



รายงานวิจัยฉบับสมบูรณ์

การตรวจหาคุณลักษณะเฉพาะระดับโมเลกุลของเชื้อไวรัสซาฟโฟลและโคซาใน
ผู้ป่วยเด็กที่มีอาการอุจจาระร่วง

Molecular detection and genetic characterization of saffold virus and
cosavirus in pediatric patients with diarrhea

โดย ดร. ภัทรา คำรินทร์ และคณะ

สิงหาคม 2558

รายงานวิจัยฉบับสมบูรณ์

การตรวจหาคุณลักษณะเฉพาะระดับโมเลกุลของเชื้อไวรัสซาฟโฟลและโคซาใน
ผู้ป่วยเด็กที่มีอาการอุจจาระร่วง

Molecular detection and genetic characterization of saffold virus and
cosavirus in pediatric patients with diarrhea

คณะผู้วิจัย

ดร. ภัทรา คำรินทร์

ภาควิชาจุลชีววิทยา คณะแพทยศาสตร์ มหาวิทยาลัยเชียงใหม่

ศาสตราจารย์ ดร. นิวัฒน์ มณีกาญจน์

ภาควิชาจุลชีววิทยา คณะแพทยศาสตร์ มหาวิทยาลัยเชียงใหม่

สนับสนุนโดยสำนักงานคณะกรรมการการอุดมศึกษา สำนักงานกองทุนสนับสนุนการวิจัย

และ มหาวิทยาลัยเชียงใหม่

Abstract

Acute gastroenteritis is one of the most common diseases in infants and young children under 5 years of age. In recent years, several viruses related to diarrheal diseases have been discovered, mostly by advanced genetic molecular techniques. The new virus species, named “Saffold virus” and “Cosavirus” were recently reported in 2007 and 2008, respectively, in children suffering from acute gastroenteritis. Thus far, it is not clear whether Saffold virus and Cosavirus are associated with the diseases, including gastroenteritis in humans and epidemiological data of both viruses are still limited. With the aim of having an overview of molecular epidemiology of Saffold virus and Cosavirus in children hospitalized with acute gastroenteritis in Chiang Mai, Thailand, therefore, the prevalence of Saffold virus and Cosavirus infections were examined by RT-PCR and the genotypes of these viruses were identified by nucleotide sequencing and bioinformatics analysis. From a total of 411 fecal specimens collected, Cosavirus was found in a single specimen (CMH-N199-11) (0.2%), while Saffold virus was detected in three cases (CMH-S052-10, CMH-N031-11, CMH-N120-11) (0.7%). Genotype identification by nucleotide sequencing revealed that the Cosavirus CMH-N199-11 strain was most similar to the Cosavirus SH1 Chinese strain and belonged to HCoV-A6 genotype. For Saffold virus genotype identification, the CMH-N120-11 was identified as SAFV-1, while other 2 strains of CMH-S052-10 and CMH-N031-11 belonged to SAFV-6. It was observed that all three Saffold virus strains isolated in this study were most closely related to the Saffold virus reference strains isolated previously in Pakistan. In addition, the complete genome sequence analysis demonstrated clearly that the viral genotypic characterization should be based on the P1 region analysis. Our data demonstrated that Saffold virus and Cosavirus are the viral agents that associated with acute gastroenteritis in Thai children, although the low rates of detection were observed. To our knowledge, the

present study is the first report of Cosavirus and Saffold virus genotype 6 in pediatric patients admitted to hospital with acute gastroenteritis in Thailand.

Key words : Saffold virus, Cosavirus, Acute gastroenteritis, Diarrhea, Thailand

บทคัดย่อ

โรคกระเพาะอาหารและลำไส้อักเสบแบบเฉียบพลันหรือโรคอุจจาระร่วง เป็นโรคที่พบได้บ่อยในเด็กทารกและเด็กเล็กที่มีอายุต่ำกว่า 5 ปี จากการนำเอาความรู้และเทคโนโลยีทางด้านอนุชีววิทยา มาประยุกต์ใช้ในการพัฒนาวิธีการตรวจวินิจฉัยสาเหตุของการเกิดโรคอุจจาระร่วง ทำให้ทราบว่า เชื้อไวรัสเป็นสาเหตุสำคัญของโรคอุจจาระร่วง และเมื่อไม่นานมานี้ในปี ค.ศ. 2007 และ 2008 ได้มีรายงานการค้นพบไวรัสชนิดใหม่สองชนิด จากผู้ป่วยที่มีอาการอุจจาระร่วง คือ “Saffold virus” และ “Cosavirus” จากการสืบค้นในฐานข้อมูลพบว่า มีข้อมูลการศึกษาที่เกี่ยวข้องกับไวรัสทั้งสองชนิดนี้อยู่ค่อนข้างน้อย เนื่องจากเป็นไวรัสชนิดใหม่ที่เพิ่งมีการรายงานการค้นพบเมื่อไม่นานมานี้ ดังนั้น การศึกษานี้จึงทำการตรวจหาความชุกและวิเคราะห์คุณลักษณะทางพันธุกรรมของไวรัสทั้งสองชนิดนี้ในผู้ป่วยเด็กที่เข้ารับการรักษาในโรงพยาบาลด้วยอาการอุจจาระร่วงในจังหวัดเชียงใหม่ด้วยวิธี RT-PCR, nucleotide sequencing, และ bioinformtic analysis ซึ่งพบว่า จากตัวอย่างอุจจาระที่ทำการศึกษาทั้งสิ้น 411 ตัวอย่าง ตรวจพบ Cosavirus อยู่ 1 ตัวอย่าง (CMH-N199-11) คิดเป็นร้อยละ 0.2 และตรวจพบ Saffold virus อยู่ 3 ตัวอย่าง (CMH-S052-10, CMH-N031-11, CMH-N120-11) คิดเป็นร้อยละ 0.7 ของตัวอย่างทั้งหมดที่ทำการวิเคราะห์ในการศึกษาในครั้งนี้ สำหรับการจัดจำแนก genotype ของ Cosavirus พบว่า CMH-N199-11 จัดอยู่ในกลุ่มของ HCoV-A6 genotype และมีความใกล้เคียงกับ Cosavirus สายพันธุ์ SH1 ที่แยกได้ก่อนหน้านี้จากประเทศจีน ส่วน Saffold virus ที่ตรวจพบอยู่ 3 ตัวอย่าง พบว่ามีอยู่ 1 ตัวอย่างคือ CMH-N120-11 เป็น SAFV-1 ส่วนอีก 2 ตัวอย่างคือ CMH-S052-10 และ CMH-N032-11 เป็น SAFV-6 โดยพบว่าเชื้อทั้ง 3 ตัวอย่างนี้เป็นเชื้อที่มีความสัมพันธ์ใกล้ชิดกับ Saffold virus ที่แยกได้ก่อนหน้านี้จากประเทศปากีสถาน และจากการวิเคราะห์เปรียบเทียบลำดับ nucleotide ของตัวเชื้อ พบว่าลำดับ nucleotide ในส่วนของโปรตีนโครงสร้างโดยเฉพาะในส่วนของ P1 region เป็น region ที่เหมาะสมในการใช้จัดจำแนก genotype ของตัวเชื้อ ซึ่งผลที่ได้แสดงให้เห็นว่าเชื้อไวรัสในกลุ่มของ Cosavirus และ Saffold virus สามารถตรวจพบได้ในเด็กที่มีอาการอุจจาระร่วงในประเทศไทยถึงแม้จะตรวจพบได้อัตราที่ค่อนข้างต่ำก็ตาม รายงานการวิจัยในครั้งนี้ถือเป็นรายงานแรกของการตรวจพบ Cosavirus และ Saffold virus genotype 6 ในผู้ป่วยเด็กที่มีอาการอุจจาระร่วงในประเทศไทย

Executive Summary

ชื่อโครงการ (ภาษาไทย) : การตรวจหาคุณลักษณะเฉพาะระดับโมเลกุลของเชื้อไวรัสซาฟโพลและ
โคซาในผู้ป่วยเด็กที่มีอาการอุจจาระร่วง

(ภาษาอังกฤษ) : Molecular detection and genetic characterization of saffold
virus and cosavirus in pediatric patients with diarrhea

ชื่อหัวหน้าโครงการ (ภาษาไทย) : ดร. ปัทรา คำรินทร์

(ภาษาอังกฤษ) : Dr. Pattara Khamrin

(ตำแหน่งทางวิชาการ) : อาจารย์

ชื่อนักวิจัยที่ปรึกษา (ภาษาไทย) : ศาสตราจารย์ ดร. นิวัฒน์ มณีกาญจน์

(ภาษาอังกฤษ) : Professor Niwat Maneeekarn

(ตำแหน่งวิชาการ) : ศาสตราจารย์

สถานที่ติดต่อ :

ภาควิชาจุลชีววิทยา คณะแพทยศาสตร์ มหาวิทยาลัยเชียงใหม่

โทรศัพท์ 053 945332 ต่อ 108

โทรสาร 053 217144

ผลงานการตีพิมพ์ในวารสารวิชาการนานาชาติ : Khamrin P, Maneeekarn N. Detection and genetic
characterization of cosavirus in a pediatric patient with diarrhea. Arch Virol. 2014. 159: 2485-
2489. (Impact factor 2.03)

สรุปความสำคัญ ที่มา และวัตถุประสงค์ของปัญหาวิจัย

โรคอุจจาระร่วงหรือโรค Acute gastroenteritis เป็นการอักเสบของกระเพาะอาหารและลำไส้แบบ
เฉียบพลัน เมื่อไม่นานมานี้ ได้มีความก้าวหน้าในการพัฒนาวิธีการตรวจวินิจฉัยเพื่อหาสาเหตุของเชื้อไวรัสที่

เกี่ยวข้องกับการเกิดโรคอุจจาระร่วงโดยใช้เทคนิคที่มีชื่อว่า metagenomic analysis ทำให้ทราบว่ายังมีไวรัสใหม่ๆ อีกหลายชนิดที่เกี่ยวข้องกับการเกิดโรคอุจจาระร่วงในผู้ป่วย ดังนั้นการศึกษาไวรัสกลุ่มใหม่ที่เพิ่งมีการค้นพบนี้จึงมีความสำคัญและเป็นหัวข้อที่น่าสนใจที่จำเป็นต้องมีการทำวิจัยต่อไป

ในปี ค.ศ. 2007 และ 2008 ได้มีรายงานการค้นพบไวรัสชนิดใหม่สองตัวจากผู้ป่วยที่มีอาการอุจจาระร่วง คือ “Saffold virus” และ “Cosavirus” จากการสืบค้นในฐานข้อมูล PubMed พบว่ามีข้อมูลการศึกษาที่เกี่ยวข้องกับไวรัสทั้งสองชนิดนี้อยู่ค่อนข้างน้อยเนื่องจากเป็นไวรัสชนิดใหม่ที่เพิ่งมีรายงานการค้นพบ ดังนั้นผู้วิจัยและคณะจึงมีความสนใจที่จะทำการศึกษาไวรัสทั้งสองชนิดนี้ โดยวัตถุประสงค์ของการวิจัยในครั้งนี้คือ ตรวจสอบ Saffold virus และ Cosavirus และ จำแนก genotypes ของ Saffold virus และ Cosavirus จากอุจจาระของผู้ป่วยเด็กที่เข้ารับการรักษาในโรงพยาบาลในจังหวัดเชียงใหม่ เพื่อศึกษาความสัมพันธ์ของเชื้อไวรัสกับการก่อโรคอุจจาระร่วงในผู้ป่วย โดยวิธีการวิจัยที่ใช้ในการศึกษาในครั้งนี้คือ RT-PCR, nucleotide sequencing, และ bioinformatics analysis

ผลการศึกษาวิจัย

จากตัวอย่างอุจจาระที่ทำการศึกษาทั้งสิ้น 411 ตัวอย่าง ตรวจพบ Cosavirus อยู่ 1 ตัวอย่าง (CMH-N199-11) คิดเป็นร้อยละ 0.2 ตรวจพบ Saffold virus อยู่ 3 ตัวอย่าง (CMH-S052-10, CMH-N031-11, CMH-N120-11) คิดเป็นร้อยละ 0.7 ของตัวอย่างทั้งหมดที่ทำการศึกษาวิเคราะห์ในการศึกษาในครั้งนี้ สำหรับการจัดจำแนก genotype ของ Cosavirus โดยการศึกษาลำดับ nucleotide sequence พบว่า CMH-N199-11 มีลำดับ nucleotide ใกล้เคียงกับ Cosavirus species A, genotype A6 (HCoV-A6) สายพันธุ์ SH1 ที่แยกได้ก่อนหน้านี้จากประเทศจีน ส่วน Saffold virus ที่ตรวจพบอยู่ 3 ตัวอย่าง พบว่ามีอยู่ 1 ตัวอย่างคือ CMH-N120-11 เป็น Saffold virus genotype 1 (SAFV-1) ส่วนอีก 2 ตัวอย่างคือ CMH-S052-10 และ CMH-N032-11 เป็น SAFV-6 โดยพบว่าเชื้อทั้ง 3 ตัวอย่างนี้เป็นเชื้อที่มีความสัมพันธ์ใกล้ชิดกับ Saffold virus ที่แยกได้ก่อนหน้านี้จากประเทศปากีสถาน และจากการวิเคราะห์เปรียบเทียบลำดับ nucleotide ของตัวเชื้อพบว่าลำดับ nucleotide ในส่วนของโปรตีนโครงสร้างโดยเฉพาะในส่วนของ P1 region เป็น region ที่เหมาะสมในการใช้จัดจำแนก genotype ของตัวเชื้อ ซึ่งผลที่ได้แสดงให้เห็นว่าเชื้อไวรัสในกลุ่มของ Cosavirus และ Saffold virus ที่พบในประเทศไทยน่าจะมีต้นกำเนิดทางวิวัฒนาการร่วมกับเชื้อที่ระบาดใน

ประเทศอื่นๆ ในทวีปเอเชีย รายงานการวิจัยในครั้งนี้ถือเป็นรายงานครั้งแรกของการตรวจพบ Cosavirus และ Saffold virus genotype 6 ในผู้ป่วยเด็กที่มีอาการอุจจาระร่วงในประเทศไทย

ประโยชน์ที่ได้รับ

1. ทราบความชุกของการติดเชื้อ Cosavirus และ Saffold virus ที่ทำให้เกิดโรคอุจจาระร่วงในผู้ป่วยเด็กที่เข้ารับการรักษาในโรงพยาบาลในจังหวัดเชียงใหม่ ประเทศไทย
2. เข้าใจถึงคุณลักษณะพื้นฐานทางพันธุกรรมและวิวัฒนาการของ Cosavirus และ Saffold virus โดยพบว่าการศึกษาเพื่อจัดจำแนกชนิดของ genotype ของไวรัสเหล่านี้ควรทำการศึกษาโดยการวิเคราะห์ความแตกต่างของลำดับ nucleotide ในส่วนของโปรตีนโครงสร้างโดยเฉพาะในส่วนของ P1 region หรือ full-length sequence
3. หลังจากเสร็จงานวิจัยโครงการนี้ ผลการศึกษาที่ได้สามารถตีพิมพ์ในวารสารวิชาการระดับนานาชาติ เพื่อเผยแพร่ข้อมูลงานวิจัยที่เกี่ยวข้องกับ Cosavirus และ Saffold virus ที่ได้จากการศึกษาในประเทศไทย

เนื้อหางานวิจัยฉบับสมบูรณ์

1. ความสำคัญและที่มาของปัญหาการวิจัย

โรคอุจจาระร่วงหรือโรคกระเพาะอาหารและลำไส้อักเสบแบบเฉียบพลัน (Acute gastroenteritis) เป็นภาวะการอักเสบของกระเพาะอาหารและลำไส้ที่เกิดขึ้นแบบเฉียบพลัน โดยสาเหตุอาจเกิดจากการติดเชื้อไวรัส แบคทีเรีย หรือโปรโตซัวก็ได้ จากการนำเอาความรู้และเทคโนโลยีทางด้านอนุชีววิทยา (Molecular biology) มาประยุกต์ใช้ในการพัฒนาวิธีการตรวจวินิจฉัยสาเหตุของการเกิดโรคอุจจาระร่วง ทำให้ทราบว่า เชื้อไวรัสเป็นสาเหตุสำคัญของโรคอุจจาระร่วง โดยมีไวรัสหลายชนิดที่เป็นสาเหตุได้แก่ Rotavirus, Norovirus, Sapovirus, Astrovirus, Adenovirus, Enterovirus, Parechovirus, Aichi virus, และ Bocavirus เป็นต้น (Chaimongkol et al., 2013)

เมื่อไม่นานมานี้ ได้มีความก้าวหน้าในการพัฒนาวิธีการตรวจวินิจฉัย เพื่อหาสาเหตุของเชื้อไวรัสที่เกี่ยวข้องกับการเกิดโรคอุจจาระร่วงโดยใช้เทคนิคที่มีชื่อว่า metagenomic analysis ทำให้ทราบว่ายังมีไวรัสใหม่ๆ อีกหลายชนิดที่เกี่ยวข้องกับการเกิดโรคอุจจาระร่วงในผู้ป่วยนอกเหนือจาก Rotavirus, Norovirus, Sapovirus, Astrovirus, Adenovirus, Enterovirus, Parechovirus, Aichi virus, และ Bocavirus ซึ่งการศึกษาไวรัสชนิดใหม่ๆ เหล่านี้ยังมีข้อมูลที่ค่อนข้างจำกัดในแง่ของระบาดวิทยาและกลไกการก่อโรคในมนุษย์ ดังนั้นการศึกษาไวรัสกลุ่มใหม่ที่เพิ่งมีการค้นพบนี้จึงมีความสำคัญและเป็นหัวข้อที่น่าสนใจที่จำเป็นต้องมีการศึกษาและทำวิจัยต่อไป

ในปี ค.ศ. 2007 และ 2008 ได้มีรายงานการค้นพบไวรัสชนิดใหม่สองชนิด จากผู้ป่วยที่มีอาการอุจจาระร่วง คือ “Saffold virus” และ “Cosavirus” โดย Saffold virus แยกได้ครั้งแรกจากอุจจาระของเด็กที่เป็นไข้โดยไม่ทราบสาเหตุ (fever of unknown origin) ส่วน Cosavirus พบครั้งแรกในอุจจาระของเด็กที่มีอาการอัมพาตที่ไม่เกี่ยวข้องกับการติดเชื้อ Poliovirus (non-polio acute flaccid paralysis) (Jones et al., 2007; Kapoor et al., 2008) จากการสืบค้นในฐานข้อมูล PubMed พบว่ามีข้อมูลการศึกษาที่เกี่ยวข้องกับไวรัสทั้งสองชนิดนี้อยู่ค่อนข้างน้อยเนื่องจากเป็นไวรัสชนิดใหม่ที่เพิ่งมีการรายงานการค้นพบเมื่อไม่นานมานี้ สำหรับในประเทศไทย จากการสืบค้นข้อมูลที่ผ่านมา มีเพียงการศึกษาเดียวที่ได้

รายงานการตรวจพบ Cosavirus ในผู้ป่วยผู้ใหญ่ที่มีอาการอุจจาระร่วงในจังหวัดเชียงใหม่ (Khamrin et al., 2012) โดยพบว่าจากผู้ป่วย 150 ราย พบการติดเชื้อ Cosavirus ในผู้ป่วยเพียง 1 รายเท่านั้น และพบว่าเป็น Cosavirus species C (HCoSV-C) สำหรับ Saffold virus จากการศึกษที่ผ่านมาพบว่าเด็กที่มีอาการอุจจาระร่วงในจังหวัดเชียงใหม่ในช่วงปี ค.ศ. 2007 และ 2012-2013 พบว่าร้อยละ 2.7 และ 1.5 เกิดการติดเชื้อ Saffold virus ตามลำดับ โดยเชื้อที่พบระบาดในจังหวัดเชียงใหม่ คือ genotype 1, 2, 3, 4 (SAFV-1 ถึง SAFV-4) (Khamrin et al., 2011; Yodmeeklin et al., 2015) ดังนั้นผู้วิจัยและคณะจึงมีความสนใจที่จะทำการศึกษาและวิเคราะห์คุณลักษณะทางพันธุกรรมของไวรัสทั้งสองชนิดนี้ในผู้ป่วยเด็กที่เข้ารับการรักษาในโรงพยาบาลด้วยโรคอุจจาระร่วงในจังหวัดเชียงใหม่

2. วัตถุประสงค์ของโครงการวิจัย

- 2.1 ตรวจหา Saffold virus และ Cosavirus จากอุจจาระของผู้ป่วยเด็กที่เข้ารับการรักษาในโรงพยาบาลในจังหวัดเชียงใหม่ เพื่อศึกษาความสัมพันธ์ของเชื้อไวรัสกับการก่อโรคอุจจาระร่วงในผู้ป่วย
- 2.2 จำแนก genotypes ของ Saffold virus และ Cosavirus เพื่อให้ทราบคุณลักษณะทางพันธุกรรมของเชื้อไวรัสทั้งสองตัวนี้ที่มีการระบาดในเขตจังหวัดเชียงใหม่

3. ระเบียบวิธีวิจัย

- ทำการเก็บตัวอย่างอุจจาระจากผู้ป่วยเด็กที่มีอาการอุจจาระร่วงที่เข้ารับการรักษาในโรงพยาบาลในจังหวัดเชียงใหม่จำนวนทั้งสิ้น 411 ราย
- โดยตัวอย่างทั้งหมดที่เก็บได้ ถูกนำไปศึกษาต่อเพื่อตรวจสอบหาสารพันธุกรรมของเชื้อ Saffold virus และ Cosavirus ด้วยวิธี RT-nested PCR ซึ่งมีขั้นตอนดังนี้
 1. สกัดสารพันธุกรรมของเชื้อไวรัสจากตัวอย่างอุจจาระ
 2. สร้างสาย complementary DNA ด้วยวิธี Reverse Transcription Reaction โดยใช้ Random primer

3. นำ cDNA ที่ได้มาทำการตรวจว่าเป็น cDNA ของเชื้อ Saffold virus หรือไม่ ด้วยวิธี Nested PCR
4. นำ cDNA ที่ได้มาทำการตรวจว่าเป็น cDNA ของเชื้อ Cosavirus หรือไม่ ด้วยวิธี Nested PCR
5. ตัวอย่างที่ให้ผลบวกต่อเชื้อ Saffold virus และ Cosavirus จะถูกนำไปทำ DNA sequencing เพื่อทำการวิเคราะห์ full-genome sequence ของตัวเชื้อ (Saffold virus มีขนาดประมาณ 7.8 kb, Cosavirus มีขนาดประมาณ 7.3 kb) เพื่อศึกษา genetic background ของเชื้อไวรัสที่พบต่อไป
6. นำ full-genome sequence ที่ได้ไปวิเคราะห์ด้วยวิธี Bioinformatic analysis

