

02738473-61

Computer Application
in Biological Science

ครั้งที่ 10 เรื่อง Bioinformatics part II

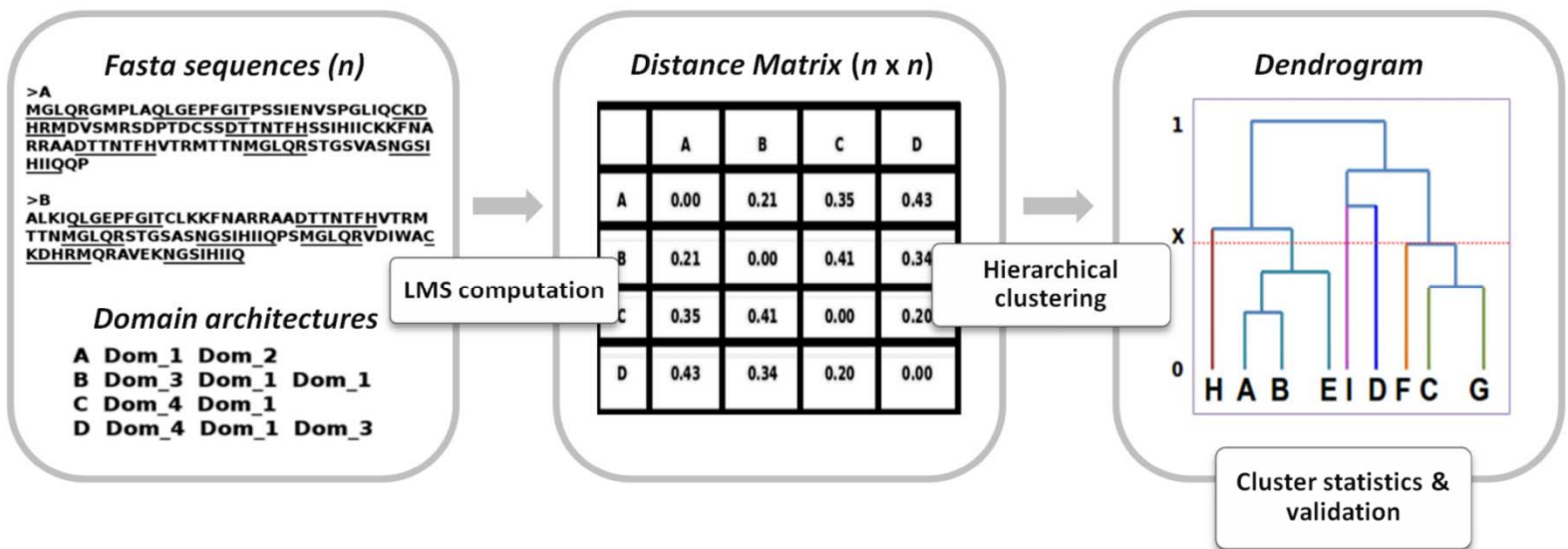
19-23 กันยายน 65

ต่อจากสัปดาห์ที่แล้ว เรากำลังจะสร้าง cladogram หรือ phylogenetic tree เพื่อเปรียบเทียบลำดับวิวัฒนาการของสิ่งมีชีวิต โดยอาศัยข้อมูลลำดับกรดอะมิโนองค์ประกอบของโปรตีนคือ เอนไซม์อะไมเลส (amylase)

เรามาดูขั้นตอนการสร้างไฟล์ text ที่ประกอบด้วยลำดับกรดอะมิโนของสัตว์ 4 ชนิด ในรูป FASTA format แล้ว

สำหรับสัปดาห์นี้ จะนำข้อมูลไปวิเคราะห์ด้วยโปรแกรม CLAP

โดยเข้าใช้งานได้ที่ <http://nslab.mbu.iisc.ernet.in/clap/index.html>



โปรแกรม CLAP ดูแลและบำรุงรักษาโดยทีมวิจัยที่ Institute of Science ประเทศอินเดีย เมื่อเข้าหน้าเว็บของ CLAP แล้ว ให้คลิกเข้าใช้งานได้ที่ลูกศรชี้ครับ

