

CLC Genomics Workbenchを用いた 分子系統樹解析

✓ CLC Genomics Workbenchには、遺伝子配列データより分子系統樹作成を行うツール、ある いはその関連ツールが多数搭載されている

系統樹解析ツール

✓ 系統樹の表示を行う際に、サンプルのメタデータも同時に表示させ、解析結果の解釈などに役立 てることが可能

Toolbox			
Processes	Toolbox	Favorites	
Kenter tool n	ame>		8
 Beno <l< td=""><th>me Finishir Mark Gene bial Genon sical Seque</th><td>ng Module Finding nics Module nce Analysis</td><th>3</th></l<>	me Finishir Mark Gene bial Genon sical Seque	ng Module Finding nics Module nce Analysis	3
	Greate A Join Alie Create F K-mer B Create T Model Te Maximur	Alignment gnments Pairwise Com ased Tree C Tree esting n Likelihood	oparison construction Phylogeny
⊕ G ⊕ G = G = G	eneral Seq ucleotide <i>A</i> rotein Anal NA Structu cular Biolog	uence Analy: Analysis ysis re gy Tools	sis

Create Alignment :

• 配列間アライメントデータの作成

Join Alignments :

• 複数のアライメントデータを統合

Create Pairwise Comparison :

ペアワイズの配列間比較データを作成

K-mer Based Tree Construction :

K-merを利用した系統樹の作成

Create Tree :

系統樹の作成

Model Testing :

最適な置換モデルの同定

Maximum Likelihood Phylogeny :

• 最大尤度を利用した系統樹の作成

Filgen

解析ワークフロー



- 配列データに始まり、「Create Tree」「Maximum Likelihood Phylogeny」では「Create Alignment」によるアライメントデータ作成を経て、系統樹を作成する
- 「K-mer Based Tree Construction」では、アライメントデータ作成を経ずに系統樹を作成する ため、アライメントデータ作成が困難な配列長の大きいデータセットに適している



配列アライメント



- ✓「Create Alignment」を使用し、複数の配列データを並べて表示したアライメントデータを作成
- ✓ 塩基の頻度を表すシークエンスロゴも同時に表示