4. ผลการวิจัยและสรุปผลการวิจัย

1) การตรวจหา Cosavirus

การตรวจหา Cosavirus ในตัวอย่างเด็กที่มีอาการอุจจาระร่วงในจังหวัดเชียงใหม่

จากตัวอย่างอุจจาระที่ทำการศึกษาทั้งสิ้น 411 ตัวอย่าง ตรวจพบ Cosavirus อยู่ 1 ตัวอย่าง (strain CMH-N199-11) คิดเป็นร้อยละ 0.2 ของตัวอย่างทั้งหมดที่ทำการวิเคราะห์ในการศึกษาในครั้งนี้ สำหรับการจัดจำแนก genotype ของ Cosavirus ในเบื้องต้นโดยการศึกษาลำดับ nucleotide sequence ในส่วน 5' untranslated region (5' UTR) พบว่าเชื้อไวรัสสายพันธุ์นี้ มีความใกล้เคียงกับเชื้อ Cosavirus species A (HCoSV-A) และมีลำดับ nucleotide sequence ใกล้เคียงกับ Cosavirus virus สายพันธุ์ SH1 ที่แยกได้จากประเทศจีน (Dai et al., 2010) ดังนั้นผู้วิจัยจึงได้ทำการวิเคราะห์เชื้อไวรัสที่พบเพิ่มเติม โดยการศึกษา full-genome sequence analysis

การวิเคราะห์ full-genome sequence

Cosavirus species A (HCoSV-A) ที่พบ 1 ตัวอย่าง (CMH-N199-11) ได้มีการ submit full-genome sequence ขึ้นทะเบียนไว้ในฐานข้อมูล NCBI (The National Center for

Biotechnology Information) โดยมี accession number คือ KJ396940 ซึ่งมีรายละเอียดและลำดับ full-genome sequence ดังนี้

> CMH-N199-11 (HCoSV-A)

Length = 7,317 base pairs

G+C content = 43.87%

A+T content = 56.13%

Nucleotide	Number	Mol%
A	2001	27.35
C	1652	22.58
G	1558	21.29
T	2106	28.78

> CMH-N199-11

```

GGCTATGTCAAAGCAATACAGTGTGTGCAAGGCGTACACGCTCAAGGCGGAGACTTAGGCC
TCACAGATTGTGTTTTGTGTTATTGGATGCTGGATGGTCACGTTGGAGACTGCATGTGGC
AGTCTTGAAACGTGTGGTTGACGTCTATCCATTACGGCAGTGGGTGGAGTACTGCAAAG
ATGTCACCGTGCTTTACACGGTTTTGAACCCACACCGGCTGTTTGGCGCTTGCAGGAC
AGCAGGTTTATTTCTTATGTTCTTTATTTCTAGCCAGCAGGGTTCTATCCTGTTGGGCG
GAGTGATACTCCCCTTCTTCTGGACAGATTGCCTCCACGACCTTTGTGGATCTCAAGG
TGATCAAGTCACTGGTGAATAGAGCGAAGGTTGAGGAGACCTGAGGAATTTCCATGTGGC
TTTGCCAGGAGTTGTAGCGATGCTGTGTGTGTGCGGATTTCCCTCATGGCAACATGA
GCCTCACAGGCCGAAAGCCCTGTCCGAAAGGACCCACACAGTGGAGCAACCCAGCTCCC
TCCTACAAAGCTTTGTGAGAATGAACTCACGTTTATTCTTCTTTATTCTCTATTTACATC
AGGCCCCAAAGATGTCCTGAAGGTACCTTGTGTATCTGGGCATGAGCACCATCAACTACC
CGGACTTGCACTTTGGTGCAGACGCATGTGGTGACCCAGCCCCTCTGCTTCGGCAGAGGG
GCTTTCGCTCGCTCAGCACGAGATCTGATCAGGAGCCCCTCCCAGTGTGCTTTACACCTG
GCGGGGGGTTAAAAATTGCCAAGGCCTGGCACAATAACCTAGGGGACTAGGTTTACCTT
TTGTTAACAATGTCTGTCTATTATGGGTGCAAACAACAGCAAAGAATCAGTGTCCAGCAAT
GGCAACGAGGGAACAATTGTTAATAACTTTTTATTCAAACCAGTATTATGCTTCTATTGAT
GCTTCTGCCAAGGTGTTGGGACCTCTACTACTCCTGAAAACGGCAACGTATCTGGCTTT
CTTGGACTTGCAAGCAGTGCCTTTAATGCTCTCTCTTTTCTCGCCTCCCCGCGAGTAGAG
AACTCAACATATCAGGAAGATAGGATTTATCGCGCTAGCAGGCAATACATCTATAAAC
TCTCAAGCTTCAGAAGGGGTTTTGTGTGCTTACGCGCATGAAACCGACGGTTCAGATCCT
ACTTCTTGTGGTGACACCCCAAGGAACTGCAGCAACAGATAGAAGTTTCATAGTA
CAACTCCAGCAATGGTCCACAACACAGGCCGCATATGATGCTCAGTGGGTCCGCTCACA
CAAGAACTGCGGGCGGATAAACAAGGTAATGTGTTTGCAAAAAATCTGAAAACACTCACTCA
TATGCAAAGATGGGATTCGAGGTAACACTTCAGGTCAACACCTCTCCGTTCCATTGTGGT
TTGGTTGGACTGTTTCTGTCCCCGAATTCACCAGATTCGGGCCAGACAACCTTGACATGG
CAGGATGTAACAACCCGACTTTCTTGTGAACAATACTGCTTTATATGACCCCCAAAAT

```

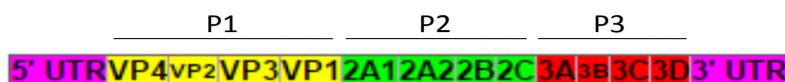
TATGCAGGAGATAGGGCTTTAGACACTGATGGCTCTTTTGATATGGCGGATTTTTGCCT
GAACAATTCATGCTGTTTCCCACCAGTTTATAAATCCCAAAGACACTAATATTGCTACT
GTTAGAGTCCCATATGTTAATATAGCCCCTACTAGTGATACCACTGTTTCATACAGTTTGG
ACAGCGGTGGTAATGGTGATAGTTCCCCTCTACTATTCATCAGGTGCTTCTCCAACAGTT
AATATGACAATGACTATTACTCCTGTGAATTCAGTGTTAATGGACTGCGTCATTGCGCG
CAGACCCCTATTCTACTAGGCCTTTTACAACCTTTAACCAGTTCAACACCACGGTGCCT
CTTCGCACCGAGCCATGCTATGGTATGACTGTGACCCCTCCAGTTGATTACATGCCTCCA
CCAGTTGAAGATCTCATGTCTTTGGCCCAGGTGCCCTTGCTTCTGTGCGTTGCCTTTTCT
GACAATAGTATTAAGGGTTTTCTTACTTTGCTGTTAGCAATACGGAACAAGGGAGAAAA
TTGTTTTCTAGTGGTGTAGTGTATCAGATCCCCACTATCAACACACTTTGCTTTCAAAT
TTGGCTCGCTTCTTCTGCAATTATCGCGGAAGCATACAATTTGATTTTGTGCGCCACA
ACTGCCCTGACCCGTGGTAAGTTGCTTATAGCTTATTCTCCTCCAGGTGCCAATGAGCCC
GCATCTATTGACCAGGCTATGATGGGGACTTATGCAATTTGGGATCTTGGACTGCAGTCA
ACTTTCAATTTTGTGTTCCTTTCATATCGGCAACTGACTTTCGCTACAATCAAGTAGCA
ATCGCTCAGCTTTGAATTCGTATGGATGGCTTACAGTTTGGATGCTTAACCCATTGACT
TACCCACCCAATACACCCCCAACCCAACAGATTGTCATGATGCTTTCTGCAGGTAAGGAC
TTCTCTTTCAGACTTCCAATTTACCAGGTATAGCACAAAGGGATTGGCGAACAAGACCCA
CATGACAATATGGAGTGTGGACAGGTTGATGACAAAGATGCGTCCATGTTCTCCGGACAT
TCTGTTTCTCTCCCCACGCCTCACACACTGTGTTCCTTTTCTATGACAGATACCGCATG
ATTGGAATTGTGCGTTCAACAAACAGAAATTCACCAAACTTAAGACACCTCTGGAAGCA
GGCAAAATAATGAATTTGGTCAAATATTTCAAGATTGAATCGGCATATACACCCCAACCA
GAACATGCTATCGGTCTAACTCCAGTACCTAACTTGAATGGAGTCCCTTGTCGGCTTTC
CTTAATGCCAAATCTACTGCATCCAAGAAAAGTTAAGACTTGTAAAGGAGACTTTTCAG
TTGAATTATTGCTGTCTTTTACTTATGCTAAGTTTGATCTTGAATTTACAGTGGTGCCC
CCTGCTAGCTTTGCATCTGATTATATAGTTCACTGGTATCCTCCAGGTGCCCTATTGAT
ACTGCTGAGATCATGTTTGAATGAATGCTACTGACACTGAGTTTCAAGACAATGGAGAA
AACATGAGTTCAGCTCAACTCTCATATAACCCCTACTTTTTACGCCAGAGGACCTACTAAA
GTCTGTGCAGTAGTACCTTTTTGTTTACCTACTTCTCTATTGCCTCTTTATTTTGTGGT
TATCCAGACTATGCAAAAACCTGGCTTGTATGGGACTTCTCCAGCTTCTACCTTTGGT
ACTATTTTTGTTGAATCAGACACAGAAGCAGCTTCCATGTTTACATCAGGTACAAAAAT
TTCAAGGGTTACCTACCTAGACCACTAATTAGAACAACACACATACCCCTGGCAAGTCGC
AACAGAAGACAGGTGGCGGGTAGTGTGCTTACTAGAGAGATTACTAGGGAAGAACGAGAA
GTTGCCAGGCAACTCCTCCTTATTTTCAAGGTGATATTGAGTCCAACCCAGGTCCTGCATTT
AATCCAGAATACACTGTACATGGCCCAGTAACAGATTTGATTCAATTGGCTAGAAAACCA
GAAACAGTTGATAATGTAACAGACTTCTGACTACTCTTAACACCCCTTATGGCTAAATGG
AACAACCTTAAAGAGACTGTTACAGATGCTGTGTTTCTTAGAGACATGGTGTGCTTCTT
GTAACCTTACCTCTTACTCTACCTGGTTACGACGACAAGGACCAGGTGCTTACTTTGCT
GCTGCTTCTATATTGATGGCTGATGGGATATCTTTCTTTGATTGGTATGAAAAGATTA
ATCTTCATGGCAAGAAAACCTCAGAGTTTCCCCCTTTCTTTCCAGTGCCCAGGGACCA
GATCTTAGAGATTTTGTGACTTTCTTCAACGCTGCGCGTGGAGCGCAGTGGATGATTGAT
TCTCTTAAATCCCTCATATCTTGGATTAACAGTGGCTTGAACCTGAGGAAGAAAATGAA
GCAGTACAACCTGAAAAGATGTTAATTGAATCTCCAGACACTGCAAGGCAATAAATGAT
TACAACAGAGGTGACTCTTTCCAACGACCGACTAACTCTTTTGAATTCATGGACAGACTT
GTGGAATGTGCTACTAACTCGGGAGGGTTCAAATTGCAACATACTTTAGAAATTTACC
ACAGCCGATTCTGATACTAGCAGACCAGAGCCGGTGGTGTGTTCTGCGCGGGAAACCA
GGCGCAGGCAAATCAGCTGCAGCTACTGTCATCGCTGCTGCAGTCTCAAATTTGTTAGTA
GGAAGTCAATCAGTTTATACACTTTCTCCAGATACAGAACACATGGATGGATATCATGGT
CAATTTGTGACCTTGTGGATGACCTTGGACAAAACCCAGACGGTGAAGACTTCAGATGT
TTCTGTCAAATGGTTTTCATGTGCTCAGTACCGCCCCGCCATGGCTGATCTTAAAGACAAG
GGAATCCTGTTTACTTCTAGACTTTTGTGTTGCTACCACCAATTTACCAGACTTCAATCCC

GTTACTATTTTCAGATCCTCGCGCCCTTGACAGGCGCATTACCTTCGATATCCACGTTACA
CCAGGCTCTGCTGCAACCAAGAATGGGAACTCGACTTGGCTGCGGCCCTCAAACCCGAT
GGACCCGGTGAACACCCCTTACACTTCTGACTGTCCCATTCTCCACACTACTGGACTTCTT
TTGAAAAACCTCAGAAACAACCAGACTATGAATGTAAAGGACCTTGTTGACCTTA
TCGTGAGAAGAATTAACATAAGAAAGAAGTAGGAAACATGCTTACTCTCTAGTTGCC
AGGGACCTACAATGATTGTTGGCTATACTAAGGATGATGATGGCATTGCGATTGTGGACT
GTCTCGAAGAATGGAACAAGATTAAGGACAAGAAAAAGAAACAGCTTGCTTTGGAAATGG
TTGCCAGGAGCTTAAAGACAAACATGAAGAACATAAGGGAACAATTAACCTTTTAAAA
TGTTTGTACAGGTTTGGGTGTCGTCGCTGCCGTTGCAGGAGCGTACGCCACAATGAAAT
ATTTCACTAAAGACAAACCAGAGGAAACCAAGGAACCAGAGGAAAAAGAAATGAAGAAAGAG
AAGAGTCCAAAGAAGCTGCTGGTCCGTACAATGGGCCAACAAAGAAAGAAATTAACAT
TGAAGCTCAAGGCCAGAGTCCTTATGGATATGAAAAAGAAATCTCTCAGAATGTTA
TGCCCTTTCAGATTTGGTACAACGGAAAGAAGTATACCCAGTCTTGTGGCAATAGGTA
AAAGAGTGATTCTTGTAAACAAACATGCTTTTGAATCAGTGGATAACAAATTCATTGTGA
ATCAGAAAGAATACACACTAGATCAGGTTACAGCTATTTCCCTTGATTGTGGGTCAGGTG
TTACAGATGTGTGTGCTATTTGCTTGCCCCAGGCCAGATTTCAAATCAATAAAGAAAC
ACTTCTACCTTTTAACTACTATGTTTCCAGGCACCCGCCTGACTATTCTCTCGAATG
ACCACTACCCTATGTCCAGAGAAGGGTCTTTCCCTTCGCTTTGAGGATGAAGTTCCACCA
ACATAGGTCAGATGCCCTTTGTGATGCTTTACAAATCAACCTTACTTTGGTATGTGTG
GTTCAGTTGTGTAGTAGATTTGTTGATGGAGGAGGAATAATTGGAATGCACTGTGCAG
GTGGAGGCGGAGTCACTGTTGGAACCCGTTTACTGCTAGAATGATTGAATCAGTTTTTG
ATTATTTCTACCCACAGTTGCTCAGGGAATTATTGAAAATACAGAGACAGGACCCCGTG
TGCATGTGCCAGACAATCCAAATTAAGAAGAACTCATGCCACTTATCCGGCAACGGACA
AGTTTGGCCAGCTGCACTCTCGCGGTACGATCCGCGTCTTAATGAAGGAGTAGACCTTG
ATACAGTGATCTTTTCAAACACACCCATAATACTTTGATTGATAAAGAATCCACTTTTA
GGAAAGCTCTTGATATGGCAGCAGAAATTTATGGTGAAGTTTCAGAGGAAATGATTTCT
CCCCTCTTTCAGTTGAAGATGCAATCTTGGAGTACCCGGACTCGACCGTTTGGACCCCA
ACACAGCTTCAGGACTTCTTACACAAAGACCAGAAGACAGATGATTGACTTCAA
CACAGGACAGATTTTGGACGATACTCTTAAGTGTGACTTGGACAATGGCTTGCTGGACG
ACCTCTCAGGAAGTACATTACCAGACATTTCTCAAGGATGAAATCAGGCCCATCGAAAA
GGTCAAAGCAGGAAAAACCAGAATAATTGATGTTCCCTCTTGTATCATGTCATTGCCTT
CAGAATGCTCTTTGGCAAATTCATTGCCACTACCACTTGAACCTTTGGCTTTAAACAGG
CTCTGCTATTGGTTGTGACCCAGATGTTGCTTGGGCTTCTTTCGGCTTTGAGCTCAGTGG
CTTTCCTTATCTGTATGATTTTGAATTAACCTTTGATGCTTCTCACAGCACTTCAAT
CTTTGAAATCTTAGATCAGAAATCTTTTCTCCAGAATTAGGTTTTGATCCTAGATGCTC
ACTTCTTGAATCCCTTGCACTTCAACCCACTGTTATGAGAACAAGAGACTCCAGAT
TGCTGGAGGGCTTCCCTCCGGCACGGCAGGTACCTCAGTACTGAACACCGTGATAAACAA
TATCATTTTTACGGTGCATTGTATCACACTTACACTAATTTTGAAGTGGATGACATCAG
TATGTTAGCCTACGGCGATGACATTGTTGTTGCCTCCAAATTCGAACTTACTTGGTTAT
GGTTAAGGCTTTCATGAACCGGATTGGCTACAAGATTACCCCTGCAGACAAAAGTGACGA
ATTCAGACCAAAGTGATGGATGACATTTGCTTCTTAAAGAGGCGTTTTGTTAAAGTTGC
TGGAGTTTGGGCTCCAGTATGGAACCTGAAAACCTCGAGGCAATGTTGTCTTGGTACAA
ACCAGGAACTTAAATGAAAAGCTCCAGAGTGTCTCAAGACTTGGCCATTTCTCAGGACG
TGACGTATATGACCGCTTTTCAAGCCCTTCAATCGTGATGGGTTTACGTCACACCTTG
GAAACAGTTACACTTGAATGGCTTAATAAGTTATCAGCTTAAAGAATTTGAATTGGCA
TTTCAAATTTATTTGAATTTGGCTTTAATTCGTTCTAATTTGGTTATTTATTGGGTA
TAATCA

ตารางที่ 1 การเปรียบเทียบ full-genome sequence ระหว่าง CMH-N199-11 กับ Cosavirus species ต่างๆ

Region*	Nucleotide sequence identity (%)						
	HCoSV-A6	HCoSV-A1	HCoSV-A19	HCoSV-A20	HCoSV-E1	HCoSV-F1	HCoSV-D6/E2
	SH1	0553	PK6187	NG263	Australia/81	PK5006	NG385
VP4	97.0	97.0	96.0	95.0	80.3	77.9	78.9
VP2	90.2	63.6	66.5	70.4	56.0	65.2	58.3
VP3	89.0	67.3	65.2	68.9	59.7	60.4	62.2
VP1	85.6	59.3	57.7	57.8	46.4	50.7	48.8
2A	82.2	60.0	59.3	60.4	47.7	37.2	42.2
2B	82.9	84.8	81.8	84.5	64.7	66.1	64.4
2C	84.4	85.8	85.8	86.8	69.5	64.7	70.7
3A	82.0	80.8	80.8	88.0	60.1	61.8	59.3
3B	85.9	87.7	84.2	84.2	77.1	75.4	80.7
3C	86.0	87.5	88.5	88.0	60.5	59.1	61.9
3D	91.6	92.6	90.6	92.6	67.8	66.8	67.5
Complete-sequence	87.3	77.2	77.2	78.6	60.6	61.8	61.4

* Gene mapping ของ Cosavirus (adapted from <http://www.picornaviridae.com>)

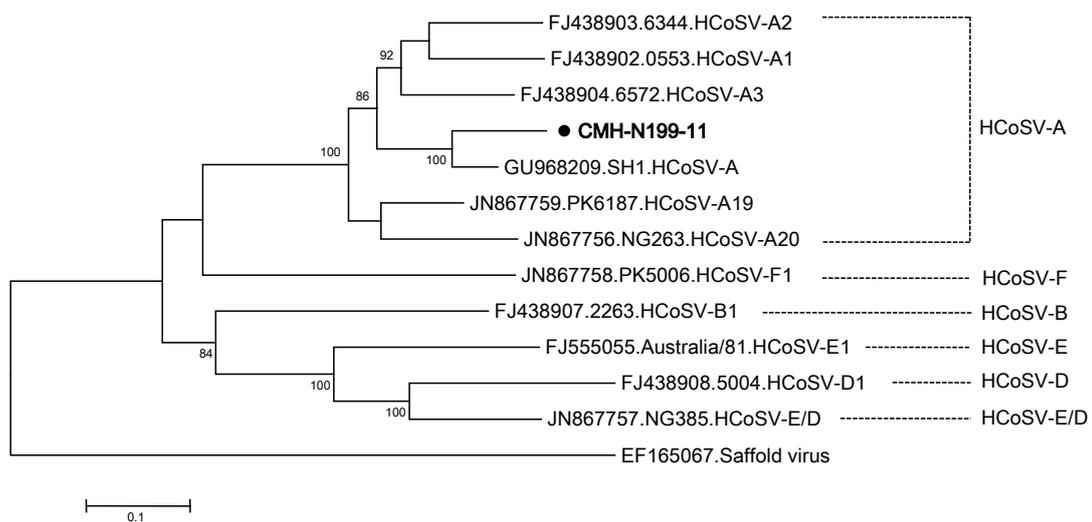


จากการวิเคราะห์เปรียบเทียบ complete coding sequence ในส่วนของ 11 gene regions (VP4, VP2, VP3, VP1, 2A, 2B, 2C, 3A, 3B, 3C, 3D) ของ CMH-N199-11 กับ Cosavirus species ต่างๆ พบว่า CMH-N199-11 มีลำดับ nucleotide ใกล้เคียงกับ Cosavirus species A, genotype A6 (HCoSV-A6) สายพันธุ์ SH1 ที่แยกได้ก่อนหน้านี้จากประเทศจีน โดยมีความเหมือนกันของลำดับ nucleotide อยู่ที่ร้อยละ 87.3 ในขณะที่เมื่อเปรียบเทียบกับ Cosavirus species A แต่เป็น genotype อื่น พบว่ามีความเหมือนกันของลำดับ nucleotide sequence ลดลงมาอยู่ที่ร้อยละ 77.2 ถึง 78.6 และพบว่าความเหมือนของลำดับ nucleotide