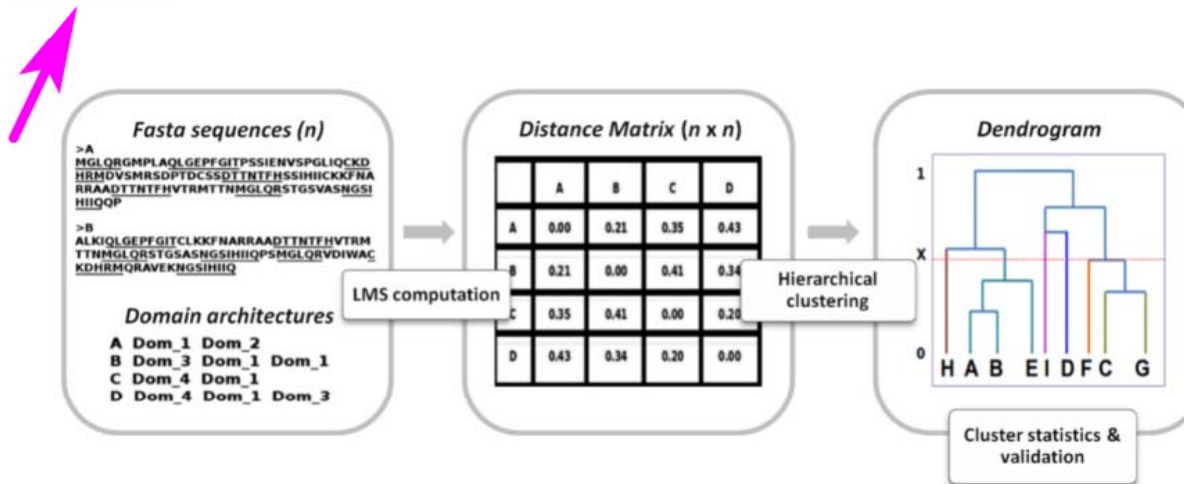
CLAP *CLAssification of Proteins*

Home CLAP Help Contact Us FAQ

Classification of Proteins (CLAP)

CLAP is a web server for automatic classification of protein sequences. It uses an alignment free approach to compute local similarities among sequences [Martin J *et al.*, 2010]. This method is particularly useful for comparing multi-domain protein sequences which are difficult to align [Bhaskara RM *et al.*, 2014]. CLAP can handle multiple sequences of varied lengths, resulting in clusters with high functional and domain architectural similarities.

Click [here](#) to use.



References

การใช้งานง่ายมาก แค่ Copy ข้อมูลจากไฟล์ text ที่เราสร้างไว้แล้วมาวางในช่องบน หรือกดปุ่มเลือกไฟล์เพื่ออัปโหลดเลยก็ได้ แล้วกดปุ่ม Submit ด้านล่างสุด ส่วนช่องด้านล่าง ใช้ในกรณีที่เรามีข้อมูล domain architecture ของโปรตีนต่าง ๆ ซึ่งขอข้ามไปไม่กล่าวถึงในวิชานี้ครับ หากนิสิตสนใจสามารถศึกษาเพิ่มเติมต่อไปได้

CLAP *CLAssification of Proteins*

Home CLAP Help Contact Us FAQ

Generate tree based on Local Matching Score (LMS)

Paste Query Sequences*

Or Upload from file* : [Sample Input](#)

*The amino acid sequences should be in FASTA format. CLAP is an alignment-free tool to compare sequences and hence requires at-least two sequences as input.

