

57 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).

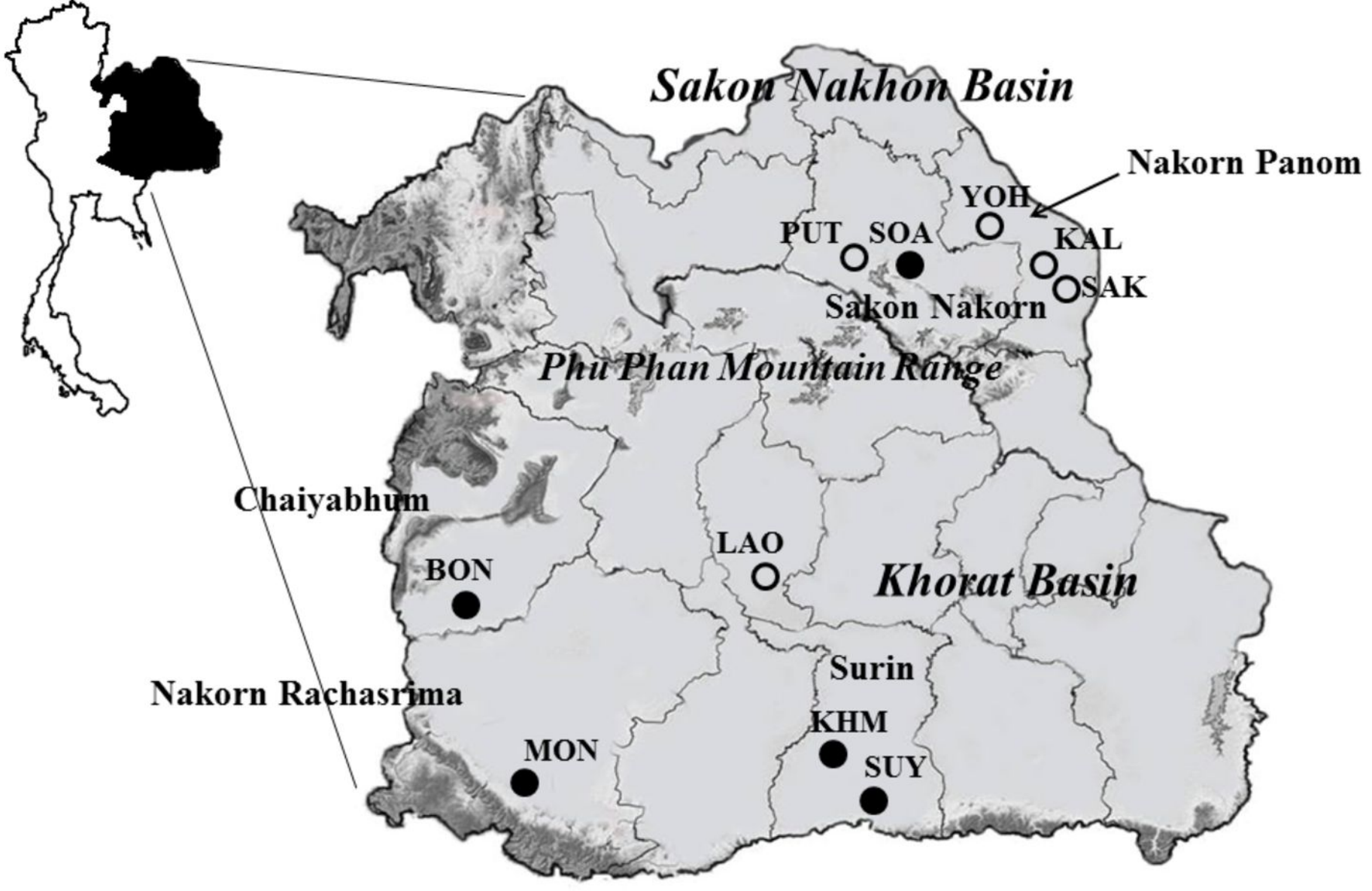
58 Sa-ard, O. *Phrase to sentence in Kuay (Surin)* (Mahidol University, Nakorn Pathom, Thailand, 1984).