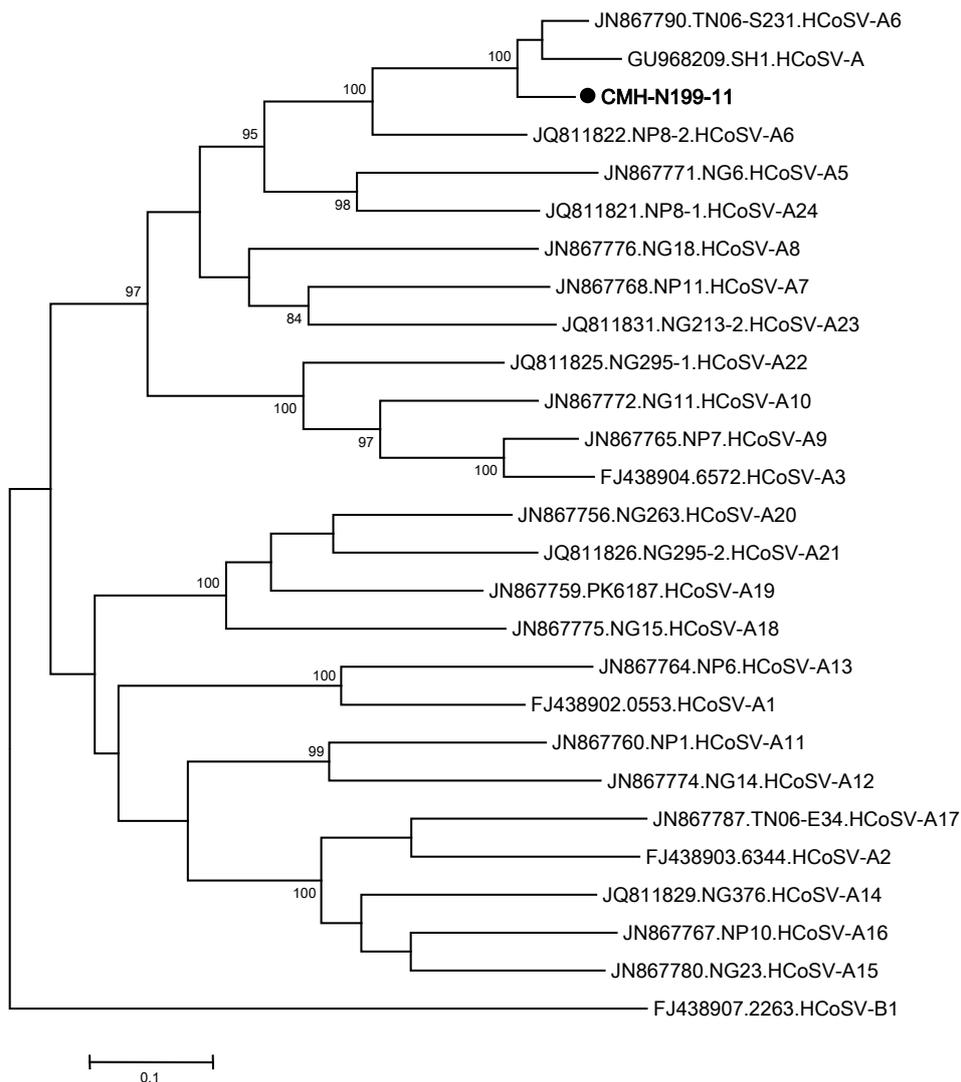
ยิ่งลดลงเมื่อเทียบกับเชื้อใน species อื่นๆ คือพบอยู่ที่ร้อยละ 60.6 ถึง 61.8 ซึ่งผลการเปรียบเทียบ ลำดับ nucleotide ที่ได้จากทั้ง full-genome sequence ช่วยยืนยันผลการ typing ในเบื้องต้นว่า CMH-N199-11 จัดจำแนกได้เป็น Cosavirus ใน species A, genotype 6 (HCoSV-A6) และ เมื่อทำการเปรียบเทียบลำดับ nucleotide ในส่วนของ genome ในแต่ละ region จาก VP4 ถึง 3D พบว่า ลำดับ nucleotide ในส่วนของ VP4 มีความ conserve มากที่สุด และ ลำดับ nucleotide ในส่วนของ VP1 และ 2A มีความแตกต่างกันในระหว่าง Cosavirus ที่ต่าง species กัน (ตารางที่ 1)

รูปภาพที่ 1 การศึกษาวิวัฒนาการของ Cosavirus CMH-N199-11 เปรียบเทียบกับ Cosavirus สายพันธุ์ อื่นๆ โดย (A) แสดงการวิเคราะห์ในส่วน of complete coding sequence และ (B) แสดงการ วิเคราะห์ในส่วน of VP1 capsid region

(A)



(B)



ผลการศึกษาวិวัฒนาการของ Cosavirus CMH-N199-11 เปรียบเทียบกับ Cosavirus สายพันธุ์อื่นๆ ด้วยวิธี Phylogenetic analysis โดยการวิเคราะห์ลำดับ nucleotide ในส่วนของ complete coding sequence ดังแสดงในภาพ A แสดงให้เห็นว่า Cosavirus CMH-N199-11 ที่แยกได้จากการศึกษาในครั้งนี้จัดกลุ่มรวมกับเชื้อมาตรฐานที่เป็น Cosavirus species A และมีความสัมพันธ์ใกล้ชิดกับ Cosavirus สายพันธุ์ SH1 ที่แยกได้ก่อนหน้านี้จากประเทศจีนมากที่สุด

สำหรับการวิเคราะห์ลำดับ nucleotide ในส่วนของ VP1 ซึ่งกำหนดการสร้าง capsid protein ของตัวเชื้อไวรัส และเป็น region ที่นิยมใช้ในการวิเคราะห์เพื่อจัดจำแนก genotype ของ

เชื้อในกลุ่มนี้ พบว่าเมื่อทำการวิเคราะห์ Cosavirus CMH-N199-11 เทียบกับเชื้อมาตรฐาน genotype A1 ถึง A24 พบว่า เชื้อที่แยกได้จากการศึกษาในครั้งนี้มีความสัมพันธ์ใกล้ชิดกับ Cosavirus genotype A6 สายพันธุ์ SH1, TN06-S231, และ NP8-2 ซึ่งให้ผลตรงกับการวิเคราะห์ โดยใช้ complete coding sequence ดังแสดงในภาพ A ข้างต้น

โดยสรุป ผลที่ได้จากการตรวจหา Cosavirus ในศึกษานี้พบว่า ไวรัสชนิดนี้พบได้อัตราที่ค่อนข้างต่ำในเด็กที่มีอาการอุจจาระร่วง จากผลการวิเคราะห์คุณลักษณะทางพันธุกรรมของตัวเชื้อ พบว่าเชื้อที่แยกได้ในประเทศไทยมีวิวัฒนาการใกล้เคียงกับเชื้อที่พบในประเทศจีน จึงเกิดข้อสันนิษฐานว่า Cosavirus ที่พบในประเทศไทยและประเทศจีนน่าจะมีต้นกำเนิดทางวิวัฒนาการร่วมกันมาก่อน โดยเชื้อที่แยกได้สามารถจัดจำแนกอยู่ใน species A, genotype 6 (HCoV-A6) ซึ่งแตกต่างจากเชื้อที่เคยแยกได้ในผู้ป่วยผู้ใหญ่ ซึ่งพบว่าเป็น Cosavirus species C รายงานการค้นพบ Cosavirus ในครั้งนี้ถือเป็นรายงานแรกของการตรวจพบ Cosavirus ในผู้ป่วยเด็กที่มีอาการอุจจาระร่วงในประเทศไทย

2) การตรวจหา Saffold virus

การตรวจหา Saffold virus ในตัวอย่างเด็กที่มีอาการอุจจาระร่วงในจังหวัดเชียงใหม่

จากตัวอย่างอุจจาระที่ทำการศึกษาทั้งสิ้น 411 ตัวอย่าง ตรวจพบ Saffold virus อยู่ 3 ตัวอย่าง (strains CMH-S052-10, CMH-N031-11, CMH-N120-11) คิดเป็นร้อยละ 0.7 โดยทั้ง 3 ตัวอย่างได้นำไปศึกษาต่อเพื่อทำการ identify genotype ด้วยการวิเคราะห์ partial sequence ของ VP1 gene ซึ่งพบว่ามีอยู่ 1 ตัวอย่างคือ CMH-N120-11 เป็น Saffold virus genotype 1 (SAFV-1) ส่วนอีก 2 ตัวอย่างคือ CMH-S052-10 และ CMH-N032-11 เป็น SAFV-6 ดังนั้นผู้วิจัยจึงเลือกตัวแทน ของ SAFV-1 และ SAFV-6 อย่างละ 1 ตัวอย่างมาทำการวิเคราะห์เพิ่มเติมโดยการศึกษา full-genome sequence analysis

การวิเคราะห์ full-genome sequence

ตัวแทนของ Saffold virus SAFV-1 และ SAFV-6 ที่ได้ทำการวิเคราะห์หาลำดับ nucleotide ของ full-genome sequence ดังนี้

> CMH-N120-11 (SAFV-1)

Length = 7,833 base pairs

G+C content = 43.10%

A+T content = 56.90%

Nucleotide	Number	Mol%
A	2249	28.71
C	1794	22.90
G	1582	20.20
T	2208	28.19

> CMH-N120-11

```

CTAATCAGAGGAAAGTCAGCATTTTCCGGCCAGGCTAAGAGAAACATATAAGTTAGATT
TGATCTAAATCCCTTGACGAATTCGAAACGAGATGTTCTCCCTCCCTTGCCGCTTGTTCA
CACCCATTTATCTAATTCGGCCTTCCGTAACAAGCCCCTCGGTGAAAGAACCTCTCTCTT
TTCGACGTGGTTGGAATTAACATCATTTCCGACGAAAAGTGCTATCATGCCTCCCCGATTA
TGTAATGTTTTCTGCCCTGCTGGGCGGAACATTCTCGGGTTGAGAAACCTTGATCTCTTT
CCTTTGGAACCTTGGTTCCCCCGGTCTAAGCCGCTTGGAAATATGACAGGGTTATTTTCTT
TATCTTATTTCTACTTTCACGGGTTCTATCCGTGAAAAGGGTACGTGTTGCCCTTCCTT
CTCCGGAGAATTCACACGGCGGTCTTCCGTCTCTCGACAAGTGTGAATGCAACATGCCG
GAAACGGTGAAGAAAGCAGTTTTCTGCGGAAGTCTAGAATGCGCATCGAAACAGCTGTAG
CGACCTCACAGTAGCAGCGGACTCCCCTCTTGGCGACAAGAGCCTCTGCGGCCAAAAGCC
CCGTGGATAAGATCCACTGCTGTGAGCGGTGCAACCCCAGCACCCCTGATTGATGAACAT
TCTTTACGGAACCAGAGGATGGCTTTCTTAAGCCCTCCGGTAGAGAAGCCAAGAATGTCC
TGAAGGTACCCCGCGTGCGGGATCTGATCAGGAGACCAATTGACTGTGCTCTACACCGTC
ACTTTGGTTTTAAAACTGTCACAGCTTCTCCAAACCAAGTGGTCTTGGTTTTCTAATTTT
ATTGACTGGCAATATGGCGTGCAAACACGGATAACCCGTTTTTGTGCCCTTTTGTACTGTC
CATTGACATCTCTTCGGATGGATCTTTTGTCTTCTATATGATGGCGAATGGTATCCAAC
TGATCTCCTGACTGTTGACTTGGACGACGATGTGTTTCATCCCCTGGATTCTATCATGGA
ATGGACTGATCTACCATTGATTCAAGATGTGTTGATGGAACCACAAGGCAACTCCAATTC
TTCAGACAAAAACAATTCACAATCTTCTGGGAATGAAGGTGTCATTATCAACAATTATTA
CTCCAACCAATACCAGAATTCAATTGACCTTTCGCCAATGCCAACGGTGTAGGCAAGGA
AAATTCTAAGCCACAAGGTCAATTAATGAATATTCTTGAAGCGCGGCTGACGCCTTCAA
AAACATTGCCCCGCTTCTCATGGACCAAGACACTGAAGAAATGACCAACCTTTCTGACAG
AGTGTATCAGACACTGCAGGCAATACGGCCACAAACACTCAGTCTACAGTGGGTCGTTT
GTTTGCATTTGGCTCGAGACACAAGGGAAAAACACCCAGCGTCTTGTGCGGACACCGCCAC

```

TGACAAAGTGCTCGCTGCGGAGAGATACTACACTCTCAAATTGGTTACTTGGACTATGAC
ACAAAAACCTTTTGACCAGTTTCGCATACCTCTCCACACGTGCTAGCAGGCAAACAGGG
TGGTGTTCGCGGCTACTCTCCGGAGACACTATCTTTGCAAATGCGGGTGGCGCATACA
GGTTCAGTGCAATGCTTCTCAATTTTCATGCAGGCTCTTTGTTAGTGTTCATGGCCCCAGA
ATTTGACACTTCTTCATATTCTGCAACCAAAGATCCCACCGGAGAAGAGCCCTTCAAGGT
TGACACCACGTGGCAGACAGGTCTCTTCGTGGTCACTCCTACGAAGACAAATCCACTCA
AACCTTCGACCATTGGCACTTAACCATCAGAACCATTGGCAATGGACAGTTTACCCTCA
TCAAATTTGAATTTGAGAACGAACACAACATGTGATCTGGAGGTACCATATGTAATAT
TTGCCAACCTTCATCTTGGACGCAGCATGCAAATTGGACCATCGTGGTGGCCGTCCTTTC
TCCTTTACAGTACACTACAGGCTCGGCCACAGAAGTGGACATTACTCTATCAATACAACC
AGTGAAACCAGTGTTCATGGTTTAAGACATTCAGTTTTAGAAAGCAGACAGTCCAATACC
AGTAAGTGTAGGGAGCATGCAGGCACCTTCTTTCTACTACCCAGATAATACTGTTC
AGTCTACGGAAAAACCGTAGCAACACCATCTGATTACATGTGTGGAGAATTTACAGACCT
GTTGTCAATTGTGCAAAATTTCAACTTTCTTGGTAATGTCTCTTCAACACAAAGGCTTCC
TTACTTTTCAGCAACAAACACAGTTCCGACTAACGCTTTAGTGACATATCAGGTCTCTCT
GTCATGTTCTTGCATGGCCAATTCATGCTGGCATCAGTTGCTAGGAATTTCAATCAATA
CAGGGGTTCTTAAATTACTTGTGTTTGTTCACAGGTAGCGCCATGACAAAAGGAAAATT
CCTTATTTCTACACACCACCTGGTGCAGGGAAACCAACAACACTAGAGACCAAGCTATGCA
AGCTACTTATGCCATTTGGGATTTAGGATTAACCTCTAGTTATAATTTCACTGTTCCCTT
CATCTCTCCAACACATTACAGACAACTTCTTACACAACCTCCATCTATTACTTCTGTAGA
TGGGTGGCTTTCTGTTTGGCAACTCACGCCTTACTTACCCTGTAGGCTGTCTTCTAA
TTCTGACATCCTCACACTCGTGTGCGGGTGGGGATGACTTCACTCTAAGGATGCCGATTTCC
TCCAACCAAGTACACCCCGCAAGGTGTGGACAATGCTGAGAAGGGAAAGGTCTCAGACGA
CAACGCATCAGCTGACTTTGTGCTGTTCCGATCCAACCTCCCGAAAACCAAAGTGAAGT
TTCTTTCTTTATGACAGAGCAGTTCTTGTTCGGCTTTGGCAGTGGGAACAAACCAGGG
CTTGAACAATTCCTTCTCTTTGGAATCTGCTTCCGGGAGTTATGAACAACATTCCTTTGAC
ACCATTACCCAGTGACAGGCTCAAAGAAAACGAGTTCGGTTTAGACGAACAACACAGATG
GTTGAGTTTCCAATCAGCAACTAGTTCAACTCCCCCTTATAGAACCAACAAGACTGGAA
TTTCTTGATGTTCTCTCTTTCACTTACTACAAGTGTGATCTGGAAGTGACCCTTTCTCT
CAACTCAGGAAATGCAGCTGGGGCTGTAGTTAGGTACGTACCTAGCGGTGCTCCGGTAGA
CATAACAACCCAAACAATGCCTTTCACTCCTTCTTTCGCGGACACTCGAGACCCACACAT
GTGGTTGGTAGGTGCTGGATCTACAAACCAAATTTCTTTGACTGCCCTACACCTCCCC
ATTATCTGTCTGCCCGCTGTGTGGTTAACGGGTCTCTGACTTTTCCAATACTTCACG
CTTTGGCGTTGCTCCAACCTCAGACTTTGGAAGATTATTTGTTCAAGGATCTGATAGATT
TTCCGTGCATTTCAAGTACAAGAAAATGCGAGTCTTTTGTCCAAGACCCACAGTTTTCTT
TCCTTGGCCTTCGAGTCTGTTTCAAAAATTCATGCCAACACACCAGTCCAAATCTTGGAA
ACTTCAAGACCTATTTCAATGTATAGAGTAGATCTTTTCATAAACTTTTTCAGATGAAAT
CATTCAAGTTCACCTTACAAAGTTCACGGTAGAACAGTTTGCCAATATGAAATCCCTGGCTT
TGGTCTATCTCGCTCTGGACGTCTTCTTGTGTGCATGGGCGAGAAGCCCTGCCAACTCCC
CATTTCCACTCCCAAGTGTTCACACATCGTTTTACCGGTTCCAGAAATCTTTTGG
TGTGTCCATTTACAAAGCAAGGTACAGACCTTGGAAACAGCCTCTTTCATGATGAACTTTA
CGACTATGGTTTCTCCACTTTCAGTACTTCTTCAAAGCTGTAGGGACTATCATGCCTC
TTATTATAAGCAAAGACTTCAGCATGATATTGAAACAAACCCAGGACCCGTCAGTCGGT
TTTTCAGTTGCAAGGTGGGGTGTAACTAAATCCCAAGCACCCATGTCGGGATTACAGAG
TATGTTACTGAGAGCAATTGGGATTGAAGCTGATTGTACAGAGTTCACCAGAGCAGTTAA
TTTGATCACAGACTTGTGCAACACCTGGGAGAGCGCTAAAACAACACTGTCTTCTCCCGA
ATTCTGGACCAAAATGGTGTGAGAATTGTCAAATGGTCGCTGCTTCAGTTCTTTATCT
ACACAATCCAGATTTGACCACCACTGTTTGTATCTTTGATGGCCGGAATTGACATTCT
CACTAACGACAGTGTTTTAACTGGTTGTCTCCAAACTTTCTAAATTTTTCCACACTCC
CGCCCCCCCCATTGTCCCACTTCTCCAACAACAGTCTCTATTAGGGAAGCAAATGACAA

TTTCAACTTGGCTAAAAACATAGAATGGGCAATAAAAAACAATAAAAAGAATAGTTGAATG
GATTACTTCATGGTTCAAACAAGAAGAGACTAGTCCTCAGGCTAACTAGACAAAATGTT
AGAAGACTTTCCAGAACACTGTGCCTCCATTCTAGCTATGAGAAATGGGAGAAAAAGCATA
CACTGATTGTACTGGTGCATTTAAGTACTTCGAACAACCTTTACAATTTGGCAGTACAATG
TAAAAGAATACCACTAGCTACTTTGTGTGAAAAATTTAAGAACAACACGACCATGCTGT
GGCCAGGCCAGAACCTGTGGTTGTTGTGTTGCGAGGAAATGCTGGACAAGGAAAAATCCGT
GACCAGTCAAATCATCGCTCAAGCGTTTTCCAACTTTTCATTTGGCCGTCAATCAGTTTA
CTCTCTTCCACCAGATTCTGATTATTTGGATGGTTATGAGAATCAATACTCAGTGATCAT
GGATGATTTAGGACAAAATCCTGACGGAGAAGATTTTAAAGTGTCTGTCAAATGGTTTC
AAGCACAAAATTTCTTCCAAATATGGCCCACTTGAAAAGAAAGGAACCCCTTTTACTTC
AAACTTCATAATAGCAACAACCAACCTTCCCAAGTTTCGACCCGTACCGTAGCTCACTA
CCCAGCTGTTGATAGAAGAATCACCTTTGATCTGACCGTTGAGGCAGGAAATGAATGTGT
TACTCACAATGGAATGTTGAATGTAGAGAAGGCTTTTGAAGAAATTCAGGAAAACCACA
ATTGGATTGTTTCAACACTGATTGCAGACTACTCCATAAGAGAGGGGTGCGGTTTGTATG
TAACCGCACAAAGAATGTCTACAACCTTCAACAAGTAGTAAAAATGGTGAAGAACCACAT
TGATAACAAGGTGAAAAATCTAAAGAAAATGAATACATTGGTTGCCAATCACCTGGTAA
TGACATGGACTATGTGTTGACTTGCTGAGACAAACAATGCTGCATTACAAGATCAAAT
TGATGAACTCCAAGAGGCCCTTAAATCAGGCTCAAGAAAGACAGAATTTCTATCAGATTG
GCTTAAAGTTTCTGCAATTGTTTTCGCATCTATTGCTTCTTTGTCTGCAGTTTGTAAT
GGTTTCCAGATTCAAAAATTTGGTGTGTCCGGCTCCTGTCCAAATACAATTATCCGAAGG
AGAACAAGCTGCTTACTCTGGAGGCAAGAAAGGAGAGAAGCAAACACTCCAGGTTTTGGA
CGTACAAGGTGGAGGAAAGATCGTAGCCCAGGCTGGGAACCCGGTCATGGACTATGAGGT
GAATATCGTAAGAACATGGTCACCCCTATTACTTTCTTTATGCAGACAAGGCACAAGT
CACACAATCCTGCTTGTGTTGTCAAAGGGCACCTTTTTGTGGTCAATAGACATGTGGCTGA
AACTGACTGGTGCCTTTTGAACCTCAAAGGCACACGGCATGAGAGAGATTTCGGTACAGAT
GAGATCAGTGAATAAATCTGGGATGGAAGTTGATTTGACTTTTGTGAAAGTTATTAAGG
ACCTCTTTTCAAAGACAATTCGAAGAAATTTTGTCCAAGAATGATGACTTTCCGGCTAG
AAATGAACTGTGACAGGGATTATGAATACTGGAGTACCCTTCGTCTTCACTGGAAAATT
CTTGATTGGTAATCAACCGTTAACACAACCACTGGTGCATGTTTCAATCATTGCATACA
CTATAGAGCCACAACACATCGAGGCTGGTGTGGGTCAGCTCTCATTGTGCATGTCAACGG
GAAAAAGGCTGTGTACGCAATGCACTCGGCTGGTGGAGGAGGCATAGCTGCCGCTACTAT
AATCACTCAGGAAATGATTGAAGCTGCTGAAAAGGCCTTAGATTGTCTAACTCCTCAAGG
TGCTATAGTTGAAATAGGCATTGACACAGTTGTGCATGTTCCAAGGAAAACCAAATTACG
TAGAACAGTGGCACAATCCTTGCTTCCAACCTAAGTTTGAACCTGCTGTCTCTCACGTTA
CGACCCAGGACCACAAAGGATGTAGACCAGGTTGCTTTTTCTAAGCACACTACTAATTT
GGAAGAACTTCCATCTGTTTTCTCCATGGTTGCAAAAGAGTATGCTAACAGAGTGTTCAC
AACAATAGGTAAAGAAAATCAAATTTTACCACCAAGCAAGCTATTTTAGGGCTGCCAGG
TATGGATCCAATGAAAAGGACACTTCGCCTGGTTTGCCTTATACTCAACAAGGCCTAAA
AAGAGCTCAGTTGGTAAATTTGAGCAAGGCACTATGACACAAAACCTAAAAGAAGCACA
TACTAAATTAAGTGTAGGAAATTAAGAGACATTTTGTACCAAAGTTTTTGAAGACGA
AATTAGGCCTATAGAGAAAATTCATGAAGCGAAGACTAGAATCGTAGACGTCCCCCTTT
TCATCACTGTATCTGGGGACGTCAACTGTTAGGGCGCTTTGCATCTAGATTTCAAATAA
CCCTGGATTGGATTTAGGCTCAGCAATTGGTACTGACCCGACACAGATTGGACTGCTTT
TGCCTTTCAACTCTTACAATACAAATATGTCTACGATGTAGATTATTCTAATTTGATGC
TTCACATTCCACTGCAATGTTTGAATTTTATTGATTGAGAACTTCTTTACGGTTGAAAATGG
ATTTGATGAAAGAATAGGAGACTACCTAAGATCCCTTGCAGTCTCCCGACACGCTTTTGA
AGAGAGAAGAGTTCTAGTTAAGGGTGGCCTCCCGTCAGGGTGCGCCGCCACGAGCATGTT
AAACTATTATAAACAATATAGTAATTAGAGCAGCATTGCATTTAACATATTCAAATTT
TGAATTTGATGACATCAAAGTTCTGTCTTATGGTGTGATTTACTCATCGCCACAAATTA
CCAAATCAATTTTAACTTGTCAAACAAGATTAGCACCCCTTCAATTACAAGATTACACC

