## <u>MEGAによるコロナウイルス系統樹作成</u>

## コロナウイルスの系統樹を描いてみよう。

- 1. コロナウイルスのsmall membrane protein遺伝子(以下、E遺伝子)のサンプル配列 💩 <u>E-protein.fasta</u>を取得する。具体的には リンクを右クリックして、メニューから「名前を付けてリンク先を保存」を選ぶ。
- 2. 保存先を聞いてくるので、保存したい場所に移動する。新規フォルダを作成するのもよい。たとえば、「新しいフォルダ」をクリックして新フォルダを作成し、「corona」という名前にする。
- 3. MEGA6(またはMEGA5)を起動して、サンプル配列のアライメントを行う。

MEGA6 を起動して、「File」→	「Open A File/Session」から	、E-protein.fastaを選択する。
---------------------	-------------------------	------------------------

M MEGA 5.05	
File Analysis Help	
Open A File/Session         The The Session Sector	• Rates • Clocks •
Bedit a Text File Bedit a Text File Format to MEGA	M
Printer Setup	
Exit MEGA Alt+X	GA
First time User? Tutorial Examples Citation MEGA Web Report a Bug Updates? Customize ToolBar → Preferences	. О
MEGA release #5110426	][E-protein.meg.]

4. 何をしたいか聞いてくるので、「Align」を選択する。

M Analy	ze or Align File?	2
2	How would you like to open this fasta file?	
	Analyze Align	

5. 入力した配列をタンパク質に翻訳する。具体的には、「Translated Protein Sequences」タブをクリックする。 (ここで使用している配列は全長がタンパク質コード領域のため、このような操作をしても問題はないが、非コード領域を含む配列を 翻訳するときにはコード領域をしてやる必要がある)

M M5: Alignment Explorer (E-protein.fasta)		
Data Edit Search Alignment Web Sequencer Display Help		
] D 😅 🖬 🎬   🧮 🌆 🎆   ₩ 💔   💥 🐍 ] ∽ 🗈 🌡 🛍 🗙 兴   物 🎒   ◀ 🕨   ] 🗛 🏘 🏘 🐴		
DNA Sequences Translated Protein Sequences		
Species/Abbrv		
1. NC_008315.1_Bat_coronavirus_(BtCoV/133/2005)		
2. NC_014470.1_Bat_coronavirus_BM48-31/BGR/2008 A TOTAC TCATICGITICAGAAGAAGCCGGIACGIIAATA		
3. NC_003045.1_Bovine_coronavirus_E-protein AIGCICAIGCICAIGCICAGECACACACACACIGIGIGGIAIGIG		
4. NC_005147.1_Human_coronavirus_0C43_E-protein AIGCCIGAIGCIIAICIICCAGACACIGIGIGIAIGIG		
5. NC_007732.1_Porcine_hemagglutinating_encepha A I G C I G A I G C I I A I I I I G C A G A C A C I G I G I G G I A I G I G I		
6. NC_012948.1_Bovine_respiratory_coronavirus_AAIGCCIGAIGCTIAIIIICCAGACACIGIGIGIAIGIG		
7. NC_012949.1_Bovine_respiratory_coronavirus_baicscigalscirlaterespiratory		
8. NC_006577.2_Human_coronavirus_HKU1_E-protein AIGGIIGAIIIIAIIIIICAAIGAIACCIGCIIGGIACAIA		
9. NC_005831.2_Human_coronavirus_NL63_E-protein AIGICCIICGAIGAIGAIGAIGAIGAIGAIGAIGAIGAIGAIGAIGAIG		
10. NC_012950.1_Human_enteric_coronavirus_straina IGCCIGAIGCIIAIIIICCAGACACIGIGIGIAIGIG		
11. NC_010800.1_Turkey_coronavirus_E-protein GAGGAGGAGGAGGITICCIAACAGCAGITIACAIATII		
12. NC_001846.1_Murine_hepatitis_virus_strain_AAIGITIAAIIIAIICCIIACAGACAGIAIGGIAIGIG		
13. NC_006852.1_Murine_hepatitis_virus_strain_JAIGITIAAIIIAIICCIIACAGACACAGIAIGETAIGIG		
14. NC_012936.1_Rat_coronavirus_Parker_E-proteina GETERATITATICCTTATAGACACAGTAIGGTACGTG		
15. NC_009020.1_Bat_coronavirus_HKU5-1_E-proteina CCIIIIIGICCAACAACAACAACAACIACAAAIAGGICCAICAAAA		
16. NC_009988.1_Bat_coronavirus_HKU2_E-protein AIGIIICIGAGAIIGAIGAIGAIGAIGAIGAIGAIICIICAII		
17. NC_009021.1_Bat_coronavirus_HKU9-1_E-proteinAIGIAIGATAIIGIAGGCACIAAIAAIICIAIICITAIIC		
18. NC 009019.1 Bat coronavirus HKU4-1 E-proteiAIGIIACCIIIIGICCAIGAACAAAIAGGIACIAICAIA		
Site # 1 🔹 🔍 with C w/o Gaps		