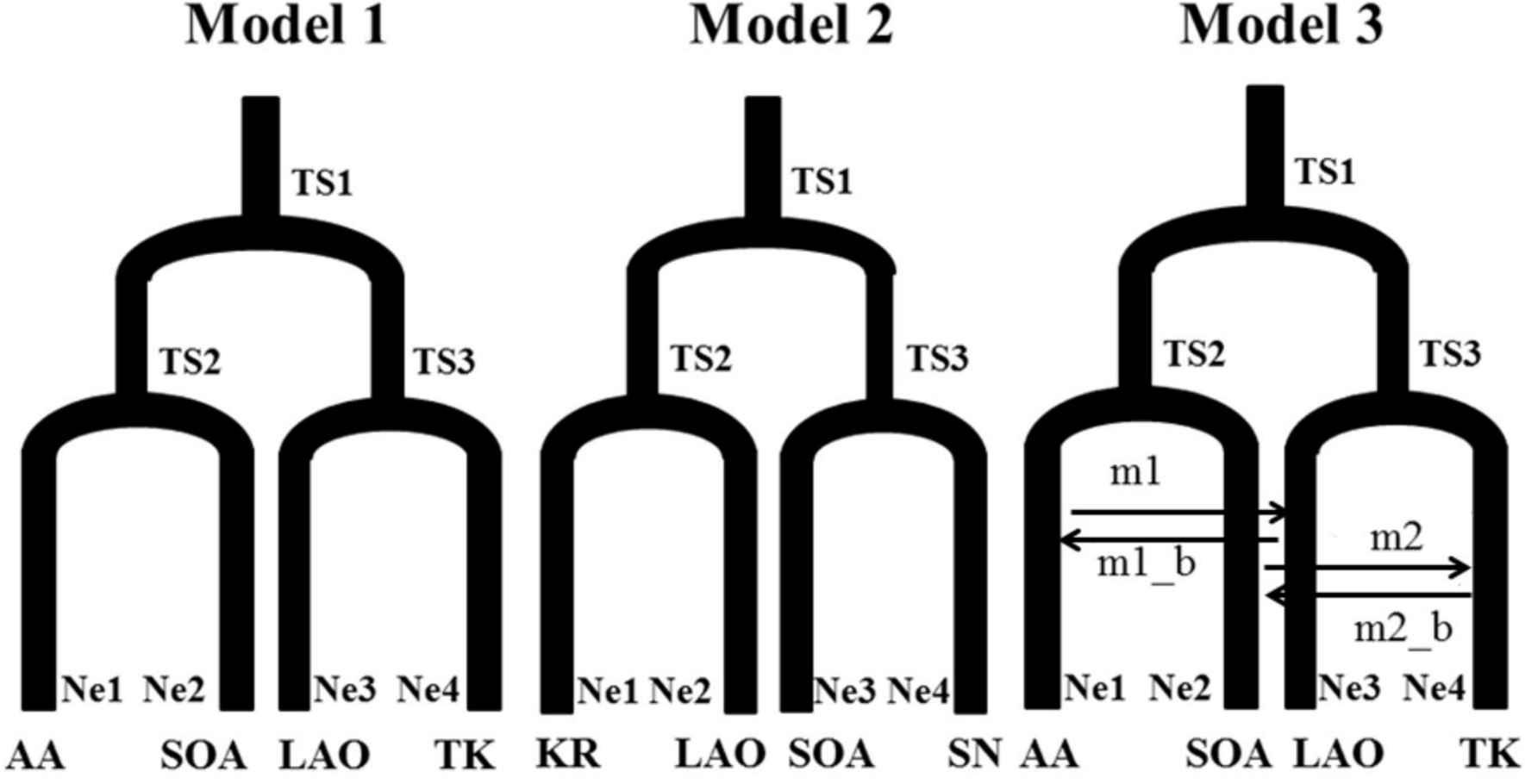
Titles and legends to figures

Fig. 1 Map of Khorat Plateau showing the locations of studied populations in different geographic areas of Northeastern Thailand. Population codes are given in Table 2. Filled circles: Austro-Asiatic linguistic family; Empty symbols: Tai-Kadai linguistic family.

Fig. 2. Schematic presentation of the three models described by language (Model 1), geography (Model 2), and recent migration (Model 3). Ne, Ts and m are the effective population sizes, separation times and the migration rates, respectively. Population codes are given in Table 2.

Fig. 3. Three dimentional scaling plot (3D-MDS) constructed based on pairwise F_{st}. Population codes are given in Table 1. Stress value for MDS = 0.0339.


Sakon Nakhon Basin
Nakorn Panom
YOH
PUT
SOA
KAL
SAK
Sakon Nakorn
Phu Phan Mountain Range
Chaiyabhum
LAO
BON
Khorat Basin
Surin
Nakorn Rachasrima
KHM
MON
SUY


Model 1
TS1
TS2
TS3
Ne1
Ne2
Ne3
Ne4
AA
SOA
LAO
TK
Model 2
TS1
TS2
TS3
Ne1
Ne2
Ne3
Ne4
KR
LAO
SOA
SN
Model 3
TS1
TS2
TS3
m1
m1_b
m2
m2_b
Ne1
Ne2
Ne3
Ne4
AA
SOA
LAO
TK