Generate clusters based on distance cut-off

No cut-off Tree parsing cut-off (0.0 to 1.0):

Obtain representative domain architecture for each cluster (Optional)

Paste domain architecture file (tab delimited format)

Or Upload from file : [Sample Input](#)

Email (Optional): e.g : user@xmail.com

กรณีที่คุณใช้วิธี Copy มา Paste ก็จะมีหน้าต่างแบบนี้ครับ
แต่ถ้ามีไฟล์ text อยู่ในคอมพิวเตอร์แล้ว ก็กดปุ่มเลือกไฟล์มาโดยตรงก็สะดวกดี
เรียบร้อยแล้วก็กดปุ่ม Submit ว่างสุดของจอ

Generate tree based on Local Matching Score (LMS)

Paste Query Sequences*

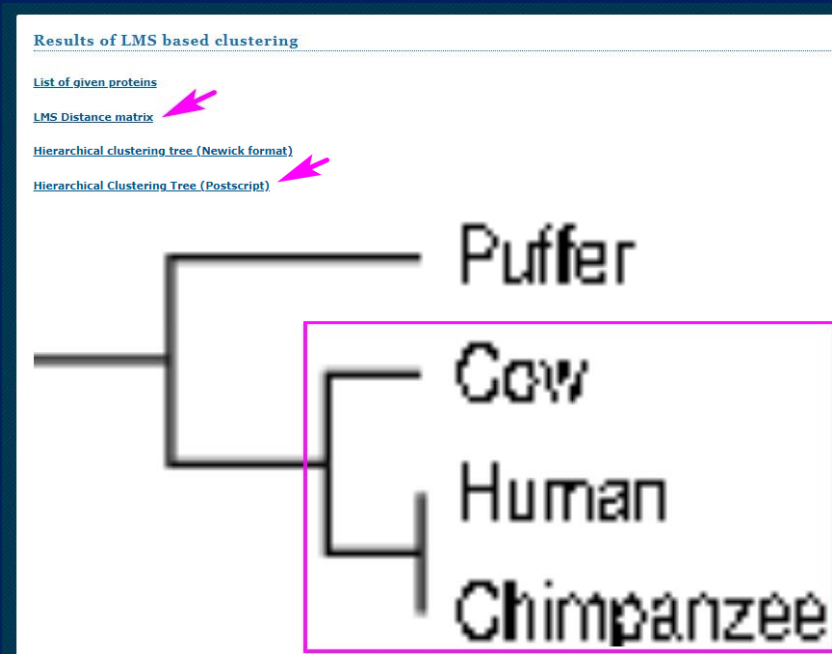
```
> Puffer  
MRFFILVTLFGLSLAQHTPHFKHGRTSIVHLFEWRWTDIAAECERFLGPNGFAGVQISPPNEHIILDSPW  
RPWWQRYQPIGYNLCSSRSGSENELRDMITRCNNVGVNIYVDVVINI IMCRASSGEGTI ISSCGSWFNAGT  
KD  
FPSVSYSLDFNDHKCRTGSGNIESYDADVNQVRDCRLENLLDLALEKDYVRGKVADYMNRLVDMGVAG  
F  
RVDACKHMWPGDLSAVYGRLNNLNTNWFPGGSRPFYQEVIDLGSELITAKEYFHLGRVTERKYGAKLGT  
VFRKWNNEKLMYTKNWGEGWGFMPNGNAVVFIDNHDNQRGHGAGGAAIVTFWDFRLHKMAVAYMLA  
HPYG
```