TGCTAACAAAGACTGTTGAATTCCTGAAATTTCTAATCTTTATGAGGTTACCTTTTTAAA
 GAGAAAATTTGTCAGACACAATTCTTGTTTATTTAAGCCTCAAATGGACACAGAAAATTT
 GAAAGCAATGGTTAGCTACTGTAGACCAGGTACGTTAAAGGAAAAGCTCAACTCTATTGC
 ACTGCTGGCGGTGCACTCAGGGAAGAGTGTCTACGATGAAATTTTCGACCCTTTTCGTGC
 TATCGGAATAATAGTTCCAGAACATAGTACCATGCTCTATAGATGGCTCAATCTATTTAG
 ATGAACATCCTCTCGATCGGATCGCAACGTGTTACCCAAGAATCCACTTGGGTGTACGCG
 GCCGTTCTGACGTTGGAATTCTGTAGATGAAAG

> CMH-N031-11 (SAFV-6)

Length = 7,717 base pairs

G+C content = 42.93%

A+T content = 57.07%

Nucleotide	Number	Mol%
A	2288	29.65
C	1758	22.78
G	1555	20.15
T	2116	27.42

>CMH-N031-11

CTAATCAGAGGAAAGTCAGCATTTCGGGCCAGGCTAAGAG
 AAACATATAAGTTAGATTTGATTTCAATCCCTTGACGAAATCGGGATGAGAAGTTCTCCC
 TCCCTTGCCGCTTGTTACACCCTCTCCTTGTAATTCGGCCCTTGACAAAGCCCTTG
 GTGAAAGAACCCTCTCTTCCCTGACGTGGTTGGAATTAACATCATTTCGACGAAAGTGC
 TATCATGCCTCCCCGATTATGTGGTGTTCCTGCCCTGCTGGGCGGAGCATTCTCGGGTT
 GAGAAATCTTGACTCTTTCTTTTGGAAACCTTGGTTCCCCGGTCTAAGCCGCTTGAAT
 ATGACAGGGTTATCTTCTATTTCTAATTTCTACTTTTCATGGGTTCTATCCATGAAAAGG
 TATGTGTGCCCTTCTTCTTTGGAGAATTCACACGGCGGTCTTCCGTCTCTCGAAAA
 GTGTGAATGCAACATGCCGGAACGGTGAAGAAAACAGTTGTCTGTGGAAGTCTAGAGTG
 CGCATCGAAACAGCTGTAGCGACCTCACAGTAGCAGCGGAATCCCCTCTTGGCGACAAGA
 GCCTCTGCGGCCAAAAGCCCCGTGGATAAGATCCACTGCTGTGAGCGGTGCAGCCCCAGC
 ACCCTGGTTCGATGATCACTCTTTACTGAACCAGAGAATGGTTTTCTTAAGCCCTCCGTT
 AGAGAAGCCAAGAATGTCCTGAAGGTACCCCGGTGCGGGATCTGATCAGGAGACCAAAT
 GAGAGTGCTCTACACTGTCACCTTTGGTTTAAAACTGTCACAGCTTCTCAAACCAAGTG
 GTCTTGGTTTTCTAATTTTGTAAATGGCAGTATGGCGTGCAAAACACGGATACCCGTTTT
 TGTGCCCTCTTTGTACGGCCATTGAAACCACTCCCGACGTTTCGTTTGTCTTGTCTGTTCC
 ATAATGAATGGTACCCAAGTATCTCCTAACCACGGACTTGGATGACGATGTATTTTATC
 CCAAGATTGTTAATGGACTGGACAGATCTACCATTAATTCACGATATAGTAATGGAAC
 CACAAGGAAATTCAACTCTTCAGACAAAACAATTCACAGTCTTCAGGCAATGAAGGCG
 TGATTATCAACAATTAATACTCTAACCATAATCAAAAATTCATAGACCTTTCTGCTAACG
 CCAATGGTGTGCGGAAAAGAAAACCTCAAACCAAGGTCAACTGATGAACATTCTCGGCA
 GCGCAGCCGATGCTTCAAAAACATTGCCCTCTTCTTATGGATCAAGACACTGAAGAAA

TGACCAACCTTTCAGACAGGGTTGCCTCTGATACAGCTGGCAATACGGCTACAAACACTC
AATCAACTGTAGGCCGTTTGTGGATACGCGATGAGGCATAAAGGAAAACACCCTACTT
CTTGCGCTGACAAAGCGACTGACAAGGTGTTGGCTGCTGAAAGATATTACACTCTGAAAC
TAGCAACTTGGACACCTACACGTAAGCTTTTTGATCAGATCAGAGTACCATTTCCCCACA
TTCTTTCCGGCGAGGATGGTGGTGTGTTCCGAGCGACTCTCAGGAGACACTATCTGTGTA
AGTGTGGCTGGAGAATCAAGTACAGTGCAATGCTTCCCAGTTCATGCTGGTTCCTACTAC
TCGTATTCATGGCCCCAGAATTTGACACTTCTCCCTATGAGTCAGAAGGAGAACCGGATA
GAAATAAGGTTTTCTCCATGGACACTGATTGGCAACAAGGTGCGAAATGTTTACGTGGGC
ATTCCTTTGAATCAAGGACTTCCCAAGAGATGAGACCACTTTCTCTAAACCACCAAAATT
TCTGGCAGTGGACTGTGTACCCCCATCAGATTTTGAATTTGAGAACAAACACCACCTGTG
ACCTGGAAGTACCTTATGTGAACATTTGCCAACATCCTCTTGGACGCAACACGCAAAATT
GGACGCTGGTCATAGCCGTTCTGACACCCTTACAATGCCCGTCAGGATCAGCGACCAATG
TGAAATCACTGCTTCTATAACAACAGTGAAGCCAGTATTTAACGGACTGAGACACGCAG
TAATGGAACCCAAACAGCCCCTTCTGTAACGGTCCGCGAACATGCTGGCACATTTTATT
CAACCACACCTGACACTACAGTCCCGTATATGGGAAGACAATCTCCACACCTGCAGACT
ATATGTGTGGTGAGTTCACCGATCTATTGTCTCTGTGCAAAATACCAACTTTTGTGGAA
ATAAATCTGATACTGATTGGAATCCGTTCTTTCTTCCACTAATACAGTAAAAGAAACAG
AAGCGCTTGTACTTACCAAGTCACTCTCTTCCACAAACCTTGGCAACACGATGCTGT
CCGCTGTGCGCAAGAACTTCAACCAGTACAGAGGATCATTGAACTTTTTGTTGTTTTCA
CTGGCAGTGAATGGTAAAAGGAAAGTTTCTCATATCTTACACCCCCCGGTGCTGACA
AGCCCAAGAACAGAAATCAGGCGATGCAGGCCACTTACGCTATTTGGGACCTTGGACTGA
ATTCTAGTTACAATTTTACTGTGCCTTTTATTTACCTACACATTTCCGTCAGACCTCCT
ACACCAGTCTACTATCACCTCGGTGGATGGTTGGATTTCTGTTTGGCAGTTGACACCAC
TGACATATCCCGCTAATACACCTGTCCATTCTGACATTTTAACTCTTGTCTGGCGGTG
AGGATTTTACACTTCGAATGCCCATCACTCCCACAAAATACGTTCTCAAGGTGTGGACA
ATGCTGAAAAAGGAAAAGTTTCTGATGATAATGCTTCTACCGATTTTGTGCGGAACCG
TCAAACCTACCAGAAAACCAAACCTGAGGTTGCTTTCTTCTACGACCGAGCTACTCAAGCTG
CGGTCATTGCACCAAGCCAAAATTTCTCTCAAACGTTCCGCCCTTAAAACGGTGGAAACTA
CGATGAATCAAGCTCTTTTACACCTTTGCCAAACTATAAAGTAAACCAGTCTAAAGAAG
AACTTGTGAAACAATACCGTTGGCTGGAAACAATTGGGACTGACAAACAGCCACCTTTTA
AGACCAACAGGACTGGAACCTTATTATGTTCTCACCACTTCACTTACTACAAATGTGATC
TGGAAGTGACCTTGACGGCGACAGCTAACGAGAGTCAAGGAGCGGTAGTACGCTACACCC
CTTGCGGAGCACCCACTGACCTAACACGCAGACAATGAGTCTGGTTCCTTCTTTGTCTG
ACACAAGAGACCCACATATGTGGTTCGTGGGACCAGGAACTACAAACCAAACCTCATTCC
TCTTACCTTTCACCTCACCTCTCTCAGTGCTACCGGCCGTGTGGTTCAATGGTTACGCTA
ACTTTAACAATACTCCAGATATGGTGTGCCCCAAATGCAGACTTTGGCAGGTTGTTC
TCCAAGGCAATGGTAATTTTCTATACACTTTAGATACAAGAAGATGAGAGTTTTCTGCC
CTAGACCAACAGTGTATTTTCCATGGCCAACCTGACGCACGAACTACTAAAATACATGCTG
AAAACCTCATCCCCATACTAGAAGTACAAAATCCAGTTTCAATCTATAGAGTTGATTTGT
TTATAAATTTCTCAGATGAAATTTCAATCACTTACAAAGTTCATGGTAGAACGGTTT
GCCAATATGAAATTTCCGGTTTCCGGCCTGTCTCGTCTGGGCGCCTTCTGTCTGCATGG
GCGAGAAACCTTGCAACTGCCAATTTCTACCCCTAAATGTTTCTACCACATTGTCTTCA
CTGGTCCAGGAATTTTGGAGTGTCCATATATAAAGCAAGATACAGACCTTGGAAAC
AGCCTCTCATGATGAACTCAACGACTATGGTTTCTCCACCTTCACTGACTTCTTTAAAG
CTGTTAGAGATTATCACGCTACCTATTACAAACAGAGACTACAACACGACATAGAGACAA
ACCTGGACCTGTCCAGTCGGTTTTTCAAGTGAAGTGGGGTGTAAACCAATCCCAAG
CACCCATGTCAGGTCTCCAAGTATGTTATTAAGAGCAATTGGAATTGAAGCTGATTGCA
CAGAATTCACCAGAGCAGTTAATTTGATTACAGATTTGTGCAATACCTGGGAGAGCGCGA
AAACGACACTCTCCTCTCCTGAGTCTGGACCAAAATGGTGTGAGGATTGTAATAATGG
TAGCTGCTCAGTCTTTACTTACACAACCCAGATTTAACCACTACTGTCTGTTTATCTT

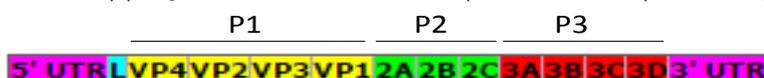
TGATGGCTGGAATTGACATTCTACTAACGACAGTGTTTTTAACTGGTTGCCTCCAAAC
TTTCTAAATTTTCCACACTCCCGCTCCCCACTGTCCACTTCTCCAACAACAGTCTC
CTATTAGGGAAGCAAATGACAATTTCAACTTGGCCAAAACATAGAATGGGCAATAAAAA
CAATAAAAAAGAATAGTTGAATGGATTACTTCATGGTTCAAACAAGAAGAGACTAGTCCTC
AGGCTAACTAGACAAAATGTTAGTAGACTTCCAGAGCACTGTGCCTCCATTTTAGCTA
TGAGAAATGGGAGAAAAGCATACACTGATTGTAAGTGGTGCATTTAAGTATTTGAAACAAC
TTTATAATTTGGCAGTACAATGTAAGAATACCACTAGCCACTTTGTGTGAAAAATTTA
AGAACAACACGACCATGCTGTGGCCAGGCCAGAACCTGTGGTTGTTGTGCTGCGAGGAA
ATGCTGGACAAGGAAAATCCGTGACCAGTCAAATCATCGCTCAAGCAGTTTCCAAACTTT
CATTTGGCCGTCAATCAGTTTACTCTCTTCCACCAGACTCTGATTATTTGGATGGTTATG
AGAATCAATACTCAGTGATCATGGATGATTTAGGACAAAATCCTGACGGAGAAGATTTTA
AAGTGTCTGTCAAATGGTTTCAAGCACAAAATTTCTTCAAATATGGCCACTTGGAAA
AGAAAGGAACCCCTTTTACTTCAAACCTCATAATAGCAACAACCAACCTTCCCAAGTTTC
GACCCGTACCCGTAGCTCACTACCCAGCTGTTGATAGAAGAATCACCTTTGATCTGACCG
TTGAGGCAGGAGATGAATGTGTCACTACAATGGAATGTTGGATGTAGAAAAGGCTTTTG
AAGAAATTCAGGAAAACCACAATTGGATTGTTTCAACTGATTGTAGACTACTCCATA
AGAGAGGGGTGCGGTTTGTATGTAACCGCACAAAAGAATGTCTACAACCTTCAACAGGTAG
TAAAAATGGTGAAGAACACCATTGATAACAAGGTGGAATCTAAAGAAAATGAATACAT
TGTTGCCCAATCACCTGGTAATGACATGGACTATGTATTGACTTGCCTGAGACAAAACAA
ATGCTGCATTACAAGATCAAATTGATGAACTCCAAGAGGCCTTTAAATCAGGCTCAAGAAA
GACAGAATTTCTATCAGATTGGCTTAAAGTTTCTGCAATTGTTTTCGCATCTATTGCCT
CTTTGTCTGCAGTTTGTAAATTAGTTTCCAGATTCAAAAATTTGGTGTGTCCGGCTCCTG
TCCAAATACAATTATCCGAAGGAGAACAAGCTGCTTACTCTGGAGGCAAGAAAGGAGAAA
AGCAAACACTCCAGGTTTTGGACGTACAAGGTGGAGGAAAGATCGTAGCCAGGCTGGGA
ACCCGGTCATGGACTATGAGGTGAATATCGCAAAGAACATGGTCACCCCTATTACTTTCT
TTTATGCAGACAAGGCACAAGTCACACAATCCTGCTTGTGATCAAAGGGCACCTCCTTG
TGGTCAATAGACATGTGGCTGAAACTGACTGGTGCCTTTTGAACCTCAAAGGCACACGGC
ATGAGAGAGATTCCGTACAGATGAGATCAGTGAATAAATCTGGAATGGAAGTTGATTTGA
CTTTTGTGAAAGTTGTTAAAGGACCTCTGTTCAAAGACAATCCAAGAAAATTTGTTCCA
AGAATGATGACTTTCCGGCTAGAAATGAACTGTCACAGGGATTATGAATACTGGAGTAC
CCTTCGTCTTCACTGAAAAATCTTGATTGGCAATCAACCGGTTAACACAACCACTGGTG
CATGTTTCAATCATTGCATACACTACAGAGCCACAACACATCGAGGCTGGTGTGGGTCAG
CTCTCATTTGCCATGTCAACGGCAAAAAGGCTGTGTACGCAATGCACTCAGCTGGTGGAG
GAGGCATAGCTGCCGCCACTATAATCACTCAGGAAATGATTGGAGCTGCTGAAAAGGCCCT
TGGATTGTTAACTCCTCAAGGTGCTATAGTTGAAATAGGCATTGACACAGTTGTGCATG
TTCCAAGGAAAACCAAATTACGTAGAACAGTGGCACATCCTTGTTCACCTAAGTTTG
AACCTGCTGCTCCTCTCACGTTACGACCCAGAACCACAAAAGATGTAGACCAGGTTGCTT
TTTCTAAGCACACTACTAATTTGGAAGAACTTCCATCTGTTTTCTCCATGGTTGCAAAAG
AGTATGCTAACAGAGTGTTCAACAATAGGTAAAGAAAATCAAATCTTGACCCAGAAC
AAGCTATTTTAGGGCTACCAGGTATGGATCCAATGGAAAAGGACACTTCGCCTGGTTTGC
CTTATACTCAACAAGGCCTAAAAAGGGCTCAGTTGGTAAATTTTGAAGCAAGGCACTATGA
CACAAAACCTAAAAGAAGCACATACTAAATTAAGTGTAGGAAATATGAAGACATTTTGT
ACCAAAGTTTTTTGAAAGACGAAATTAGGCCTATAGAGAAAATTCATGAAGCGAAGACTA
GAATCGTAGACGTCCCCCTTTTCACTACTGTATCTGGGGACGTCAACTGTTAGGGCGCT
TTGCATCTAGATTTCAAACCTAACCCCGGATTGGATTTAGGCTCAGCAATTGGCACTGATC
CTGACACAGATTGGACTGCTTTTGCCCTTCAACTCTTACAATACAAATATGTCTACGATG
TAGATTATTCTAATTTGATGCTTACATTCCACTGCAATGTTTGAATTTTGGTTGAAA
ACTTCTTACAGTTGAAAATGGATTTGATGAAAAGAATAGGAGACTACCTAAGATCCCTTG
CAGTCTCCCGACACGCCCTTTGAAGAGAGAAGATTCTAGTTAAGGGTGGCCTCCCGTCAG
GGTGCGCCGCCACGAGCATGTTAAACACTATTATAACAATATAGTAATTAGAGCAGCAT

TGCATTTGACATATTCAAATTTTGAATTTGATGACATCAAAGTTCTGTCTTATGGTGATG
 ATTTACTCATCGCCACAAATTACCAAATCAATTTTAATCTTGTCAAACAAAGATTAGCAC
 CCTTCAATTACAAGATTACACCTGCTAACCAAGACTGTTGAATTTCTGAAATTTCTAATC
 TTTATGAGGTTACCTTTTTAAAGAGAAAATTTGTCAGACACAATTCTGTTTATTTAAGC
 CTCAAATGGACACAGAAAATTTGAAAGCAATGGTTAGCTACTGTAGACCAGGTACGTAA
 AGGAAAAGCTCAACTCTATTGCACTGCTGGCGGTGCACTCAGGGAAGAGTGTCTA
 CGACGAAATCTTCGACCCGTTTCGCCGTATTGGAATAATAGTTCCAGAACATAGTACCATG

ตารางที่ 2 การเปรียบเทียบ full-genome sequence ของ SAFV-1 (CMH-N120-11) และ SAFV-6 (CMH-N031-11) กับ Saffold virus genotype ต่างๆ

Reference strains	Nucleotide sequence identity							
	Full-length*		P1		P2		P3	
	CMH-N031-11	CMH-N120-11	CMH-N031-11	CMH-N120-11	CMH-N031-11	CMH-N120-11	CMH-N031-11	CMH-N120-11
SAFV-1 (AB747248)	83.2	92.6	69.4	93.7	89.5	91.3	91.9	91.6
SAFV-2 (EU376394)	80.5	82.4	67.8	72.3	86.6	87.8	88.0	88.1
SAFV-3 (HQ902242)	80.6	80.7	68.9	69.0	86.8	87.2	87.0	86.8
SAFV-4 (AB747251)	82.7	84.4	68.8	72.5	88.6	90.3	91.2	91.2
SAFV-5 (FJ463615)	84.6	83.2	75.3	69.5	90.0	91.5	88.3	88.0
SAFV-6 (FJ463617)	88.1	83.9	83.6	69.1	89.0	90.9	91.6	91.4
SAFV-7 (AB747254)	82.3	83.1	69.4	70.1	89.1	90.3	88.9	88.7
SAFV-8 (AB747255)	82.4	83.4	68.8	69.8	89.0	90.6	89.6	89.4
SAFV-9 (AB747256)	82.0	82.6	69.6	69.5	89.0	91.2	88.0	87.7
SAFV-10 (AB747257)	82.2	83.5	67.8	70.0	88.6	91.0	91.5	91.1
SAFV-11 (AB747258)	82.3	82.4	70.0	68.5	88.5	89.0	88.5	88.2

* Gene mapping ของ Saffold virus (adapted from <http://www.picornaviridae.com>)

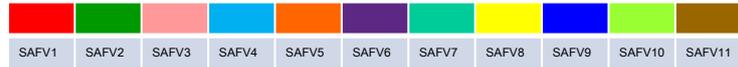


จากการวิเคราะห์เปรียบเทียบ full-genome sequence ในของ CMH-N031-11 และ CMH-N120-11 กับ Saffold virus ต่างๆ พบว่า CMH-N031-11 มีลำดับ nucleotide ใกล้เคียงกับเชื้อมาตรฐานของ Saffold virus genotype 6 (SAFV-6) โดยมีค่าความเหมือนของลำดับ nucleotide อยู่ที่ร้อยละ 88.1 ในขณะที่เมื่อเปรียบเทียบกับ Saffold virus genotype อื่นๆ คือ SAFV-1 ถึง SAFV-5 และ SAFV-7 ถึง SAFV-11 พบว่ามีความเหมือนกันของลำดับ nucleotide ลดลงมาอยู่ที่ร้อยละ 80.5 ถึง 84.6 และเมื่อทำการเปรียบเทียบ full-genome sequence ของ CMH-N120-11 พบว่า ลำดับ nucleotide ที่ได้มีความใกล้เคียงกับเชื้อมาตรฐานของ Saffold virus genotype 1 (SAFV-1) โดยมีค่าความเหมือนของลำดับ nucleotide อยู่ที่ร้อยละ 92.6 ในขณะที่เมื่อเปรียบเทียบกับ Saffold virus genotype อื่นๆ คือ SAFV-2 ถึง SAFV-11 พบว่ามี