MEGAによるコロナウイルス系統樹作成 - bioinfo2014

6. Genetic Code について聞いてくるが、ここではそのまま「Yes」をクリックする。 Use the selected genetic code for rest of the Alignment S...



7. 「Alignment」→「Align by Muscle」を選ぶ。

M5: Alignment Explorer (E-protein.fasta)	X
Data Edit Search Alignment Web Sequencer Display Help	
📔 📽 🖬 🎬 🦉 Align by ClustalW 👔 🐰 💼 🗙 👘 💥 👘 🖉 🖛 🦊 🏘	ŝ <u></u>
DNA Sequences Tran W Align By Muscle	
Species/Abbry	
1. NC_008315.1_Bat	
2. NC_014470.1_BatUnmark All Sites FVSEEIGULIVNSVLLFLAFVVFLLVULAILIAL	¢
3. NC_003045.1_Bovi X Delete Gap-Only Sites	
IA. NC_005147.1_Humai Auto-Fill Gaps	
6. NC 012948.1 Bovine respiratory coronavirus AMA A FADIVEVENTIFIVATULIVITYVAFLATE	
7. NC_012949.1 Bovine_respiratory_coronavirus_bMABAYFADIV_YVGQIIFIVAICLLVIIVVVAFLAIF#I	
8. NC_006577.2_Human_coronavirus_HKU1_E-proteinMV_LFFNDIA_YIGQILVLVLFCLISLIFVVAFLAIIYI	(
9. NC_005831.2_Human_coronavirus_NL63_E-protein MFLELIDDNGIVLESILELLVMIFFFVLAMEFICLIDL	1
10. NC_012950.1_Human_enteric_coronavirus_straiMA_AYFADIV_YVGQIIFIVAICLLVIIVVVAFLAFFAI	(
11. NC_010800.1_Turkey_coronavirus_E-protein	
13. NC 006852.1 Murine hepatitis virus strain JMF LFL DIVEVEVEGITIFICAVELAVIIVVAFLASI	
14. NC 012936.1 Rat coronavirus Parker E-proteiMFLLFLIDIV YVGQIIFIVAVCLMVIIIVVAFLASIV	
15. NC_009020.1_Bat_coronavirus_HKU5-1_E-proteinL_FV0E0IGAFIVNFFILSVVCAVILVVCLAILIAI	(
16. NC_009988.1_Bat_coronavirus_HKU2_E-protein MFL IVEDOGLFI.IVL LLVLILVLLVAI VIKLIGL	1
17. NC_009021.1_Bat_coronavirus_HKU9-1_E-proteiMT IVGINNS ILIANVLVLIIICLLVVIVGCALLLIL	-
18. NC 009019.1 Bat coronavirus HKU4-1 E-proteiMLEFVHECIGEIIVEFFILEVVCAIELVVCLAILEAIEI	
Site # 1 🗘 🐨 with C. w/o Gaps sabled for translated protein data.	.:

8. Nothing selected for alignment. Select all? と聞いてくるので「OK」を選択する。



9. すると、以下のようにパラメーターを聞いてくるが、そのまま「Compute」ボタンを押してアライメントを開始する。

Option	Selection
Presets	None 💌
Gap Penalties	
Gap Open	-2.9
Gap Extend	0
Hydrophobicity Multiplier	1.2
Memory/Iterations	
Max Memory in MB	4095
Max Iterations	8
More Advanced Options	
Clustering Method (Iteration 1,2)	UPGMB
Clustering Method (Other Iterations)	UPGMB
Min Diag Length (lambda)	24
Additional MUSCLE Options	
Genetic Code (when using cDNA)	Standard
Alianment Info	MUSCLE Citation: Edgar. Robert C. (2004). MUSCL