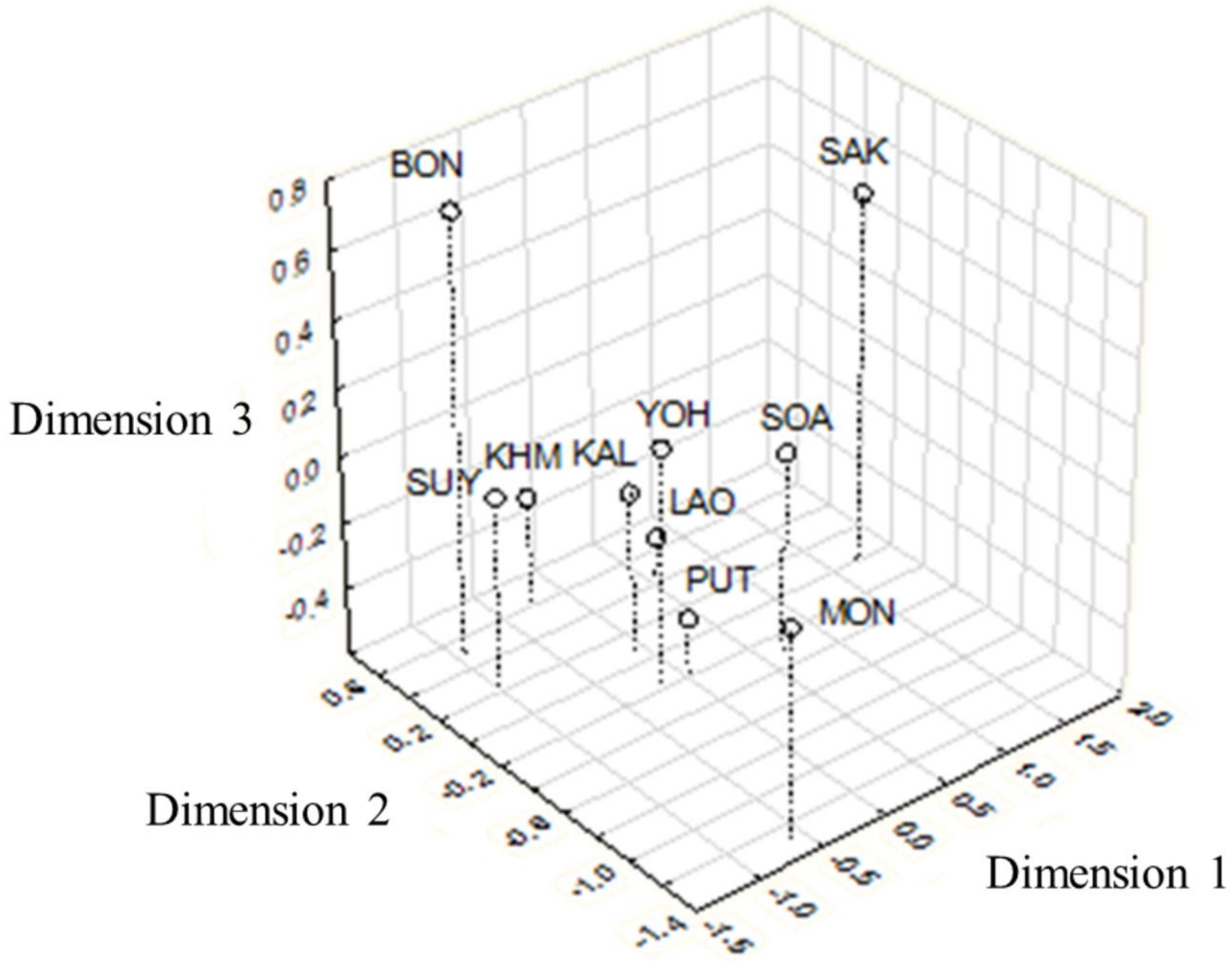

BON
SAK
YOH
SOA
KHM
KAL
SUY
LAO
PUT
MON
Dimension 3
Dimension 2
Dimension 1

Table 1. General information of studied populations and summary statistics

Populations	Khmer	Mon	Suay	Chaobon	So	Lao Isan	Phutai	Nyaw	Saek	Kaleung
Code	KHM	MON	SUY	BON	SOA	LAO	PUT	YOH	SAK	KAL
Latitude	14.90	14.69	15.01	15.59	17.37	15.62	17.28	17.55	17.45	17.33
Longtitude	103.49	102.06	103.94	101.46	104.30	103.50	103.65	104.09	104.74	104.59
Language	AA	AA	AA	AA	AA	TT	TT	TT	TT	TT
Geography	KR	KR	KR	KR	SN	KR	SN	SN	SN	SN
Location (District, Province)	Sangkla, and Chumpolburi, Surin	Pakthongchai, Nakorn Rachasrima	Sam Rong Tap, Surin	Thepsatit, Chaiyabhum	Kusuman, Sakon Nakorn	Kaset Wisai, Roiet	Waritchabhum, Sakon Nakorn	Nawa, Sakon Nakorn	Mueang, Nakorn Panom	Kuruku, Nakorn Panom
Sample size	68	44	44	42	47	35	38	41	28	46
[a]Population size	1,266,828	1,000	407,724	6,283	71,532	11,135,493	457,411	406,738	3,535	68,431
Haplotype	37	23	22	12	27	30	23	20	11	21
Unique	24	19	12	10	16	21	14	9	6	11
Single unique	20	12	11	6	12	17	10	7	4	9
Multiple unique	4	7	1	4	4	4	4	2	2	2
Non-unique	13	4	10	2	11	9	9	11	5	10
h	0.9583	0.9545	0.9397	0.8583	0.9584	0.9899	0.9573	0.9402	0.7920	0.9063
π	0.0130	0.0098	0.0143	0.0121	0.0141	0.0149	0.0153	0.0131	0.0114	0.0115
Intra MPD	7.3995	5.5254	8.1057	6.8269	8.0324	8.4924	8.6956	7.4317	6.4929	6.5266
Polymorphic site	54	40	47	23	48	54	47	39	33	35
Tajima's D (p-value)	-1.0596 (0.1370)	-1.3277 (0.0710)	-0.9632 (0.1730)	1.1116 (0.8990)	-0.9327 (0.1850)	-1.3016 (0.0780)	-0.8134 (0.2270)	-0.6458 (0.2900)	-0.8067 (0.2200)	-0.5512 (0.3170)
Fu's *Fs* (p-value)	-17.1136 (0.0000)	-8.3834 (0.0080)	-3.8913 (0.1130)	1.4256 (0.7610)	-8.3333 (0.0080)	-19.0744 (0.0000)	-5.6044 (0.0400)	-3.3709 (0.1190)	0.3691 (0.5630)	-4.3474 (0.0760)
r	0.0204	0.0195	0.0332	0.0572	0.0154	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

Table 2. Linguistic distance matrix (below diagonal) and geographic distance matrix (above diagonal) using the Mantel test. See the population abbreviation key in Table 1.

	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	