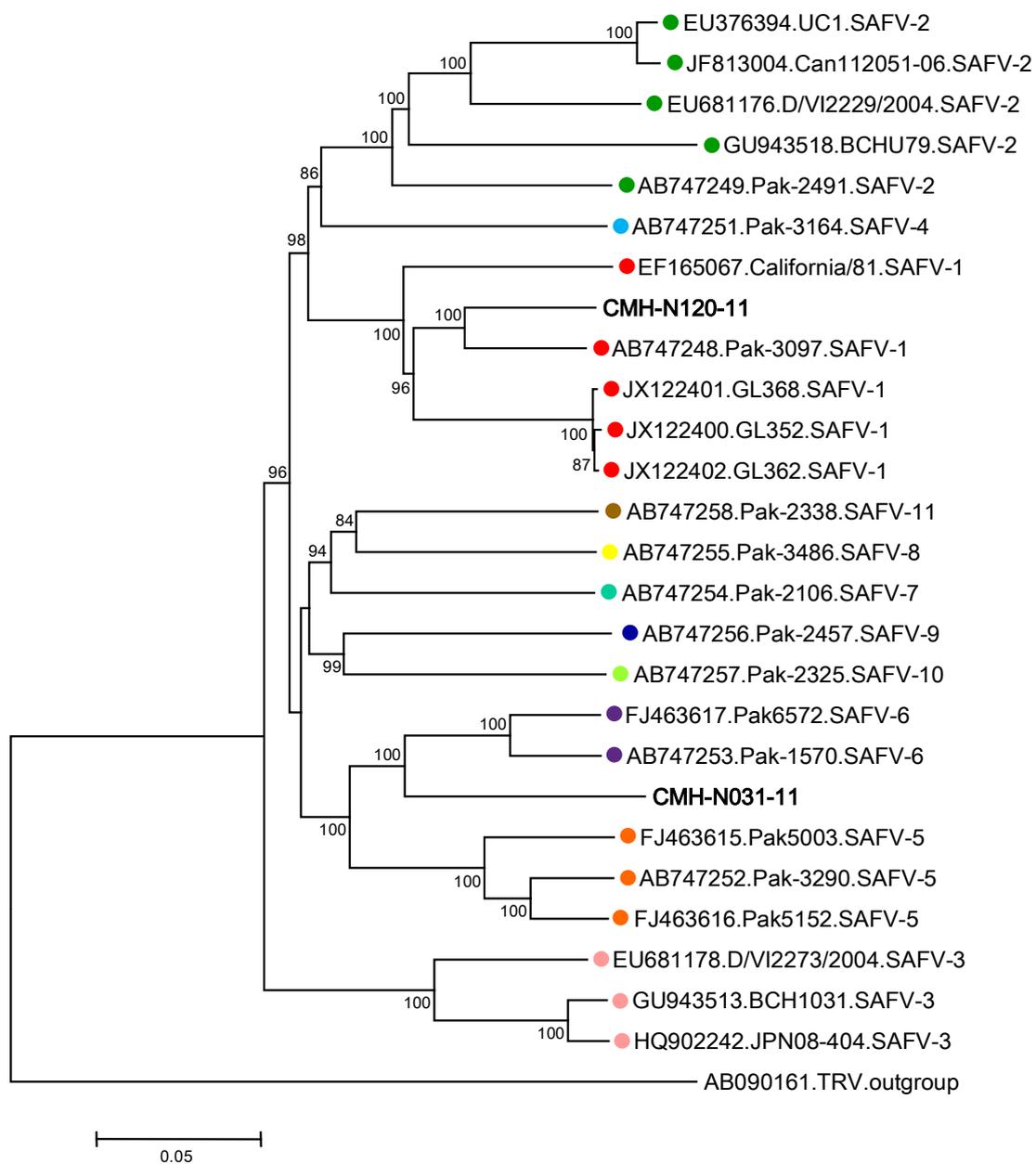
ความเหมือนกันของลำดับ nucleotide ลดลงมาอยู่ที่ร้อยละ 80.7 ถึง 84.4 ซึ่งผลการเปรียบเทียบลำดับ nucleotide ที่ได้จาก full-genome sequence ช่วยยืนยันผลการ typing โดยการวิเคราะห์ partial sequence ของ VP1 gene ในเบื้องต้นว่า CMH-N031-11 จัดจำแนกได้เป็น SAFV-6 และ CMH-N120-11 จัดจำแนกได้เป็น SAFV-1 (ตารางที่ 2)

สำหรับการวิเคราะห์ความเหมือนกันของลำดับ nucleotide ในแต่ละ region คือ P1, P2, และ P3 พบว่า ค่าความเหมือนกันของลำดับ nucleotide ในส่วนของ P1 region ซึ่งเป็นส่วนที่กำหนดการสร้างโปรตีนโครงสร้าง (structural protein) ของตัวเชื้อนั้น ให้ผลการในจัดจำแนก genotype ที่ตรงกับการวิเคราะห์โดยใช้ full-genome sequence ในขณะที่ค่าความเหมือนกันของลำดับ nucleotide ที่ได้จาก P2 และ P3 regions ไม่สามารถนำมาวิเคราะห์เพื่อบ่งบอกชนิดของ genotype ได้ (ตารางที่ 2) ซึ่งผลที่ได้แสดงให้เห็นว่า การจัดจำแนก genotype ของเชื้อในกลุ่ม Saffold virus สามารถทำได้โดยการวิเคราะห์ full-genome sequence หรือ P1 region (ตารางที่ 2)

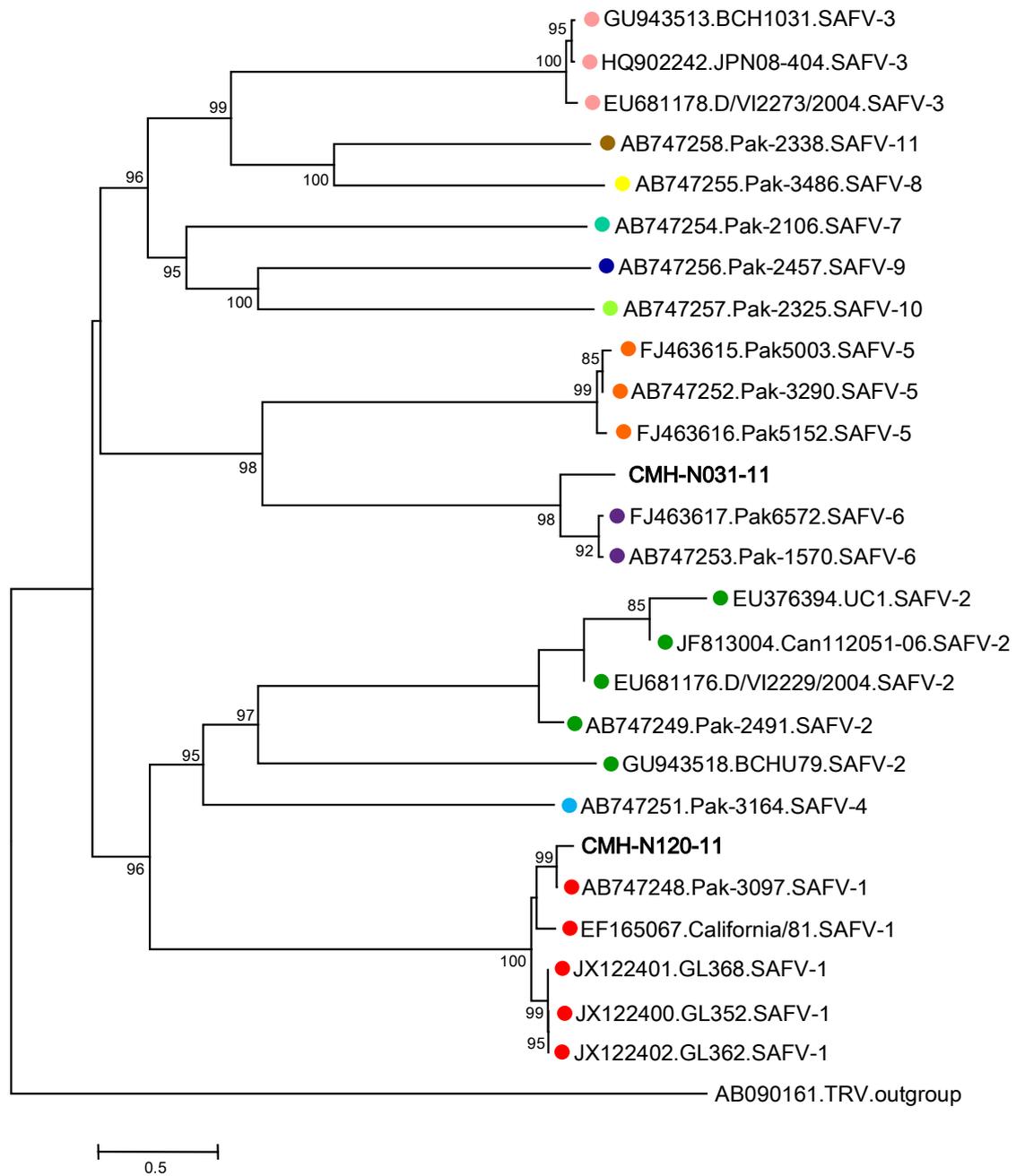
รูปภาพที่ 2 การศึกษาวิวัฒนาการของ Saffold virus CMH-N031-11 และ CMH-N120-11 เปรียบเทียบกับ Saffold virus สายพันธุ์อื่นๆ โดยตัวแทนของเชื้อจากแต่ละ genotype จะแสดงโดยใช้สีที่แตกต่างกัน และ (A) แสดงการวิเคราะห์ในส่วนของ complete coding sequence, (B) แสดงการวิเคราะห์ในส่วนของ P1 region, (C) แสดงการวิเคราะห์ในส่วนของ P2 region, และ (D) แสดงการวิเคราะห์ในส่วนของ P3 region



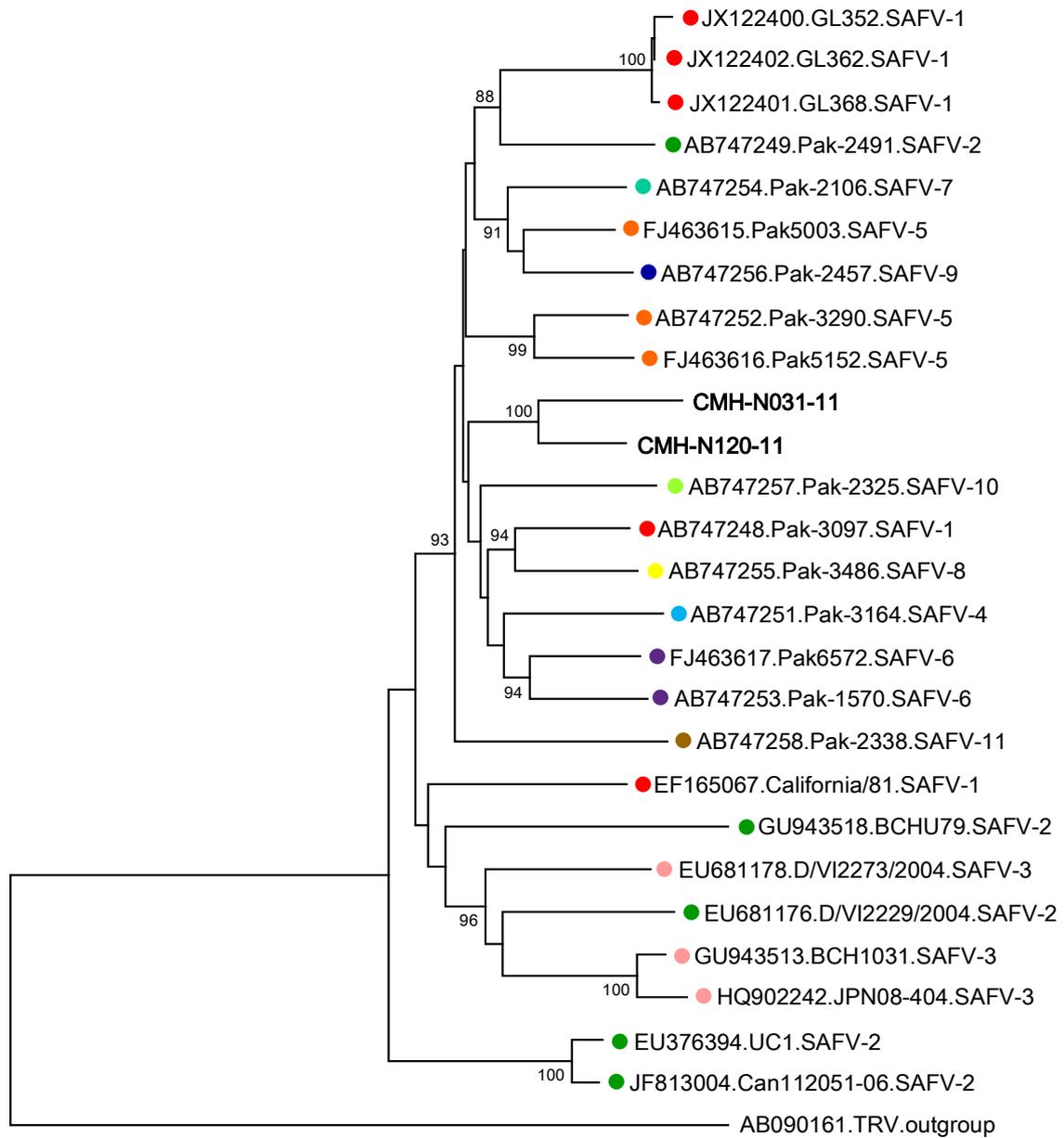
(A)



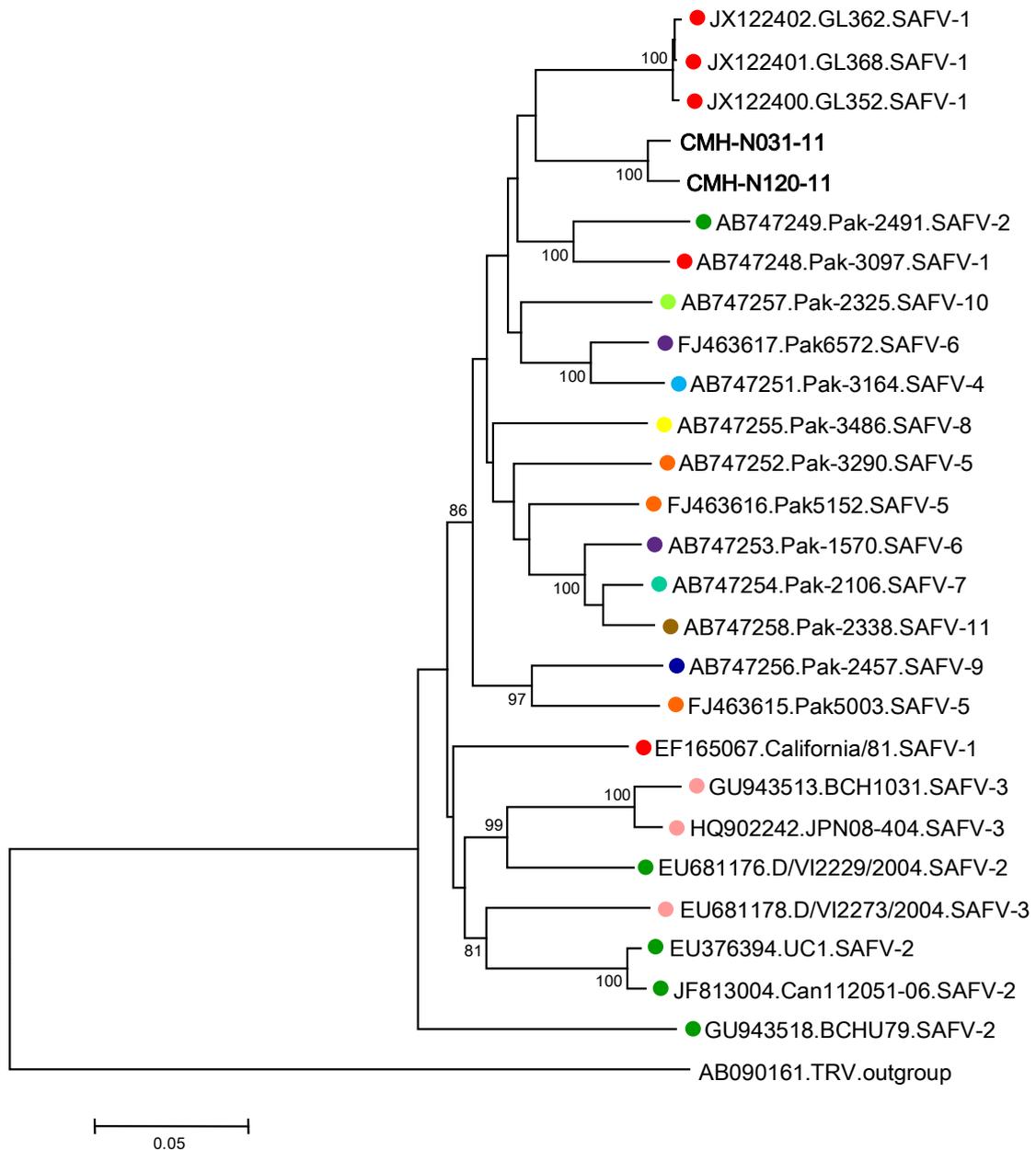
(B)



(C)



(D)



ผลการศึกษาวิวัฒนาการของ Saffold virus CMH-N031-11 และ CMH-N120-11 เปรียบเทียบกับ Saffold virus สายพันธุ์มาตรฐานอื่นๆ ด้วยวิธี Phylogenetic analysis โดยการวิเคราะห์ลำดับ nucleotide ในส่วนของ full-length sequence ดังแสดงในรูปภาพที่ 2 (A) แสดงให้เห็นว่า Saffold virus สามารถแบ่งกลุ่มออกได้เป็น 11 genotypes โดยเชื้อไวรัสที่เป็นตัวแทนของ

แต่ละ genotype จะ form cluster อยู่ด้วยกัน และเมื่อดูการจัดกลุ่มของ CMH-N031-11 และ CMH-N120-11 ที่แยกได้จากการศึกษาในครั้งนี้ พบว่า CMH-N031-11 เป็นเชื้อที่มีความสัมพันธ์ใกล้ชิดกับ SAFV-6 สายพันธุ์ Pak6572 และ Pak-1570 ที่แยกได้ก่อนหน้านี้จากประเทศปากีสถาน (Blinkova et al., 2009) สำหรับ CMH-N120-11 พบว่าเป็นเชื้อที่มีความสัมพันธ์ใกล้ชิดกับ SAFV-1 สายพันธุ์ Pak-3097 จากประเทศปากีสถานเช่นกัน

สำหรับ Phylogenetic analysis ที่ทำการวิเคราะห์ gene ในส่วนของ P1, P2, และ P3 พบว่า ผลการวิเคราะห์ลำดับ nucleotide เฉพาะในส่วนของ P1 เท่านั้นที่ให้ผลการจัดจำแนก genotype ตรงกับการวิเคราะห์โดยใช้ full-length sequence คือ CMH-N031-11 จัดจำแนกได้เป็น SAFV-6 และ CMH-N120-11 จัดจำแนกได้เป็น SAFV-1 (รูปภาพที่ 2 (B)) ซึ่งแตกต่างจากผลการศึกษาในส่วนของ P2 และ P3 regions (รูปภาพที่ 2 (C) และ (D)) ที่ให้ผลการจัดจำแนกไม่สัมพันธ์กับชนิดของ genotypes ดังเช่นการพบว่า SAFV-2 และ SAFV-3 จัดกลุ่มอยู่ร่วมกัน ในขณะที่ CMH-N031-11 และ CMH-N120-11 ที่แยกได้จากการศึกษาในครั้งนี้ มีลำดับ nucleotide ที่ใกล้เคียงกัน ในส่วนของ P2 และ P3 regions โดยพบว่าผลที่ได้สอดคล้องกับค่าความเหมือนกันของลำดับ nucleotide ดังแสดงในตารางที่ 2 ข้างต้น ซึ่งแสดงให้เห็นว่า การจัดจำแนก genotype ของ Saffold virus สามารถทำได้โดยการวิเคราะห์ full-genome sequence หรือ P1 region เท่านั้น

โดยสรุป ผลที่ได้จากการตรวจหา Saffold virus ในศึกษานี้พบว่า ไวรัสชนิดนี้พบได้อัตราที่ค่อนข้างต่ำในเด็กที่มีอาการอุจจาระร่วง จากผลการวิเคราะห์คุณลักษณะทางพันธุกรรมของตัวเชื้อพบว่าเชื้อที่แยกได้ในประเทศไทยมีวิวัฒนาการใกล้เคียงกับเชื้อที่พบในประเทศปากีสถาน โดยเชื้อที่พบในการศึกษาในครั้งนี้ คือ CMH-N120-11 เป็น SAFV-1 ส่วนอีก 2 ตัวอย่างคือ CMH-S052-10 และ CMH-N032-11 เป็น SAFV-6 รายงานการค้นพบ Saffold virus ในครั้งนี้ถือเป็นรายงานแรกของการตรวจพบ SAFV-6 ในผู้ป่วยเด็กที่มีอาการอุจจาระร่วงในประเทศไทย นอกเหนือจาก SAFV-1 ถึง SAFV-4 ที่เคยมีรายงานแล้วก่อนหน้านี้

5. ประโยชน์ที่ได้รับ

นอกเหนือจากเชื้อไวรัสที่เราทราบกันดีว่าเป็นสาเหตุสำคัญของโรคอุจจาระร่วง ดังเช่น Rotavirus, Norovirus, Sapovirus, Astrovirus, Adenovirus, Enterovirus, Parechovirus, Aichi virus, และ Bocavirus แล้ว การศึกษาในครั้งนี้ได้แสดงให้เห็นว่า Cosavirus และ Saffold virus ก็ถือเป็นไวรัสที่พบได้ในเด็กที่มีอาการอุจจาระร่วงในจังหวัดเชียงใหม่ ประเทศไทย ถึงแม้อุบัติการณ์การตรวจพบจะค่อนข้างต่ำก็ตาม จากรายงานการค้นพบไวรัสทั้งสองชนิดนี้ครั้งแรกในปี ค.ศ. 2007 และ 2008 จนกระทั่งถึงปัจจุบัน ข้อมูลการศึกษาด้านระบาดวิทยา ความรุนแรงในการก่อโรค รวมถึงคุณลักษณะทางพันธุกรรมของไวรัสเหล่านี้ยังมีอยู่อย่างจำกัด ข้อมูลที่ได้จากการศึกษาในครั้งนี้ นอกจากจะได้ข้อมูลทางด้านระบาดวิทยาแล้ว ยังทำให้ผู้วิจัยได้เข้าใจถึงคุณลักษณะพื้นฐานทางพันธุกรรมของไวรัสทั้งสองชนิดนี้ โดยพบว่าการศึกษาเพื่อจัดจำแนกชนิดของ genotype ของไวรัสเหล่านี้ควรทำการศึกษาโดยการวิเคราะห์ความแตกต่างของลำดับ nucleotide ในส่วนของโปรตีนโครงสร้างโดยเฉพาะในส่วนของ P1 region หรือ full-length sequence ซึ่งผลการศึกษาทั้ง Cosavirus และ Saffold virus ในประเทศไทยพบว่าเชื้อที่พบระบาดมีความหลากหลายทางพันธุกรรม โดยสามารถตรวจพบชนิดของ genotype ที่แตกต่างกัน และเชื้อส่วนใหญ่มีต้นกำเนิดทางวิวัฒนาการร่วมกับเชื้อที่ระบาดในประเทศอื่นๆ ในแถบทวีปเอเชีย

6. References

- Blinkova O, Kapoor A, Victoria J, Jones M, Wolfe N, Naeem A, Shaukat S, Sharif S, Alam MM, Angez M, Zaidi S, Delwart EL. Cardioviruses are genetically diverse and cause common enteric infections in South Asian children. *J Virol.* 2009;83:4631-4641.
- Chaimongkol N, Khamrin P, Suantai B, Saikhreang W, Thongprachum A, Malasao R, Ukarapol N, Kongsricharoern T, Ushijima H, Maneekarn N. A wide variety of

diarrhea viruses circulating in pediatric patients in Thailand. *Clin Lab*. 2012;58:117-123.