DK-5151	ATGGAATGGA	ACACTTTCTT	CTTGGTGATC	TTGATCATCA	TCATAAAGAG	CACCACACCA	60
DK-6137	ATGGAATGGA	ACACTTTCTT	CTTGGTGATC	TTGATCATCA	TCATAAAGAG	CACCACACCA	60
DK-6045	ATGGAATGGA	ACACTTTCTT	CTTGGTGATC	TTGATCATCA	TCATAAAGAG	CACCACACCA	60
DK-7974	ATGGAATGGA	ACACTTTCTT	CTTGGTGATC	TTGATCATCA	TCATAAAGAG	CACCACACCA	60
DK-7380	ATGGAATGGA	ATACTTTCTT	CTTGGTGATC	TTGATCATCA	TCATAAAGAG	CACCACACCA	60
DK-9595168	ATGGAATGGA	ATACTTTCTT	CTTGGTGATC	TTGATCATCA	TCATAAAGAG	CACCACACCA	60
DK-9695377	ATGGAATGGA	ACACTTTTTT	CTTGGTGATC	TTGATCATCA	TCATAAAGAG	CACCACACCA	60
DK-9895024	ATGGAATGGA	ACACTTTTTT	CTTGGTGATC	TTGATCATCA	TCATAAAGAG	CACCACACCA	60
DK-5741	ATGGAATGGA	ACACTTTTTT	CTTGGTGATC	TTGATCATCA	TCATAAAGAG	CACCACACCA	60
DK-9895174	ATGGAATGGA	ACACTTTTTT	CTTGGTGATC	TTGATCATCA	TCATAAAGAG	CACCACACCA	60
DK-3946	ATGGAATGGA	ACACTTTTTT	CTTGGTGATC	TTGATCATCA	TCATAAAGAG	CACCACACCA	60
Fil3	ATGGAATGGA	ACACTTTTTT	CTTGGTCATC	TTGATCATCA	TCATAAAGAG	CACCACACCA	60
DK-3592B	ATGGAATGGA	ACACTTTTTT	CTTGGTGATC	TTGATCATCA	TCATAAAGAG	CACCACACCA	60
DK-3971	ATGGAATGGA	ACACTTTTTT	CTTGGTGATC	TTGATCATCA	TCATAAAGAG	CACCACACCA	60
AU-8-95	ATGGAATGGA	ACACTTTTTT	CTTGGTGATC	TTGATCATCA	TCATAAAGAG	CACCACACCA	60
FR-1458	ATGGAATGGA	ACACTTTTTT	CTTGGTGATC	TTGATCATCA	TCATAAAGAG	CACCACACCA	60
GE-1.2	ATGGAATGGA	ACACTTTTTT	CTTGGTGATC	CTGATCATCA	TCATAAAGAG	CACCACACCA	60
DK-4p168	ATGGAATGGA	ATACTTTTTT	CTTGGTGATC	TTGATCATTC	TCATAAAGAG	CACCACATCA	60
UK-MLA98-6PT11	ATGGAATGGA	ATACTTTTTT	CTTGGTGATC	TTGATCATTC	TCATAAAGAG	CACCACATCA	60
UK-H17-5-93	ATGGAATGGA	ATACTTTTTT	CTTGGTGATC	TTGATCATTC	TCATAAAGAG	CACCACATCA	60
UK-H17-2-95	ATGGAATGGA	ATACTTTTTT	CTTGGTGATC	CTGATCATTC	TCATAAAGAG	CACCACATCA	60
IR-F13.02.97	ATGGAATGGA	ATACTTTTTT	CTTGGTGATC	TTGATCATTC	TCGTAAAGAG	CACCACATCA	60
DK-4p101	ATGGAATGGA	ATACTTTTTT	CTTGGTGATC	TTGATCATTC	TCATAAAGAG	CACCACATCA	60
FR-L59X	ATGGAATGGA	ATACTTTTTT	CTTGGTGATC	TTGATCATTA	TCATAAAGAG	CACCACATCA	60
UK-860-94	ATGGAATGGA	ATACTTTTTT	CTTGGTGATC	TTGATCATTA	TCATAAAGAG	CACCACATCA	60
DK-1p53	ATGGAATGGA	ATACTTTTTT	CTTAGTGATT	TTGATCATCA	TCATAAAAAG	CACCACGTCA	60
DK-1p55	ATGGAATGGA	ATACTTTTTT	CTTAGTGATT	TTGATCATCA	TCATAAAAAG	CACCACGTCA	60
US-Makah	ATGGAATGTA	ATACTTTTTT	CTTGGTGATT	CTGATCATCA	TCATAAAAAG	CACCACATCA	60
US-Goby1-5	ATGGAATGGA	ATACTTTTTT	CTTGGTGATT	CTGATCATCA	TCATAAAAAG	CACCACATCA	60
Consensus	ATGGAATGGA	ATACTTTTT	CTTGGTGATC	TTGATCATCA	TCATAAAGAG	CACCACACCA	
100%							
Conservation							
2.0bits Secure contactor				TATATAT		$\Delta T = \Delta T = \Delta T = \Delta T$	
Sequence 10g0 0.00lts	ATUGAATUGA	AFAVIIITII	VIIGUIGAIC	TIGATCATCA	IVATAAAGAG	UAUUAUACUA	

✓ 計算アルゴリズムの異なる「Create Tree」「K-mer Based Tree Construction」
 「Maximum Likelihood Phylogeny」の3種類のツールを使用可能

系統樹作成

✓ 作成された系統樹データは、レイアウトの編集や属性データなどの追加表示のカスタマイズが可能



Filgen

biosciences & nanoscience

計算パラメータ



Gx Create Tree	×
1. Choose where to run	Tree Construction
2. Select alignments of same type	Tree construction method Neighbor Joini Nucleotide distance measure Jukes-Cant
8. Tree Construction	Protein distance measure Jukes-Cantor 👻
4. Result handling	Bootstrapping Perform bootstrap analysis Replicates 100
Help Reset	Previous Next Einish Qancel