Table 3. Genetic distance based on pairwise F_{st} (below diagonal) and shared haplotype in each pairwise comparison (above diagonal). Population codes are given in Table 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**	

Bold letters indicate statistical significance at $P< 0.01$

Table 4. SAMOVA analysis. Population codes are given in Table 1.

Group category	Group of population								F_{ct}
2	SAK	KHM,MON,SUY,BON,SOA,LAO,PUT,YOH,KAL							0.1276
3	SAK	MON	KHMSUY,BON,SOA,LAO,PUT,YOH,KAL						0.0849
4	SAK	MON	SUY, BON	KHM,SOA,LAO,PUT,YOH,KAL					**0.0809**
5	SAK	MON	SUY, BON	KHM	SOA,LAO,PUT,YOH,KAL				**0.0713**
6	SAK	MON	SUY	BON	KHM	SOA,LAO,PUT,YOH,KAL			**0.0728**
7	SAK	MON	SUY	BON	KHM	SOA, KAL	LAO,PUT,YOH		**0.0693**
8	SAK	MON	SUY	BON	KHM	SOA	KAL	LAO,PUT, YOH	0.0702
9	SAK	MON	SUY	BON	KHM	SOA	KAL	YOH	LAO ,PUT	0.0664

Bold letters indicate statistical significance at $P < 0.01$

F_{ct} = Fixation index among groups

Table 5. AMOVA analysis

	No. of groups	No. of populations	% of variance			F_{st}	F_{sc}	F_{ct}
			Within populations	Among populations Within groups	Among groups			
Geography								
All samples	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**
Language								
All samples	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

Bold letters indicate statistical significance at $P< 0.01$

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

Table 6. Posterior probabilities of three population models computing by acceptance-rejection procedure (AR) and weighted multinomial logistic regression (LR) approaches.

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

Table 7. Type one error results for three best model emerging from an ABC analysis.

AR					LR				
probability threshold	probability of recognize the right model				probability threshold	probability of recognize the right model			
	Model 1 (true)	Model 2	Model 3	Not Assigned		Model 1 (true)	Model 2	Model 3	Not 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 1	Model 2 (true)	Model 3	Not Assigned		Model 1	Model 2 (true)	Model 3	Not 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 1	Model 2	Model 3 (true)	Not Assigned		Model 1	Model 2	Model 3 (true)	Not 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