Dai XQ, Hua XG, Shan TL, Delwart E, Zhao W. Human cosavirus infections in children in China. *J Clin Virol*. 2010;48:228-229.

Jones MS, Lukashov VV, Ganac RD, Schnurr DP. Discovery of a novel human picornavirus in a stool sample from a pediatric patient presenting with fever of unknown origin. *J Clin Microbiol*. 2007;45:2144-2150.

Kapoor A, Victoria J, Simmonds P, Slikas E, Chieochansin T, Naeem A, Shaukat S, Sharif S, Alam MM, Angez M, Wang C, Shafer RW, Zaidi S, Delwart E. A highly prevalent and genetically diversified Picornaviridae genus in South Asian children. *Proc Natl Acad Sci USA*. 2008;105:20482-20487.

Khamrin P, Chaimongkol N, Malasao R, Suantai B, Saikhruang W, Kongsricharoern T, Ukarapol N, Okitsu S, Shimizu H, Hayakawa S, Ushijima H, Maneekarn N. Detection and molecular characterization of cosavirus in adults with diarrhea, Thailand. *Virus Genes*. 2012;44:244-246.

Khamrin P, Chaimongkol N, Nantachit N, Okitsu S, Ushijima H, Maneekarn N. Saffold cardioviruses in children with diarrhea, Thailand. *Emerg Infect Dis*. 2011;17:1150-1152.

Yodmeeklin A, Khamrin P, Chuchaona W, Saikruang W, Malasao R, Chaimongkol N, Kongsricharoern T, Ukarapol N, Maneekarn N. Saffold viruses in pediatric patients with diarrhea in Thailand. *J Med Virol*. 2015;87:702-707.

Output ที่ได้จากโครงการ

1. ผลงานการตีพิมพ์ในวารสารวิชาการนานาชาติ : Khamrin P, Maneekarn N. Detection and genetic characterization of cosavirus in a pediatric patient with diarrhea. *Arch Virol*. 2014. 159: 2485-2489. (Impact factor 2.03)
2. การนำเสนอผลงานวิจัยในที่ประชุมวิชาการ : Khamrin P, Maneekarn N. Detection and genetic characterization of cosavirus in a pediatric patient with diarrhea. The Thailand Research Fund (TRF) Seminar Series in Basic Research CII “From Molecular to Market”, Chiang Mai, Thailand, 2 May, 2014. (Oral presentation)
3. การเผยแพร่ข้อมูลทางพันธุกรรมของ Cosavirus ในฐานข้อมูล NCBI :
<http://www.ncbi.nlm.nih.gov/nucore/KJ396940.1>

ภาคผนวก

1. เอกสารแนบหมายเลข 1 : Reprint ผลงานการตีพิมพ์งานวิจัย
2. เอกสารแนบหมายเลข 2 : PowerPoint การนำเสนอผลงานในที่ประชุมวิชาการโดย
สำนักงานกองทุนสนับสนุนการวิจัย (สกว.)
3. เอกสารแนบหมายเลข 3 : การเผยแพร่ข้อมูลทางพันธุกรรมของ Cosavirus ในฐานข้อมูล
NCBI (<http://www.ncbi.nlm.nih.gov/nuccore/KJ396940.1>)

Detection and genetic characterization of cosavirus in a pediatric patient with diarrhea

Pattara Khamrin · Niwat Maneekarn

Received: 5 February 2014 / Accepted: 13 April 2014 / Published online: 1 May 2014
© Springer-Verlag Wien 2014

Abstract Human cosavirus (HCoSV) is a newly discovered virus of the family *Picornaviridae*. A total of 411 fecal specimens were collected from children admitted to hospitals with acute gastroenteritis symptoms in Chiang Mai, Thailand, in 2010–2011 and screened for HCoSV by an RT-nested PCR method. HCoSV was found in a single specimen (CMH-N199-11) collected from a 3-year-old boy. This represents the first report of HCoSV infection in a pediatric patient with diarrhea in Thailand. Analysis of the complete coding sequence revealed that this HCoSV was most similar to the Chinese HCoSV-A reference strain SH1, and belonged to genotype A6. The data imply that HCoSVs detected in Thailand and China share the same evolutionary ancestor. Our results emphasize the need for further research to understand the distribution, genetic diversity, and association of the HCoSV with acute gastroenteritis in humans.

Keywords Cosavirus · Diarrhea · Children · Thailand

Acute gastroenteritis is one of the most common diseases in children and adults and continues to be a significant cause of morbidity and mortality worldwide [4]. Recently, human cosavirus (HCoSV), a new member of the family *Picornaviridae*, was originally identified from stool samples of both healthy children and non-polio acute flaccid paralysis (AFP) patients in Pakistan and Afghanistan [7]. Subsequently, HCoSVs were also isolated from feces of diarrhea or non-polio AFP patients in several countries

worldwide [2, 3, 10, 11, 13]. However, in several molecular epidemiological studies, HCoSVs have also been detected in stool samples of asymptomatic control subjects [3, 7, 8, 13]. In addition, contaminations by HCoSVs have been found in treated and untreated sewage water and in river water samples [1, 5]. Presumably, the HCoSVs are transmitted primarily via the fecal-oral route, either by person-to-person contact or by ingestion of contaminated food or water.

The genome of HCoSV is a positive-sense, single-stranded RNA of approximately 7.0 to 7.6 kilobases in length. The genome consists of four structural viral proteins (VP1–VP4) and seven nonstructural proteins (2A, 2B, 2C, 3A, 3B, 3C, 3D) [6, 7]. HCoSVs can be divided into at least 6 genetic groups (A to F), with more than 30 genotypes [9]. For HCoSV pathogenesis and evolution, while the viruses are thought to be gastrointestinal pathogens because they have been found in fecal samples, the viral replication process and evolutionary mechanism are not yet understood. Viral genome recombination is thought to be one of the mechanisms by which HCoSV evolves [9]. Therefore, analysis of full-length genome sequences of HCoSV may be helpful for investigating the genetic diversity and evolution of this virus. Here, we report the detection and genetic characterization of HCoSV in a pediatric patient with diarrhea in Thailand.

A total of 411 stool specimens were collected from children under 5 years of age with diarrhea at two hospitals in Chiang Mai, Thailand. The study period was 2 years, from January, 2010 to December, 2011. Only pediatric patients who had a clinical diagnosis of acute gastroenteritis with watery stool samples were included in this study. In addition, the same set of specimens was also screened for other diarrheal viruses, including group A, B, and C rotaviruses, adenovirus, norovirus GI and GII,

P. Khamrin · N. Maneekarn (✉)
Department of Microbiology, Faculty of Medicine, Chiang Mai University, 110 Intawarorose, Sriphoom, Muang,
Chiang Mai 50200, Thailand
e-mail: nmaneeaka@mail.med.cmu.ac.th; niwat.m@cmu.ac.th

Table 1 Comparison of the complete coding region of HCoV CMH-N199-11 to those of HCoV-A, HCoV-E, HCoV-F, and HCoV-D/E

Region	Nucleotide sequence identity (%)						
	HCoV-A SH1	HCoV-A1 0553	HCoV-A19 PK6187	HCoV-A20 NG263	HCoV-E1 Australia/81	HCoV-F1 PK5006	HCoV-D6/E2 NG385
VP4	97.0	97.0	96.0	95.0	80.3	77.9	78.9
VP2	90.2	63.6	66.5	70.4	56.0	65.2	58.3
VP3	89.0	67.3	65.2	68.9	59.7	60.4	62.2
VP1	85.6	59.3	57.7	57.8	46.4	50.7	48.8
2A	82.2	60.0	59.3	60.4	47.7	37.2	42.2
2B	82.9	84.8	81.8	84.5	64.7	66.1	64.4
2C	84.4	85.8	85.8	86.8	69.5	64.7	70.7
3A	82.0	80.8	80.8	88.0	60.1	61.8	59.3
3B	85.9	87.7	84.2	84.2	77.1	75.4	80.7
3C	86.0	87.5	88.5	88.0	60.5	59.1	61.9
3D	91.6	92.6	90.6	92.6	67.8	66.8	67.5
Coding sequence	87.3	77.2	77.2	78.6	60.6	61.8	61.4

sapovirus, astrovirus, Aichi virus, human parechovirus and enterovirus, using RT-multiplex PCR protocols described previously [12, 15, 16]. The study was conducted with the approval of the Ethical Committee for Human Rights Related to Human Experimentation, Faculty of Medicine, Chiang Mai University (No. 181/2554).

The HCoV genome was first extracted from the supernatant of a 10 % fecal suspension using a QIAamp Viral RNA Mini Kit (QIAGEN, Germany). The presence of HCoV was detected by an RT-nested PCR that targeted and amplified the 5' untranslated region (UTR) [7].

The amplified PCR product of HCoV was subjected to direct sequencing using a BigDye Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA). The sequence was compared with those of reference strains available in the NCBI GenBank database using the BLAST server (www.ncbi.nlm.nih.gov/blast). In addition, the complete coding sequence (VP4, VP2, VP3, VP1, 2A, 2B, 2C, 3A, 3B, 3C, 3D) of the detected HCoV strain was determined. cDNA generated by reverse transcription using random hexamer primers (Takara, Shiga, Japan) was used as a template for amplification of the complete coding region of the genome. The accession number of the HCoV sequence, available in the GenBank database, is KJ396940. The complete coding sequence as well as the individual regions encoding VP4, VP2, VP3, VP1, 2A, 2B, 2C, 3A, 3B, 3C, and 3D were compared with those of HCoV reference strains available in the NCBI GenBank database. Phylogenetic trees of the complete coding sequence and the VP1 coding region were constructed by the maximum-likelihood algorithm, using MEGA (v5.20) software [14].

Out of a total of 411 fecal specimens collected, HCoV, designated CMH-N199-11, was found in one specimen

(0.2 %) collected from a child admitted to hospital with diarrhea. HCoV CMH-N199-11 was collected in December 2011 from a 3-year-old boy admitted with acute gastroenteritis at Nakornping Hospital, Chiang Mai province, Thailand. At the time of admission, the patient had low-grade fever, nausea, and vomiting without abdominal pain. Mild dehydration with watery diarrhea was recorded and the number of diarrheal episodes per day was up to three. It was interesting to note that a mixed infection of HCoV with group A rotavirus and enterovirus was found in this patient.

Initially, the HCoV CMH-N199-11 strain was characterized by analyzing the partial 5' UTR region. This CMH-N199-11 strain had a highest pairwise nucleotide sequence identity (98 %) with the HCoV-A SH1, 5192, 5008, 5046, and 0553 strains across the partial 5' UTR region. This HCoV strain was analyzed further by amplification of the complete coding sequence and compared with those of seven complete sequences of HCoV strains available in the NCBI GenBank database (Table 1). It was observed that the polyprotein coding sequence from the VP4 to 3D region (excluding the 5' and 3' untranslated sequences) was 6,372 nucleotides long and encoded a potential polyprotein precursor of 2,123 amino acids. Analysis of the full-length sequence of CMH-N199-11 showed that it was closely related to HCoV-A (SH1, 0553, PK6187, NG263), with nucleotide sequence identity ranging from 77.2 % to 87.3 %, and it displayed the highest nucleotide sequence identity to the HCoV-A Chinese SH1 strain (GU968209), at 87.3 %. In contrast, comparison with HCoV-E, HCoV-F, HCoV-D/E revealed sequence identities ranging only from 60.6 % to 61.8 %. Analysis of each genome region from VP4 to 3D revealed that the VP4

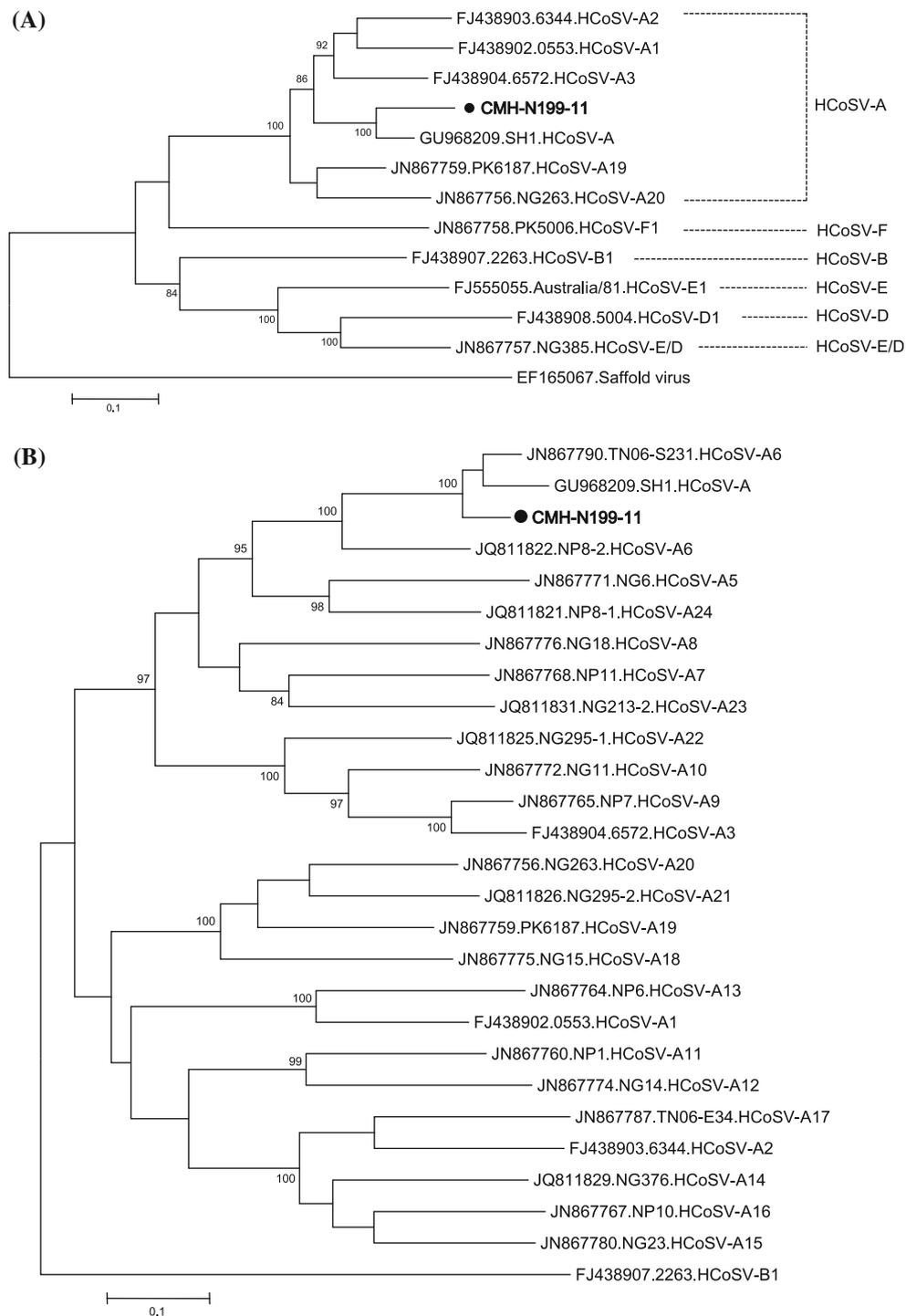


Fig. 1 Phylogenetic analysis of the complete coding region (A) and VP1 (B) of HCoSV CMH-N199-11 and reference strains. Phylogenetic trees were generated using the maximum-likelihood algorithm in MEGA v5.20. Scale bars indicate nucleotide substitutions per site,

and bootstrap values (>80) are indicated for the corresponding nodes. The HCoSV strain detected in the present study is presented in boldface with a black dot symbol

coding region was the most conserved among the different HCoSVs. The nucleotide sequence identity of the VP4 of CMH-N199-11 to other HCoSV-A isolates ranged from 95.0 % to 97.0 %, while they varied from 77.9 % to

80.3 % when compared with those of HCoSV-E, HCoSV-F, and HCoSV-D/E. As seen in picornaviruses, the VP1 coding sequence is the most exposed and immunodominant part of the capsid protein and is the most variable among

different HCoSVs [9]. In our study, comparison between the VP1 sequence of CMH-N199-11 and those of other HCoSV strains (HCoSV-A1, HCoSV-A19, HCoSV-A20, HCoSV-E, HCoSV-F, HCoSV-D/E) revealed nucleotide sequence identities ranging from 46.4 % to 59.3 %. It was found that the VP1 of CMH-N199-11 showed a high level of nucleotide sequence identity only to the VP1 sequence of the SH1 reference strain, at 85.6 %.

Eleven full-length sequences of HCoSV were analyzed together with the CMH-N199-11 strain by nucleotide multialignment and phylogenetic analysis. The results confirmed that the HCoSV detected in this study belonged to the species *Cosavirus A* and formed an exclusive monophyletic branch with the HCoSV-A SH1 strain (Fig. 1A). Based on VP1 analysis, the species *Cosavirus A* includes at least 24 genotypes [9]. In order to identify the genotype of the HCoSV detected in this study, the VP1 of CMH-N199-11 was compared with those of the established HCoSV-A1 to A24 reference strains. It was observed that the nucleotide sequence of CMH-N199-11 shared the highest level of identity with the HCoSV-A6 genotype strain TN06-S231 (JN867790), at 89.9 %, followed by the SH1 strain at 87.5 %. However, when compared with the remaining genotypes, the nucleotide sequence identities ranged from 57.4 % to 69.2 %. In addition, a phylogenetic tree was constructed on the basis of the VP1 nucleotide sequences of CMH-N199-11 and HCoSV-A1 to HCoSV-A24 (Fig. 1B). This analysis clearly showed that the CMH-N199-11 strain is located in a separate branch together with the TN06-S231 and SH1 HCoSV reference strains. Therefore, our results indicate that CMH-N199-11 together with the SH1 strain belong to the HCoSV-A6 genotype.

The present study reports the prevalence and characteristics of HCoSV for the first time in a pediatric patient admitted to hospital with diarrhea in the Chiang Mai area of northern Thailand. This report also describes the genetic background of HCoSV detected in the country. The low rate of detection of HCoSV in this studied population suggests that HCoSV is an uncommon cause of acute gastroenteritis in hospitalized patients in Chiang Mai, Thailand. The HCoSV strain CMH-N199-11 was isolated from the sporadic case of a 3-year-old boy admitted to hospital with acute gastroenteritis in December 2011. In addition to acute gastroenteritis, this pediatric patient presented with clinical symptoms of vomiting and fever. From a literature search, the only other initial report of HCoSV in Thailand was carried out in 2008 [10]. In that study, a single case of HCoSV infection was found in an adult patient with diarrhea, while the specimens collected from children admitted to the same hospital with diarrhea were all negative for HCoSV. Analysis of the nucleotide sequence of the 5' UTR revealed that this virus was HCoSV-D, whereas the HCoSV found in the present study belonged to HCoSV-A. Since the

detection rate of HCoSV in our study was relatively low and the virus was found in a coinfection with other diarrheal viruses including group A rotavirus and enterovirus, this findings raise concern about an etiologic role of HCoSV in gastroenteritis. Further study using a quantitative RT-PCR method to monitor the viral load of each virus and to evaluate viral shedding in follow-up stool samples may be necessary. Although HCoSV may not act as the major causative agent of diarrhea, the question whether or not this virus might be a co-factor for other viral agents or may require helper viruses to establish disease in human remains unclear. In order to clarify this point, further extensive epidemiological surveillance with larger numbers of clinical specimens and screening for other types of viruses is necessary. The study may provide a better understanding of the distribution, genetic diversity, and association of viral agents with acute gastroenteritis in humans.

The complete coding sequence and phylogenetic analysis of the HCoSV isolate detected in this study showed it to be highly similar to the HCoSV Chinese SH1 strain [3]. These viruses shared the complete coding sequence at a high level (87.3 %), and both of them belonged to HCoSV-A6. The data imply that HCoSVs isolated in Thailand and China share the same evolutionary ancestor. As revealed by analysis of complete coding sequences, VP4 was highly similar among the different HCoSVs, while the VP1 region showed the highest level of nucleotide sequence divergence. Therefore, our data support the report of Kapusinszky et al. [9] that VP1 is particularly suitable for distinguishing genotypes of HCoSV.

Since the detection rate of HCoSV in stool specimens was relatively low, and because the study was not performed in healthy individuals, the association of this viral agent with diarrhea in humans is still unclear. In addition, to obtain further information on HCoSV infections, seroprevalence studies may help to estimate the anti-HCoSV antibody prevalence in human populations. In summary, this study presents the detection rate and genetic background information on an HCoSV strain in a pediatric patient with acute gastroenteritis in Thailand.

Acknowledgments This research was supported by grants from the Thailand Research Fund (TRF), the Office of Higher Education Commission (OHEC), and the Faculty of Medicine, Chiang Mai University (CMU), Thailand.