Or Upload from file* : ไม่ได้เลือกไฟล์ใด

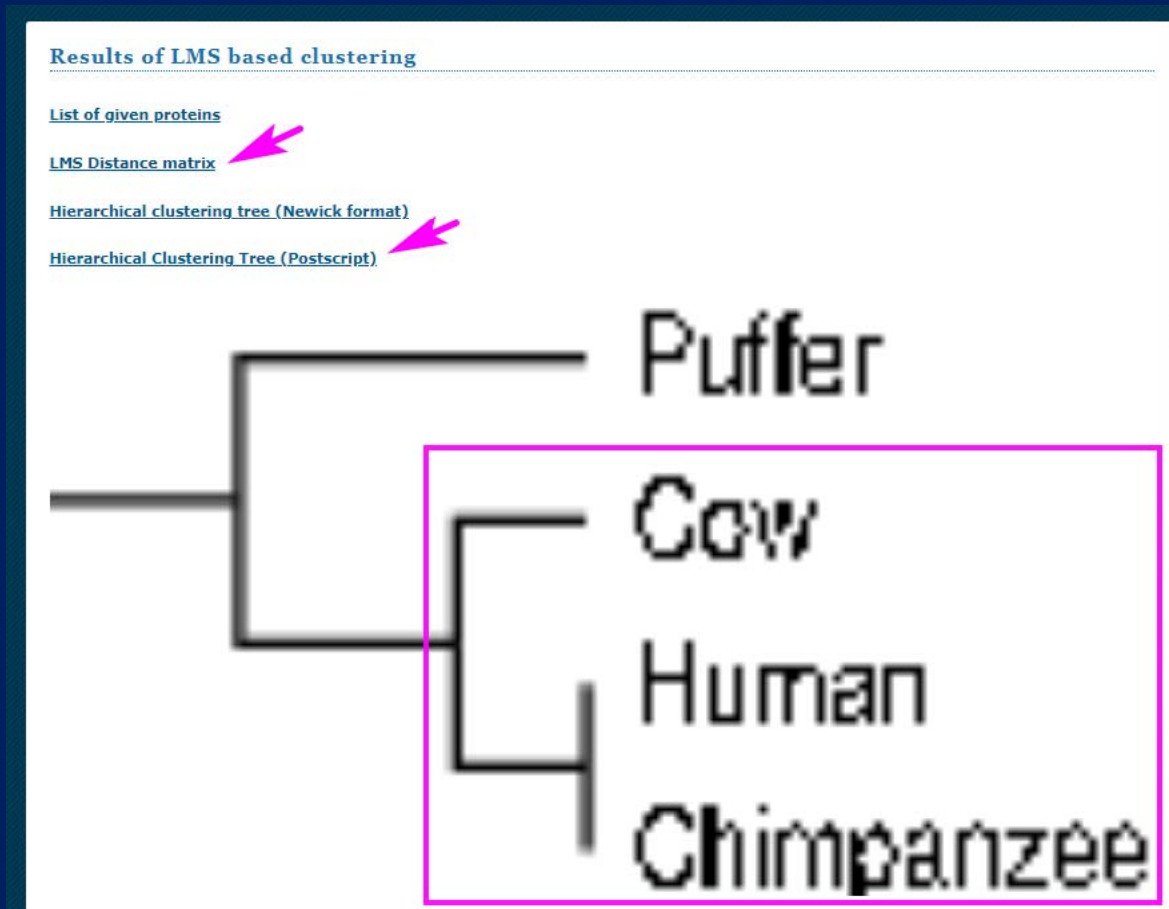
[Sample Input](#)

ก็จะได้ cladogram ออกมา จะเห็นว่า ปลาปักเป้า ไม่เข้าพวก แยกกลุ่มกับ 3 ตัวที่เหลือ ตั้งแต่ช่วงแรก ๆ ของวิวัฒนาการ เป็นปลา กับ สัตว์เลี้ยงลูกด้วยนม ส่วนคนกับลิงชิมแปนซีก็แยกออกจากวัว ในเวลาต่อมา วัวไม่มีนิ้วมีนิ้วตีน ส่วนคนกับลิงเป็นไพรเมต มีนิ้ว และหัวแม่มือพับเข้าหาฝ่ามือได้ ฯลฯ คงพอเป็นไอเดียเบื้องต้นนะครับ ว่าเขาสร้างแผนผังแบบนี้ได้อย่างไร โดยใช้ข้อมูลระดับโมเลกุล