10. 計算が終わったら、「Data」→「Export Alignment」→「MEGA Format」を選択して、MEGA形式でデータを出力する。

M5: Alignment Explorer (E-pro	tein.fasta)
Data Edit Search Alignment	Web Sequencer Display Help
Create New	🥹   💥 🖫   ∽ 🖻 % 🖻 🗙 🗡 🖮 🕼   ◀ 🕨  ] 🗛 🏘 🏘 🤻
🞬 Open	Sequences
Close	
	- BM48-31/BGR/2008 CHIVNVSLVKPIFYVYSRVKSLNSSQEVPEFL
Phylogenetic Analysis	us_E-protein CGMCNTLVLSPSIYVFNRGROFYEFYN-DVKPPVL
Save Session Ctrl+S	B_OC43_E-proteinCGMCNTLVLSPSIYVFNRGROFYEFYN-DVKPPVL
Export Alignment	MEGA Format CNILVLSPSIYVFNRGROFYEFYN-EVKPPVL
✓ DNA Sequences	FASTA format <sup>14</sup> CNILVLSPSIYVFNRGROFYEFYN-DVKPPVL
Protein Sequences	
떂 Translate/Untranslate	
😇 Select Genetic Code Table	pronavirus straid GMCNTLVLSPSIVVFNRGROFYEFYN-SVKPPVL
T   Reverse Complement	rus_E-protein CLFWYIWIVVPGAKGAAFVYNYIYGK-KLNKPE
Reverse	s_virus_strain_A <mark>CGLONTLVLSPSIYLYDRSKQLYKYYN-EE</mark> M
Complement	s_virus_strain_JCGLCNTLLLSPSIYLYNRSKQLYKYYNEEVRPPL
Evit AlpEvploror	- Parker_E-proteinCGLCNILLLSPSIYVYNRSKQLYKYYNSFVRPPPL
	_HKUS-1_E-proteifVSGCHTLVFLPAVHIYHDGRAAYVKFQ-ESHPPYP
16. NC_009988.1_Bat_coronaviru	s_HKU2_E-protein HRLMSNTIYIPVYNAYLVYKSYMVPCP
18 NC 009019 1 Bat coronaviru	
10. NC_003013.1_Bac_coronaviru	
•	4
Site # 72 🚔 @ with C w/o G	ans sabled for translated protein data
	aposabled for translated protein data.

11. 今回は、E-protein\_translated.meg という名前でデータを保存する。

🕅 名前を付けて保	存			×
保存する場所(I):	👢 corona	•	+ 🗈 🖻	▼ ▼
Ca	名前	*		更新日時
最近表示した場 所	E-protein.m	eg		2012/05/28 13:44
デスクトップ				
<b>(</b> ライブラリ				
<b>(人)</b> コンピューター				
	·	III		۲
ネットワーク	ファイル名(N):	E-protein_translated.meg		
	ファイルの種類(T):	MEGA Files		<ul> <li>キャンセル <sup>トッ</sup></li> </ul>

12. タイトルの入力を催されるが、無視して OK を押してよい。

Title	X
Input title of the data	
I	
ок Д	Cancel

13. アライメントが終了したので、「Data」→「Exit AlnExplorer」を選択して Alignment Explorer を終了する。

M5: Alignment Explorer (E-pro	tein.fasta)
Data Edit Search Alignment	Web Sequencer Display Help
Create New	♥  兼 [4] ∽ № ※ @ × ×   物 @   ◀ ▶  ] 桷 静 韩 韩
Reopen	Sequences
Close	
Phylogenetic Analysis	BM48-31/BGR/2008 CHIVNVSLVKPIFYVYSRVKSLNSSQEVPEFL
	us_E-protein CGMCNTLVLSPSIYVFNRGROFYEFYN-DVKPPVL
Save Session Ctrl+S	<pre>s_OC43_E-protein CGMCNILVLSPSIYVFNRGROFYEFYN-DVKPPVL</pre>
Export Alignment	tinating_encephaCGMCNTLVLSPSIYVFNRGRQFYEFYN-EVKPPVL
✓ DNA Sequences	ry_coronavirus_ACGMCNTLVLSPSIYVFNRGRQFYEFYN-DVKPPVL
Protein Sequences	ry_coronavirus_bCGMCNTLVLSSIYVFRGROFYFYFY-VKPVL
37 Select Genetic Code Table	
TIE	
A Reverse Complement	s virus strain ACGLCNTLVLSPSIVLYRSKOLYKYVN-FM
Reverse	s virus strain JCGLCNTLLLSSIYLYNRSKOLYKYYNE VRPPPL
Complement	Parker E-proteiCGLCNTLLLSPSIYVYNRSKOLYKYYNEEVRPPPL
Exit AlnExplorer	HKU5-1 E-proteiVSGCHTLVFLPAVHIYNIGRAAYVKFQ-ESHPPY
16. NC_009988.1_Bat_coronavirus	s_HKU2_E-protein HRLMSNTIYIPVYNAYLVYKSYM-EVEPCP
17. NC_009021.1_Bat_coronavirus	s_HKU9-1_E-proteinCGFVFKFVCKFTILVYNKFRNESLLNEREELL-
18. NC_009019.1_Bat_coronaviru	s_HKU4-1_E-proteitASGVIILLFV2AFYIYNIGRIAYFKFQ-INRPPF
· · · · · · · · · · · · · · · · · · ·	
Site # 72 🚔 @ with 🔿 w/o G	apssabled for translated protein data.