References

1. Blinkova O, Rosario K, Li L, Kapoor A, Slikas B, Bernardin F, Breitbart M, Delwart E (2009) Frequent detection of highly diverse variants of cardiovirus, cosavirus, bocavirus, and circovirus in sewage samples collected in the United States. *J Clin Microbiol* 47:3507–3513

2. Campanini G, Rovida F, Meloni F, Cascina A, Ciccocioppo R, Piralla A, Baldanti F (2013) Persistent human cosavirus infection in lung transplant recipient, Italy. *Emerg Infect Dis* 19:1667–1669
3. Dai XQ, Hua XG, Shan TL, Delwart E, Zhao W (2010) Human cosavirus infections in children in China. *J Clin Virol* 48:228–229
4. Dennehy PH (2011) Viral gastroenteritis in children. *Pediatr Infect Dis J* 30:63–64
5. Haramoto E, Otagiri M (2014) Occurrence of human cosavirus in wastewater and river water in Japan. *Food Environ Virol* 6:62–66
6. Holtz LR, Finkbeiner SR, Kirkwood CD, Wang D (2008) Identification of a novel picornavirus related to cosaviruses in a child with acute diarrhea. *Virol J* 5:159
7. Kapoor A, Victoria J, Simmonds P, Slikas E, Chieochansin T, Naeem A, Shaikat S, Sharif S, Alam MM, Angez M, Wang C, Shafer RW, Zaidi S, Delwart E (2008) A highly prevalent and genetically diversified Picornaviridae genus in South Asian children. *Proc Natl Acad Sci USA* 105:20482–20487
8. Kapusinszky B, Minor P, Delwart E (2012) Nearly constant shedding of diverse enteric viruses by two healthy infants. *J Clin Microbiol* 50:3427–3434
9. Kapusinszky B, Phan TG, Kapoor A, Delwart E (2012) Genetic diversity of the genus cosavirus in the family Picornaviridae: a new species, recombination, and 26 new genotypes. *PLoS One* 7:e36685
10. Khamrin P, Chaimongkol N, Malasao R, Suantai B, Saikhruang W, Kongsricharoern T, Ukarapol N, Okitsu S, Shimizu H, Hayakawa S, Ushijima H, Maneekarn N (2012) Detection and molecular characterization of cosavirus in adults with diarrhea, Thailand. *Virus Genes* 44:244–246
11. Maan HS, Chowdhary R, Shakya AK, Dhole TN (2013) Genetic diversity of cosaviruses in nonpolio acute flaccid paralysis cases of undefined etiology, Northern India, 2010–2011. *J Clin Virol* 58:183–187
12. Pham NT, Trinh QD, Chan-It W, Khamrin P, Shimizu H, Okitsu S, Mizuguchi M, Ushijima H (2010) A novel RT-multiplex PCR for detection of Aichi virus, human parechovirus, enteroviruses, and human bocavirus among infants and children with acute gastroenteritis. *J Virol Methods* 169:193–197
13. Stöcker A, Souza BF, Ribeiro TC, Netto EM, Araujo LO, Corrêa JI, Almeida PS, de Mattos AP, Ribeiro Hda C Jr, Pedral-Sampaio DB, Drosten C, Drexler JF (2012) Cosavirus infection in persons with and without gastroenteritis, Brazil. *Emerg Infect Dis* 18:656–659
14. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731–2739
15. Yan H, Nguyen TA, Phan TG, Okitsu S, Li Y, Ushijima H (2004) Development of RT-multiplex PCR assay for detection of adenovirus and group A and C rotaviruses in diarrheal fecal specimens from children in China. *Kansenshogaku Zasshi* 78:699–709
16. Yan H, Yagyu F, Okitsu S, Nishio O, Ushijima H (2003) Detection of norovirus (GI, GII), sapovirus and astrovirus in fecal samples using reverse transcription single-round multiplex PCR. *J Virol Methods* 114:37–44

Detection and genetic characterization of cosavirus in a pediatric patient with diarrhea

การตรวจหาและการวิเคราะห์ลักษณะทางพันธุกรรมของเชื้อ
cosavirus ที่แยกได้จากผู้ป่วยเด็กที่มีอาการอุจจาระร่วง

Pattara Khamrin and Niwat Maneeakarn

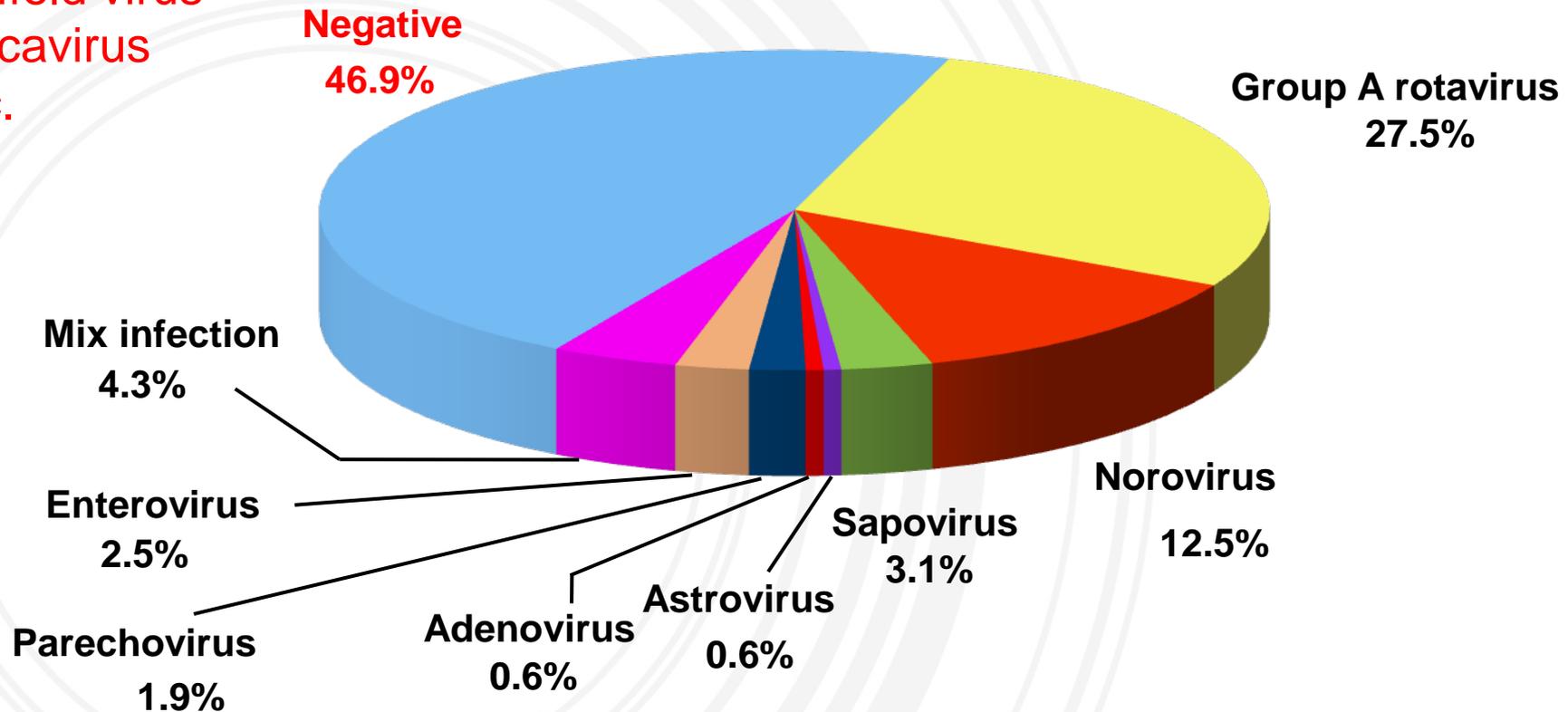
Department of Microbiology, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

Diarrheal virus infections in pediatric patients with acute diarrhea, Chiang Mai, Thailand

Chaimongkol et al., Clin Lab. 2012;58:117-23.

Unknown / Novel viruses ?

- **Cosavirus**
- Saffold virus
- Bocavirus
- etc.



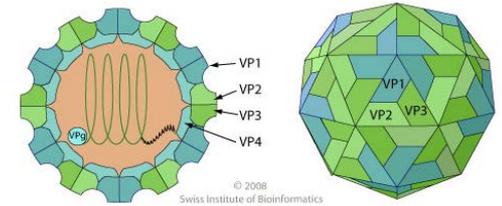
Cosavirus derived from common stool-associated picornavirus

- Human cosavirus (HCoSV), a **new member** in *Picornaviridae* family, was originally identified from stool samples of both healthy children and non-polio acute flaccid paralysis patients in Pakistan and Afghanistan.

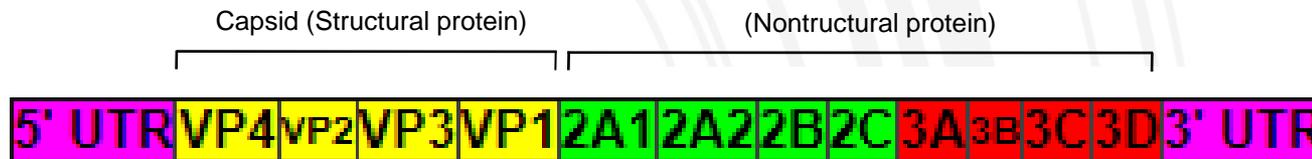
(Kapoor et al., Proc Natl Acad Sci U S A. 2008, 105:20482-7)

Human Cosavirus (HCoSV)

- Icosahedral symmetry, ~20-30 nm in diameter
- Genome : positive-sense, single-stranded RNA ~ 7.0-7.6 kilobases in length
- Genome structure



http://www.expasy.org/viralzone/all_by_species/97.html



- 6 species (HCoSV-A to HCoSV-F)
- HCoSV-A : 24 genotypes (VP1 sequence)

Detections of HCoSV

- A stool sample from a 64-year-old woman collected in Scotland was positive for HCoSV by RT-PCR [Kapoor et al., 2008].
- A report from the United States / Japan demonstrated that HCoSVs were also detected in untreated sewage samples [Blinkova et al., 2009 / Haramota et al., 2014].
- In China, HCoSV was reported from healthy (1.6%) and diarrheic children (3.2%) [Dai et al., 2010].
- In Chiang Mai, Thailand, a single case of HCoSV was reported from adult patient with diarrhea (0.6%) [Khamrin et al., 2012].

To detect and characterize HCoSV in children admitted to hospitals with diarrhea in Chiang Mai, Thailand

➤ Specimen collection

Children admitted to 2 hospitals in Chiang Mai, Thailand
(Maharaj Nakorn Chiang Mai / Nakornping hospitals)

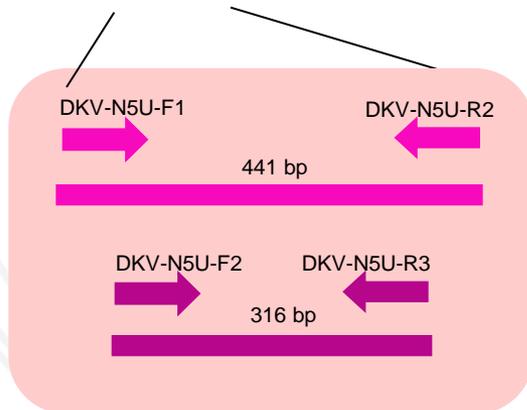
No. : 411 stool samples

Symptom : Diarrhea

Col. date : Jan 2010 – Dec 2011

➤ RT-nested PCR : HCoV-229E

(Kapoor et al Proc Natl Acad Sci U S A. 2008, 105:20482-7)

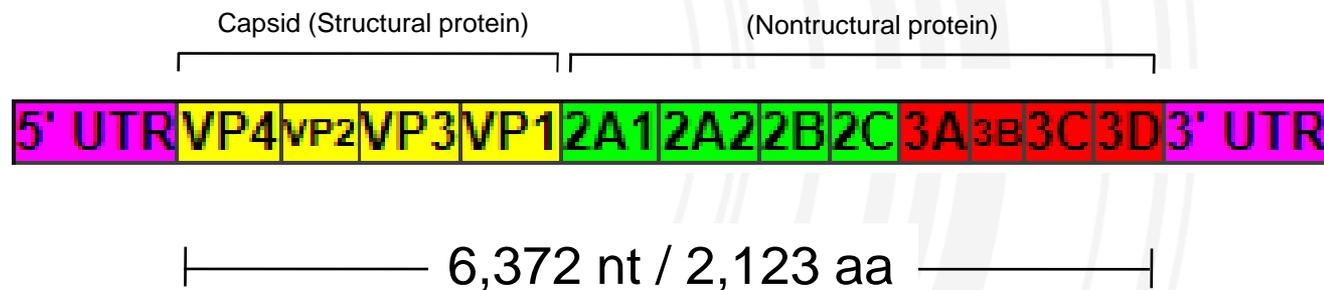


➤ Human Cosavirus detection : 1/411 (0.2%)

CMH-N199-11 3-year-old boy admitted with acute gastroenteritis at Nakornping Hospital, December 2011.

Clinical symptoms Low-grade fever, nausea, vomiting, mild dehydration, diarrheal episodes = 3/day

➤ CMH-N199-11 : Complete genome sequence analysis



NCBI accession number : KJ396940

<http://www.ncbi.nlm.nih.gov/nuccore/613487339?report=genbank>

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Search Limits Advanced Help

Display Settings: GenBank Send:

Human cosavirus A strain CMH-N199-11 polyprotein gene, complete cds

GenBank: KJ396940.1
[FASTA](#) [Graphics](#)

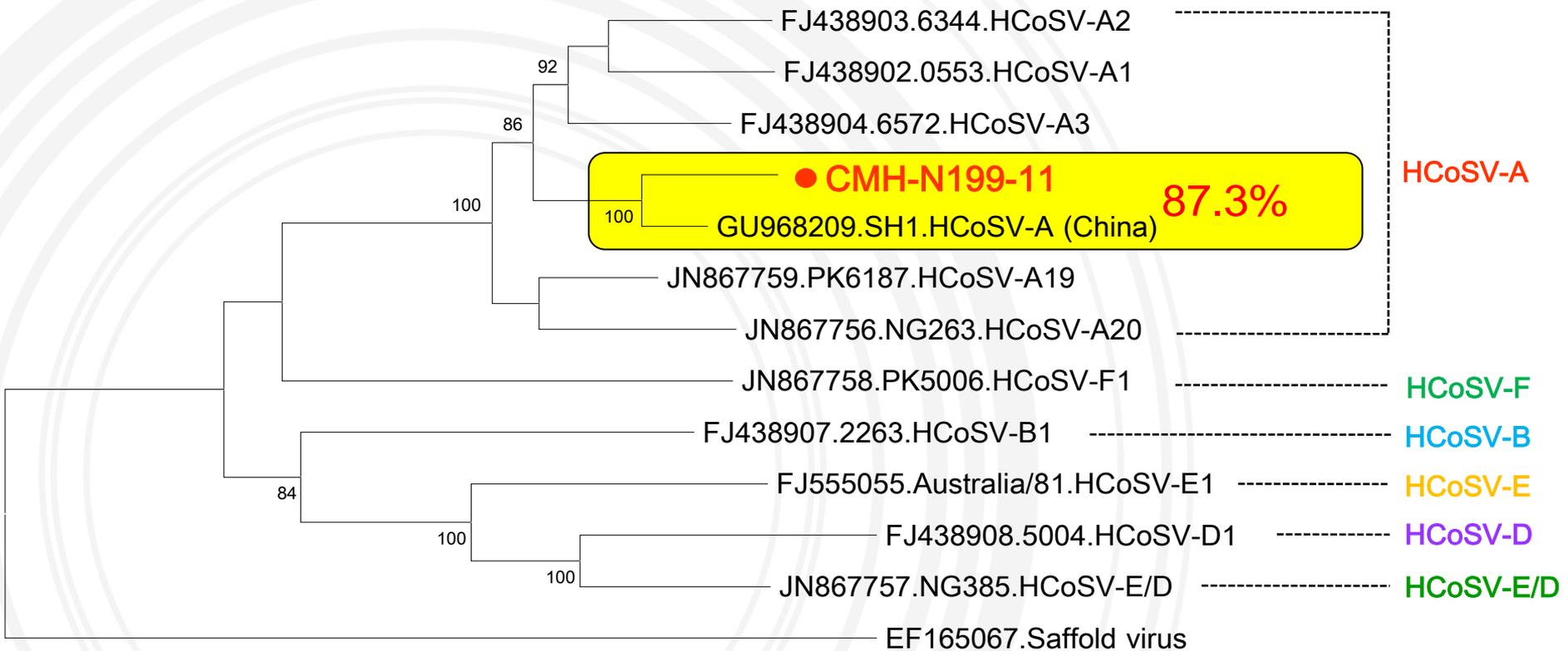
Go to:

LOCUS KJ396940 7317 bp RNA linear VRL 23-APR-2014
DEFINITION Human cosavirus A strain CMH-N199-11 polyprotein gene, complete cds.
ACCESSION KJ396940
VERSION KJ396940.1 GI:613487339
KEYWORDS .
SOURCE Human cosavirus A
ORGANISM [Human cosavirus A](#)
Viruses; ssRNA positive-strand viruses, no DNA stage; Picornavirales; Picornaviridae; Cosavirus.
REFERENCE 1 (bases 1 to 7317)
AUTHORS Khamrin,P. and Maneekarn,N.
TITLE Detection and genetic characterization of cosavirus in a pediatric patient with diarrhea
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 7317)
AUTHORS Khamrin,P. and Maneekarn,N.
TITLE Direct Submission
JOURNAL Submitted (31-JAN-2014) Department of Microbiology, Faculty of Medicine, Chiang Mai University, 110 Intawarorose, Sripoom, Muang, Chiang Mai 50200, Thailand
COMMENT ##Assembly-Data-START##

Change region shown
Customize view
Analyze this sequence
Run BLAST
Pick Primers
Highlight Sequence Features
Find in this Sequence
Related information
Protein
RefSeq Genome for Species
Taxonomy
Recent activity
Turn Off Clear
Human cosavirus A strain CMH-N199-11 polyprotein gene, complete cds Nucleotide
Phthonerodes scotarcha voucher CCDB-

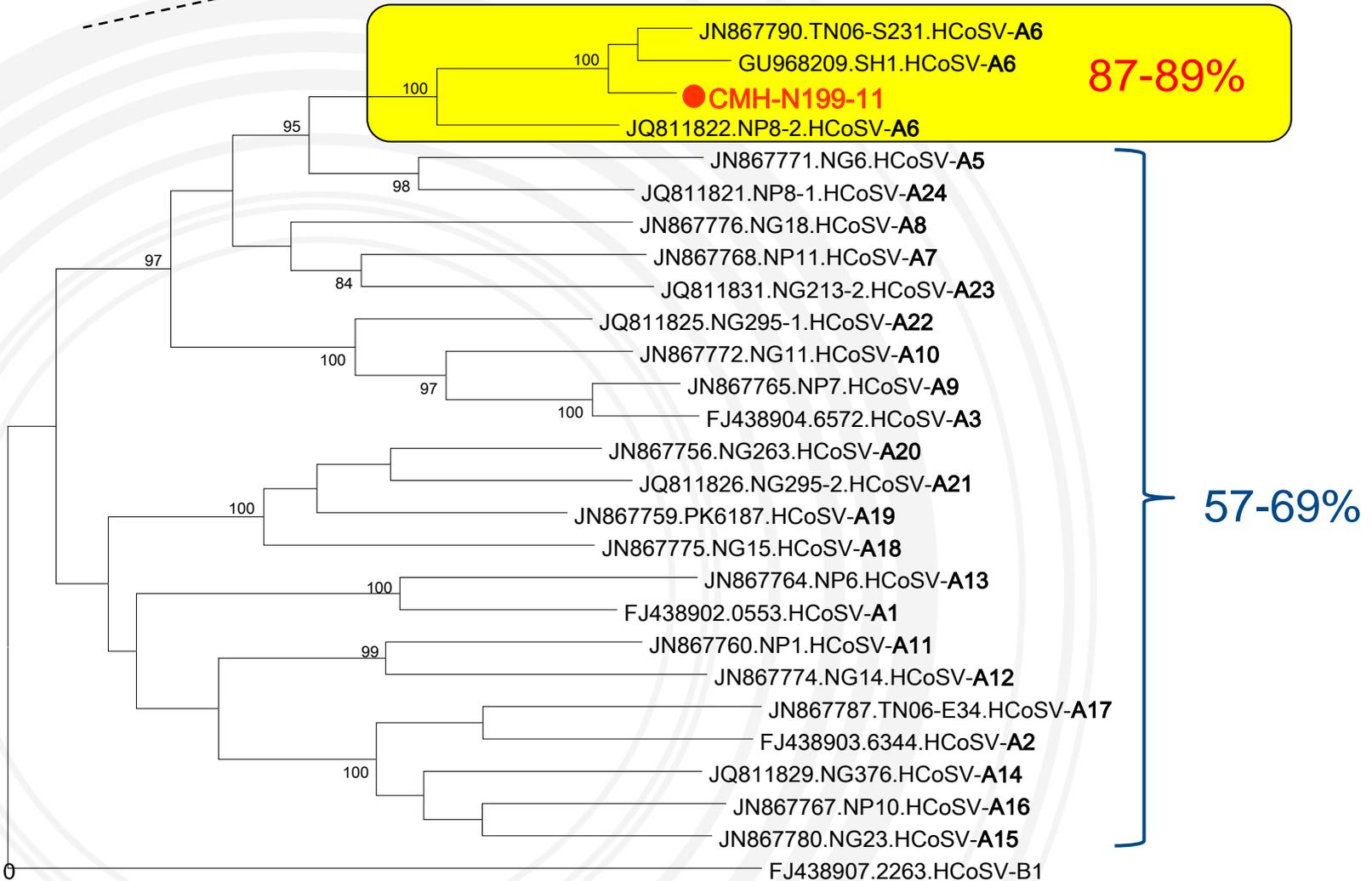
Phylogenetic analysis of complete coding sequence

VP4 VP2 VP3 VP1 2A 2A2 2B 2C 3A 3B 3C 3D



0.1

Phylogenetic analysis of VP1 sequence for HCoSV-A genotyping

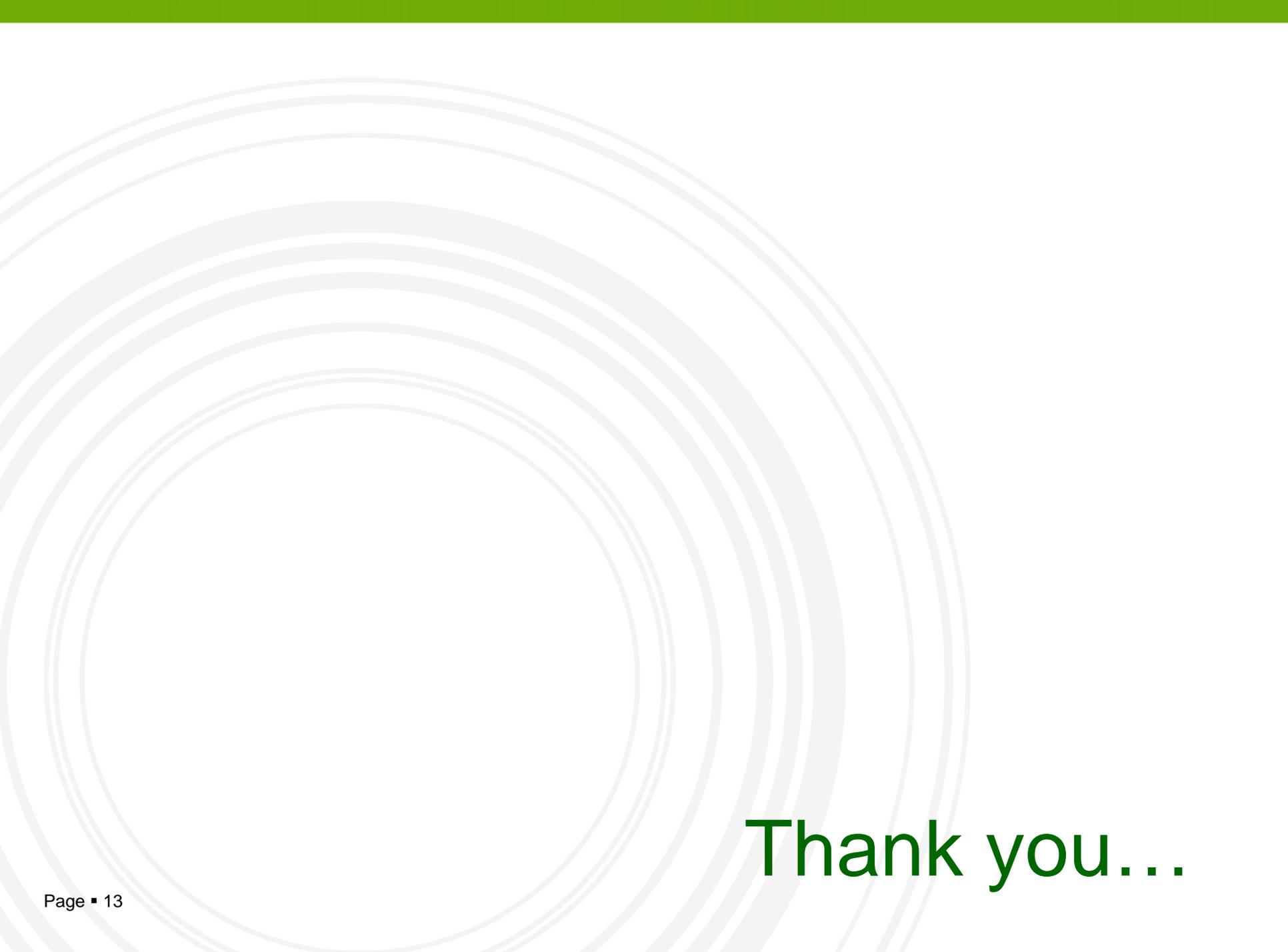


Conclusions

- HCoV was found in a single specimen collected from a 3-year-old boy with acute gastroenteritis.
- Analysis of complete genome sequence revealed that this HCoV strain was HCoV-A6 and most closely related to the Chinese HCoV strain.
- This study is the first report of HCoV infection in children with diarrhea in Thailand.
- Our results emphasize the need for further research to understand the distribution, genetic diversity, and viral pathogenicity of HCoV in humans.