อ้อ เวลาอ่านแผนผังแบบนี้ อย่าพูดว่า คนวิวัฒนาการมาจากวัวนะครับ แต่คนกับวัวมีบรรพบุรุษร่วมกัน ซึ่งไม่ทราบว่าเป็นสัตว์อะไร ในทำนองเดียวกัน คนกับลิงชิมแปนซีก็มีบรรพบุรุษร่วมกัน (Common ancestor)



สำหรับลิงก์ที่ทำลูกศรชี้ไว้ อันบนคือ LMS Distance matrix
จะเป็นข้อมูลค่าที่คำนวณได้ ในลักษณะเปรียบเทียบแต่ละคู่เป็น matrix
ตัวเลขยิ่งมาก แสดงว่า คู่ นั้นห่างไกลกันทางวิวัฒนาการมาก (มีโปรตีนที่ต่างกันมาก)
ลองคลิกดูครับ



ก็จะได้ข้อมูลในรูปตารางออกมา แต่ตารางจะโย้ไปเย้มา ดูยาก
 นิสิตอาจจะ Copy ไปวางในโปรแกรมเอกสารเช่น Word ก็ได้
 เพราะข้อมูลแต่ละตัวมี Tab คั่นอยู่ เอาไปเคาะ ๆ จัดใหม่ได้ครับ ก็จะสวยแบบภาพล่าง

	Puffer	Cow	Human	Chimpanzee
Puffer	0.000000	0.638791	0.631773	0.627564
Cow	0.638791	0.000000	0.247483	0.252782
Human	0.631773	0.247483	0.000000	0.016573
Chimpanzee	0.627564	0.252782	0.016573	0.000000

	Puffer	Cow	Human	Chimpanzee
Puffer	0.000000	0.638791	0.631773	0.627564
Cow	0.638791	0.000000	0.247483	0.252782
Human	0.631773	0.247483	0.000000	0.016573
Chimpanzee	0.627564	0.252782	0.016573	0.000000