14. アライメントセッションを保存するかどうか聞いてくるが、今回は保存せずに終了したいので「いいえ」を選択する。 ■ Do you want to save the file?

?	Closing Alignment Explorer. Would you like to save the current alignment session to file?
	(はい(Y) いいえ(N) キャンセル

15. 先ほど出力したMEGA形式のデータを開く。

保存先のフォルダから E-protein\_translated.meg を選択してダブルクリックする。 🕽 🗢 🚺 « linuxhome 🕨 corona ★ ← coronaの検索 Q  $\mathbf{C}$ 整理 ▼ ライブラリに追加 ▼ 共有 ▼ 書き込む » .... ? . 名前 更新日時 種類 🍃 ライブラリ 🕙 E-protein.fasta 2012/05/28 13:28 FASTA File 📑 ドキュメント E-protein\_translated.meg 2012/05/28 15:59 MEG File Ξ 🔤 ピクチャ • 種類: MEG File 🔣 ビデオ サイズ: 2.68 KB 🎝 ミュージック + + ш 更新日時: 2012/05/28 15:59 Þ 2 個の項目

まずはBootstrap検定なしの系統樹を作成する。

16. MEGAが起動したら、「Analysis」→「Phylogeny」→「Construct/Test Neighbor-Joining Tree ...」を選択する。



17. Would you like to use the currently active data? (ファイル名) と聞いてくるので、ファイル名を確認して「Yes」ボタンを押す。 Use the active file?

?	Would you like to use the currently active data (E-protein_translated.meg)?
	Remember to reuse currently active data.
	Yes No

18. パラメーターを聞いてくるので、「Test of Phylogeny」が「None」になっていることを確認して「Compute」ボタンを押し、計算を開始する。

M M5: Analysis Preferences		
Options Summary		
Option	Selection	
Analysis	Phylogeny Reconstruction	
Scope	All Selected Taxa	
Statistical Method	Neighbor-joining	
Phylogeny Test		
Test of Phylogeny	None 🗸 🗸	
No. of Bootstrap Replications	None	
Substitution Model	Interior-branch test	
Substitutions Type	Amino acid	
Model/Method	Poisson model	
Rates and Patterns		
9393 among Sites	Uniform rates	
Gamma Parameter	Not Applicable	
Pattern among Lineages	Same (Homogeneous)	
Data Subset to Use		
Gaps/Missing Data Treatment	Complete deletion	
Site Coverage Cutoff (%)	Not Applicable	
Compute X Cancel ? Help		

19. しばらく待つと系統樹ができあがる。



次に、Bootstrap検定をしてみる。

20. 16からの操作を再び行い、系統樹作成のパラメーターを聞いてくる箇所で、「Test of Phylogeny」に「Bootstrap method」を選ぶ。 「No of Bootstrap Replications」が500になっていることを確認して「Compute」ボタンを押す。

M5: Analysis Preferences	
Options Summary	
Option	Selection
Analysis	Phylogeny Reconstruction
Scope	All Selected Taxa
Statistical Method	Neighbor-joining
Phylogeny Test	
Test of Phylogeny	Bootstrap method 🔹 💌
No. of Bootstrap Replications	None
Substitution Model	Interior-branch test
Substitutions Type	Amino aciu
Model/Method	Poisson model
Rates and Patterns	
Rates among Sites	Uniform rates
Gamma Parameter	Not Applicable
Pattern among Lineages	Same (Homogeneous)
Data Subset to Use	
Gaps/Missing Data Treatment	Complete deletion
Site Coverage Cutoff (%)	Not Applicable
Compute X Cancel ? Help	

21. しばらく待つと系統樹ができあがる。



22.「Caption」を押すと図の説明や参考文献が示される。論文に図を掲載する際の助けとなる。

## 応用課題:構築したコロナウイルスの系統樹にSARSウイルスのデータを追加して、その系統樹を描く

- SARSウイルスの配列 <sup>■</sup> <u>sars\_E-protein.fasta</u> をダウンロードし、先程ダウンロードしたデータにこのデータを加えて系統樹を作成 せよ。
  - SARS配列を追加するには、MEGA の Alignment Explorer から追加する方法もあるが、TeraPad などのテキストエディタで fasta 形式のファイルを開いて加工するのが簡単である。