Acknowledgements

- ❑ This research was supported by grants from The Thailand Research Fund (TRF), The Office of Higher Education Commission (OHEC), and the Faculty of Medicine, Chiang Mai University (CMU), Thailand.
- ❑ Professor Niwat Maneeakarn, Department of Microbiology, Faculty of Medicine, Chiang Mai University, Thailand



Thank you...

Nucleotide

Display Settings: [GenBank](#)

Human cosavirus A strain CMH-N199-11 polyprotein gene, complete cds

GenBank: KJ396940.1

[FASTA](#) [Graphics](#)

[Go to:](#)

LOCUS KJ396940 7317 bp RNA linear VRL 23-SEP-2014
 DEFINITION Human cosavirus A strain CMH-N199-11 polyprotein gene, complete cds.
 ACCESSION KJ396940
 VERSION KJ396940.1 GI:613487339
 KEYWORDS .
 SOURCE Human cosavirus A
 ORGANISM [Human cosavirus A](#)
 Viruses; ssRNA viruses; ssRNA positive-strand viruses, no DNA stage; Picornavirales; Picornaviridae; Cosavirus.
 REFERENCE 1 (bases 1 to 7317)
 AUTHORS Khamrin,P. and Maneekarn,N.
 TITLE Detection and genetic characterization of cosavirus in a pediatric patient with diarrhea
 JOURNAL Arch. Virol. 159 (9), 2485-2489 (2014)
 PUBMED [24788846](#)
 REFERENCE 2 (bases 1 to 7317)
 AUTHORS Khamrin,P. and Maneekarn,N.
 TITLE Direct Submission
 JOURNAL Submitted (31-JAN-2014) Department of Microbiology, Faculty of Medicine, Chiang Mai University, 110 Intawarorose, Sriphoom, Muang, Chiang Mai 50200, Thailand
 COMMENT ##Assembly-Data-START##
 Sequencing Technology :: Sanger dideoxy sequencing
 ##Assembly-Data-END##

FEATURES Location/Qualifiers
 source 1..7317
 /organism="Human cosavirus A"
 /mol_type="genomic RNA"
 /strain="CMH-N199-11"
 /host="Homo sapiens"
 /db_xref="taxon:[586419](#)"
 /country="Thailand"
 /collection_date="Dec-2011"
 CDS 863..7234
 /codon_start=1
 /product="polyprotein"
 /protein_id="[AHX26969_1](#)"
 /db_xref="GI:613487340"
 /translation="MGANNSKESVSSNGNEGTIVNNFYSNQYYASIDASAQVGTSTT
 PENGNVSGFLGLASSAFNALSFLASPRVENSTYQEDRILSRVAGNTSINSQASEGVLC
 AYAHETDGSDPTSCGDTPSKGTAAATDRSFIVLQQLQWSTTQAAYDAQWVRLTQELRADN
 KGNVFLKNLKTHSYAKMGFEVTLQVNTSPFHCLVGLFLVPEFTFRFGPDNLTWQDVT
 RLSLVNNTALYDPQNYAGDRALD TDGSDMADFSPEQFMLFPHQFINPKDNIATVRV
 PYVNIAPTSDTTHVTWTA VVMVIVPLYYSSGASPTVMNMTITPVNSVFNGLRHSQAQ
 TPIPTRPFHNFNFQNTT VPLRTEPCYGMTVTPPVDMYPPVEDLMSLAQVPCFLSVAF
 SDNSIKGFPYFVAVSNT EQGRKLFSSGVVSDPHYQHTLLSNLARFFCNYRGSIQDFV
 AATTALTRGKLLIAYSPPGANEPASIDQAMMGTYAIWDLGLQSTFNFVVPFISATDFR
 YNQVAIASALNSDGLWTVWMLNPLTYPPNTPPTQQIVMMLSAGKDFSRFLPISPGIAQ
 GIGEQDPHDNMECGQVDDK DASMFGSHSVSLPTPHTLCSFFYDRYRMIGIVRSTNRNS
 PKLKTPEAGKIMNLVKYFKIESAYTPQPEHAIGLTPVPLNGVPLSAFLNAKSTASK
 KRLRLVKGFQDLYNCCPFYAKFDLEFTVVPASFASDYIVHWYPPGAPIDTAEIFG
 MNATDTEFQDNGENMSSAQLSYNPTFYARGPTKVCVAVVPCLPTSLPLLYFDGYPDYA
 KTPGLYGTSPASTFGTIFVESDTEARFHVYIRYKFKGYLPRPLIRTTIHLASRNRR
 QVAGSVLTREITREERVARQLLLISGDIESNPGPAFNPEYTVHGPVTDLIQLARKPE
 TVDNVNRLLTTLNLMKWNLLKETVTDVAVFLRDMVCLLVKLTSLLYLVHGQGPAYF
 AAASILMADGISFFDWEYKIKIFMARKLRVSPFPFSAQGPDLRDFVTFNAARGAQW
 MIDSLKSLISWIKQWLELEEENEAVQLEKMLIESPRHCKAINDYNRGDSFQRPTNSFE
 FMDRLVECATKLRVQIATYFRNFTTADSDTSRPEPVVVLVRGKPGAGKSAAATVIAA
 AVSKLLVGSQSVYTLSPDTEHMDGYHGQFVTLMDDLGQNPDGEDFRFCQMVSCAQYR
 PAMADLKDKGILFTSRLLIATTNLPDFNPVTISDPRALDRRITFDIHVTPGSAATKNG
 KLDLAAALKPDGPGEPHYTSDCPILHTTGLLLKLNLRNNQTMNVKDLVLDIVRRIKHKK
 EVGNMLDSLVAQGPTMIVGYTKDDDGIAIVDCL EEWNKIKDKKKKQLALEMVAQELKD
 KHEEHKGTIKLKMFTVGLGVVAAVAGAYATMKYFTKDKPEETKEPEEKNEERESKE
 AAGPYNGPTKKEIKTLKKAQSPLMDMEKKSQNVMPFQIWIYNGKKYQSCLAIGKRV
 ILVNKHAFAFESVDNKFIVNQKEYTLQVTAISLDCGSGVTDVCAICLPPGPDFKSIKKH
 FLFPNTTMFPGTRLTILSNHYPMREGSFLRF EDEVPTNIGQMPFVMLYKSTSYFGM
 CGSVVCSRFVDDGGGIGMHCAGGGGVS VGTRLTARMIESVFDYFYPVAQGIIENTET
 GPRVHVPRQSKLKRTHATYPATDKFGPAALSRYDPRLNEGVDLDTVIFSKHHTNLTID
 KESTFRKALDMAAEIYGEKFRGNDFSPLSVEDAILGVPGLDRLDNPNTASGLPYTKTRR
 QMIDFNTGQLDLDLTKCRLLGQWLAGRPPQEVHYQTF LKDEIRPIEKVKAGKTRIIDVP
 PLDHVIAFRMLFGKFAHYHLNFGFKTGSAGCDPDVAWASFGFELSGFPYLYDFDYS
 NFDASHSTSFELDQKFFSPELGFDPKCSLLKSLAVSTHCYENKRLQIAGGLPSGT
 AGTSVLNTVINNIIFHGALYHTYTNFEWDDISMLAYGDDIVVASKFELDLVMVKAFMN

RIGYKITPADKSDFRPKMDDICFLKRRFVKVAGVWAPVMTENLEAMLSWYKPGTL
NEKLQSVSRLAHFSGRDVEDLRFKPFIRDGFDVTPWKQLHLEWLNKLSA"

ORIGIN

1 ggctatgtca aagcaatata gttgtgtcaa ggcgtacacg ctcaaggcgg agacttaggc
61 ctacacagatt gttgtttgtg ttattggatg ctggatggtc acgttggaga ctgcatgtgg
121 cagtcttgaa acgtgtggtt tgacgtctat ccattacggc agtgggtgga gtactgcaaa
181 gatgtcaccg tgccttacac ggtttttgaa cccccaccg gctgtttggc gcttgcagga
241 cagcagggtt attttcttat gttctttatt tctagccagc agggttctat cctgttgggc
301 ggagtatac tcccgttctt tcttggacag attgctcca cgaccttgtt ggatctcaag
361 gtgatcaagt cactgggtgaa tagagcgaag gttgaggaga cctgaggaat ttccatgtgg
421 ctttgccagg agttgtagcg atgctgtgtg tgtgtgcgga tttccccca tggcaacatg
481 agcctcacag gccgaaagcc ctgtccgaaa ggaccacac agtggagcaa cccagctcc
541 ctctacaaa gctttgtgag aatgaactca cgtttattct tctttattct ctatttcat
601 caggcccaaa agatgtcctg aaggtaacct gtgtatctgg gcatgagcac catcaactac
661 cgggacttgc actttgtgtc agacgatgt ggtgaccag cccctctgtc tcggcagagg
721 ggctttcgct ccttcagcac gagatctgat caggagcccc tcccagtggt ctttacacct
781 ggcggggggt taaaaattgc ccaaggcctg gcacaataac ctaggggact aggtttacct
841 tttgttaaca atgtctgtca ttatgggtgc aaacaacagc aaagaatcag tgtccagcaa
901 tggcaacgag gaaacaattg ttaataactt ttattcaaac cagtattatg cttctattga
961 tgcttctgcc caagggtttg ggacctctac tactcctgaa aacggcaacg tatctggctt
1021 tcttggactt gcaagcagtg cttttaaagc tctctctttt ctgcctccc cgcgagtaga
1081 gaactcaaca tatcaggaag ataggatttt atcgcgcgta gcaggcaata catctataaa
1141 ctctcaagct tcagaagggg ttttgtgtgc ttacgcgatg gaaacgacg gttcagatcc
1201 tacttcttgt ggtgacaccc ccagcaagg aactcagca acagatagaa gtttcatagt
1261 acaactccag caatgggtcca caacacagcg cgcataatgat gctcagtggg tccgcctcac
1321 acaagaactg cggggggata acaaggtaa tgtgtttgca aaaaatctga aaactcactc
1381 atatgcaaag atgggattcg aggtaaacct tcaggtaaac acctctccgt tccattgtgg
1441 tttggttggc ctgtttcttg tccccgaatt caccagattc gggccagaca acttgacatg
1501 gcaggatgta acaaccgac ttttcttctg gaacaatact gctttatattg acccccaaaa
1561 ttatgcagga gataggcctt tagacactga tggctctttt gatatggcgg atttttcgcc
1621 tgaacaattc atgtctgttc cccaccagtt tataaatccc aaagacacta atattgctac
1681 tgttagagtc ccatatgtta atatagcccc tactagtgat accactgttc atacagtttg
1741 gacagcggtg gtaatgttga tagttccccct ctactattca tcagggtgctt ctccaacagt
1801 taatatgaca atgactatta ctctctgtgaa ttcagtggtt aatggactgc gtcattcggc
1861 gcagaccctt atcttacta ggctttttca caactttaac cagttcaaca ccacggtgcc
1921 tcttcgacc gagccatgct atggatgac tgtgacccct ccagttgatt acatgcctcc
1981 accagttgaa gatctcatgt ctttggccca ggtgccttgc ttctgtcgg ttgccctttc
2041 tgacaatagt attaaagggt ttcttactt tgctgttagc aatacggaac aagggagaaa
2101 attgttttct agtgggttag tgttatcaga tccccactat caacacactt tgccttcaaa
2161 tttggctcgc ttcttctgca attatcgcgg aagcatacaa tttgattttg ttgccgccc
2221 aactgcctg acccgtggta agttgcttat agcttattct cctccaggtg ccaatgagcc
2281 cgcatctatt gaccaggcta tgatggggac ttatgcaatt tgggatcctg gactgcagtc
2341 aactttcaat tttgtgttcc ctttcatatc ggcaactgac tttcgtaca atcaagtagc
2401 aatcgtctca gctttgaatt ctgatggatg gcttacagtt tggatgctta acccattgac
2461 ttaccaccc aatacacccc caaccaca gattgtcatg atgcttctg caggtaaagga
2521 cttctcttc agacttccaa tttcaccagg tatagcaca gggattggcg aacaagacc
2581 acatgacaat atggagtggt gacaggttga tgacaaagat gcgtccatgt tctccggaca
2641 tctgttttct ctccccacgc ctacacact gtgttctttt tctatgaca gataccgat
2701 gattggaatt gtgcgttcaa caaacagaaa ttcacaaaaa cttaaagacac ctctggaagc
2761 aggcaaaata atgaaatttg tcaaatattt caagattgaa tcggcatata caccacaacc
2821 agaacatgct atcgggtctaa ctccagtagc taacttgaat ggagttccct tgcctcctt
2881 ccttaatgcc aaacttactg catccaagaa aaggttaaga cttgttaaag gagactttca
2941 gttgaattat tgcgtctctt ttacttatgc taagtttgat cttgaattta cagtggtgcc
3001 ccctgctagc tttgcatctg attatattatg tctactggtat cctccaggtg cccctattga
3061 tactgctgag atcatgtttg gaatgaatgc tactgacact gagtttcaag acaatggaga
3121 aaacatgagt tcagctcaac tctcatataa ccctactttt tacgccagag gacctactaa
3181 agtctgtgca gtagtacctt tttgtttacc tacttctcta ttgcctcttt attttgatgg
3241 ttatccagac tatgcaaaaa ctctgggctt gtatgggact tctccagctt ctacccttgg
3301 tactattttt gttgaatcag acacagaagc acgcttccat gtttacatca ggtcaaaaaa
3361 tttcaagggt tactacccta gaccactaat tagaacaaca cacataacct tggcaagtcg
3421 caacagaaga caggtggcgg gtagtgtgct tactagagag attactaggg aagaacgaga
3481 agttgccagg caactcctcc ttatttcagg tgatattgag tccaaccag gtcctgcatt
3541 taatccagaa tacactgtac atggcccagt aacagatttg attcaattgg ctagaaaacc
3601 agaaaacagt gataatgtaa acagacttct gactactctt aacaccctta tggctaaatg
3661 gaacaacctt aaagagactg ttacagatgc tgtgtttctt agagacatgg tgtgtcttct
3721 tgaataactt acctctttac tctacctggg tcacggacaa ggaccaggtg cttactttgc
3781 tgctgtctct atattgatgg ctgatgggat atctttcttt gattggtatg aaaagattaa
3841 aatcttcagc gcaagaaaaa tcagagtttc ccccccttc tttccagtg cccagggacc
3901 agatcttaga gattttgtga ctttcttcaa cgctgcgctg ggagcgcagt ggatgattga
3961 ttctcttaaa tccctcatat cttggattaa acagtggtct gaacttgagg aagaaaaatg
4021 agcagtacaa gttgaaaaga tgtttaattga atctcccaga cactgcaagg caataaatga
4081 ttacaacaga cgtgactctt tccaacgacc gactaactct tttgaattca tggacagact
4141 tgtggaatgt gctactaaac tcgggggggt tcaaatgca acatacttta gaaattcac
4201 cacagccgat tctgatacta gcagaccaga gccggtggtg cttgttctgc gcgggaaacc
4261 aggcgcaagg caactcagctg cagctactgt catcgcctgt cagctctcca aattgttagt
4321 aggaagtcaa tcagtttata cactttctcc agatacagaa cacatggatg gatcatgg
4381 tcaatttgtg acctgtatgg atgaccttgg acaaaaacca gacggtgaag acttcagatg
4441 tttctgtcaa atggtttcat gtgctcagta ccgccccgcc atggctgatc ttaaagacaa
4501 gggaaatcctg tttacttcta gacttttgat tgctaccacc aatttaccag acttcaatcc
4561 cgttactatt tcagatcttc gcgcccctga caggcgcatt accttcgata tccacgttac
4621 accaggtctc gctgcaacca agaattgggaa actcgcacttg gctgcggccc tcaaacccga
4681 tggaccgggt gaacacctt acacttctga ctgtcccatt ctccacacta ctggacttct
4741 tttgaaaaac ctcaaaaaa accagactat gaatgtaaa gaccttggg acccttctgt
4801 gagaagaatt aaacataaga aagaatgagg aaacatgctt gactctctag ttgccaggg
4861 acctacaatg atgttggct atactaaagga tgatgatggc attgcgattg tggactgtct
4921 cgaagaatgg aacaagatta aggacaagaa aaagaacacag cttgctttgg aaatggttg
4981 ccaggagctt aaagacaaac atgaagaaca taagggaaca attaaacttt tgaaaatgtt
5041 tgtcacaggt ttgggtgtcg tcgctgcctg tgcaggagcg tacgccaaa tgaaatattt
5101 cactaaagac aaaccagagg aaaccaagga accagaggaa aagaatgaag aaagagaaga

```

5161 gtccaaagaa gctgctggtc cgtacaatgg gccacaag aaagaaatta aaacattgaa
5221 gctcaaggcc cagagtcctc ttatggatat ggaaaagaag atctctcaga atgttatgcc
5281 ctttcagatt tgggtacaacg gaaagaagta taccagctct tgtttggcaa taggtaaaag
5341 agtgattcct gtaaacaaac atgcttttga atcagtgatg aacaaattca ttgtgaatca
5401 gaaagaatac acactagatc aggttacagc tatttccctt gattgtgggt caggtgttac
5461 agatgtgtgt gctatttgcg tgcccccagg cccagatttc aaatcaataa agaaacactt
5521 cttacctttt aacactacta tgtttccagg caccgcctg actattctct cgaatgacca
5581 ctaccctatg tccagagaag ggtctttcct tcgctttgag gatgaagtgc ccaccaacat
5641 aggtcagatg ccctttgtga tgctttacaa atcaacctct tactttggta tegtgtggtc
5701 agttgtgtgt agtagatttg ttgatggagg aggaataatt ggaatgcact gtgcaggtgg
5761 aggcggagtc agtgttggaa cccgtttgac tgctagaatg attgaatcag tttttgatta
5821 tttctacca cccagttgctc aggggaattat tgaaaataca gagacaggac cccgtgtgca
5881 tgtgcccaga caatccaaat taaaaagaac tcatgccact tatccggcaa cggacaagtt
5941 tggcccagct gcactctcgc ggtacgatcc gcgtcttaat gaaggagtag accttgatac
6001 agtgatcttt tcaaaacaca cccataatac ttgtattgat aaagaatcca ctttaggaa
6061 agctcttgat atggcagcag aaatttatgg tgaaaagttc agaggaaatg atttctcccc
6121 tctttcagtt gaagatgcaa ttcttggagt acccggactc gaccgtttgg accccaacac
6181 agcttcagga cttccttaca caaagaccag aagacagatg attgacttca acacaggaca
6241 gattttggac gatactctta agtgtcgact tggacaatgg cttgctggac gacctcctca
6301 ggaagtacat taccagacat ttctcaagga tgaaatcagg cccatcgaag aggtcaaagc
6361 aggaaaaaac agaataattg atgttctctc tcttgatcat gtcattgcct tcagaatgct
6421 ctttggcaaa ttcattgccc actaccactt gaactttggc tttaaaacag gctctgctat
6481 tggttgtgac ccagatgttg cttgggcttc tttcggcttt gagctcagtg gctttcctta
6541 tctgtatgat tttgattact caaactttga tgcttctcac agcacttcaa tctttgaaat
6601 cttagatcag aaattctttt ctccagaatt aggttttgat cctagatgct cacttctctt
6661 gaaatccctt gcagtttcaa cccactgtta tgagaacaag agactccaga ttgctggagg
6721 gcttccctcc ggcacggcag gtacctcagt actgaacacc gtgataaaca atatcatttt
6781 tcacggtgca ttgtatcaca cttacactaa ttttgagtgg gatgacatca gtatgttagc
6841 ctacggcgat gacattgttg ttgcctccea attcgaactt gacttgggta tggttaaggc
6901 tttcatgaac cggattggct acaagattac ccctgcagac aaaagtgacg aattcagacc
6961 aaagtgtatg gatgacattt gcttcttaaa gaggcgtttt gttaaagttg ctggagtttg
7021 ggctccagtg atggaactg aaaaactcga ggcaatgttg tcttggtaga aaccaggaaac
7081 tcttaatgaa aagctccaga gtgtctcaag acttgcccat ttctcaggac gtgacgtata
7141 tgaccgccct ttcaagccct tcattcgtga tgggtttgac gtcacacctt ggaacagatt
7201 acacttggaa tggcttaata agttatcagc ttaaagaatt ttgaattggc atttcaaatt
7261 tattttgaat ttggctttta attcgttctt aatttgggta tttattgggt ataataca

```

//