ข้อมูลนี้บอกอะไรเราบ้าง

ลองดูความต่างระหว่างวัวกับคน คือ 0.247483

ส่วนความต่างระหว่างวัวกับลิงชิมแปนซี คือ 0.252782

แสดงว่า วัวใกล้เคียงกับคน มากกว่า ลิงชิมแปนซี

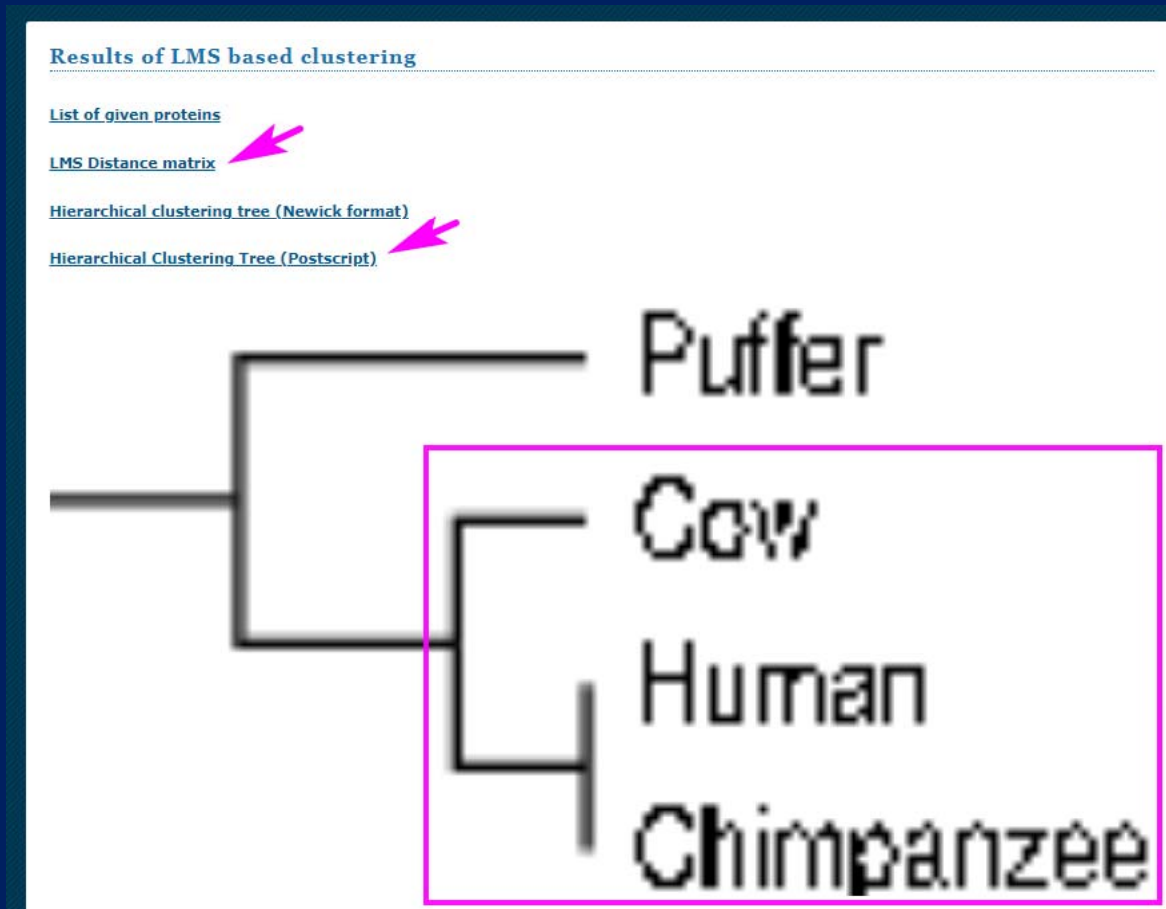
นั่นคือ คนใกล้เคียงกับวัว มากกว่าที่ลิงชิมแปนซีใกล้เคียงกับวัว (จะดีใจดีไหมครับ ที่เราเป็นญาติใกล้กับวัว ส่วนลิงชิมแปนซี ห่างออกไปอีกนิด)

ส่วนคนกับลิงชิมแปนซี ใกล้เคียงกันมาก ห่างกันแค่ 0.016573

ถ้าเริ่มต้นจากวัว ดูเหมือนว่าลิงชิมแปนซีจะออกห่างจากคนไปอีกทีในลำดับวิวัฒนาการนะครับ อันนี้คิดเล่น ๆ เพราะใช้แค่ข้อมูล amylase เพียงอย่างเดียว ไม่ต้องซีเรียสครับ

	Puffer	Cow	Human	Chimpanzee
Puffer	0.000000	0.638791	0.631773	0.627564
Cow	0.638791	0.000000	0.247483	0.252782
Human	0.631773	0.247483	0.000000	0.016573
Chimpanzee	0.627564	0.252782	0.016573	0.000000

ส่วนลิงก์กลาง คือ ไฟล์ภาพแผนผัง cladogram ในรูป Postscript ไม่แตก ไม่เบลอ แม้จะขยายใหญ่ ๆ คล้าย ๆ กับ Windows metafile ที่เราเคยเรียนตอนใช้โปรแกรม Chems sketch แต่จะเป็นไฟล์นามสกุล .eps แต่เนื่องจาก Postscript ต้องใช้โปรแกรมเฉพาะในการเปิดแก้ไขขนาดภาพ เช่น Adobe Photoshop จึงขอไม่พูดถึงรายละเอียด นิสิตสามารถศึกษาต่อได้เองถ้าสนใจ



ถ้าดาวน์โหลดไฟล์ .eps มาเปิดด้วยโปรแกรม Photoshop สามารถย่อขยายได้แบบนี้
ครับ โดยมีความสวยงาม ไม่แตก ไม่เบลอ