K-mer Based Tree Construction								
 Choose where to run Select input sequences K-mer Based Tree Construction Result handling 	K-mer Based Tree Construction Tree construction Construction method Neighbor Joini K-mer settings K-mer length (the value k) 15 Distance measure Mahalanobis							
Help Rese	t <u>Previous Next</u> Einish <u>Qancel</u>							

Create Tree

K-mer Based Tree Construction

Gx Maximum Likelihood Phyle	ogeny 📃 🗙
1. Choose where to run	Maximum Likelihood Phylogeny
2. Select alignment(s)	Construction method Neighbor Joini Evisting start tree
3. Maximum Likelihood Phylogeny	
4. Bootstrapping Parameters	Nucleotide substitution model
5. Result handling	Protein substitution model WAG - Transition / transversion ratio 2.0
Class Clas Cla	Rate variation Include rate variation Number of substitution rate categories Gamma distribution parameter 1.0
Contraction of the second second	Estimation
0117810	Estimate substitution rate parameter(s) Estimate topology Estimate gamma distribution parameter
Help Reset	Previous <u>N</u> ext <u>F</u> inish <u>Qancel</u>

- ✓ 3ツールとも、系統樹作成にNeighbor Joining法かUPGMA法を選択可能
- ✓ その他、ツールによってはブートストラップ値の 計算や、塩基置換モデルの選択、K-mer値 の設定などを行うことが可能

Maximum Likelihood Phylogeny





JK-9643

DK-1p55

FR-L59X UK-H17-2-9 DK-4p168 FR-2375 DK-9995144 DK-3971 DK-3946 DK-200079-DK-6137 DK-9895174 DK-5131 VO-A16368G DK-1p40 DK-6p403 SE-SVA14 KRRV9601







✓ 系統樹作成の際にブートストラップ値の計算を行った場合、ブートストラップ値の表示や、閾値を 設定してのノードの折り畳み、あるいはハイライト表示の設定を行うことが可能



メタデータインポート



	A B	C	D	E	F	G						
1	Sequence Strain	Host	Water	Country	ACCNo	Year \						
2	ATGGAATGGAACACTTTTTTAU-8-95	Rainbow trout	Fresh water	AU	AY546570.1	1995	\					
3	ATGGAATGGAACACTTTTTTCH-FI262BF	H Rainbow trout	Fresh water	CH	AY546571.1	1999	1					
4		Atlantic Horring	Sea water	DK	AT 540575.1	1006	1					
5		Sprat	Sea water	DK	AT540577.1	1990	1					
7	ATGGAATGGAATACTTTTTDK-1p35	Atlantic Herring	Sea water	DK	AY546573 1	1996	1					
8		Sprat	Sea water	DK I	A1010070.1	1550	1					
9	ATGGAATGGAATACTTTCT1DK-200027	3 Rainbow trout	Eresh water	DK	Rows: 121	ph	nylogenet	ic tree: Ph	nyl			Filter 🔻
10	ATGGAATGGAATACTTTCTTDK-200079-	1 Rainbow trout	Fresh water	DK		_				_		
11	ATGGAATGGAACACTTTTTDK-200098	Rainbow trout	Fresh water	DK	Name	Node t	tvpe	Branch	Bootstr	Size	Start of sequence	
12	ATGGAATGGAATACGCTCC1DK-2835	Rainbow trout	Fresh water	DK			VF-					
13	ATGGAATGGAACACTTTTTTDK-3592B	Rainbow trout	Fresh water	DK		Root		0.00		U		<u> </u>
14	ATGGAATGGAACACTTTTTDK-3946	Rainbow trout	Fresh water	DK	UK-9643	Leaf		1.§5E-3		1524	ATGGAATGGAATACTTTTTTCTTGGTG	ATCTTGAT
15	ATGGAATGGAACACTTTTTDK-3971	Rainbow trout	Fresh water	DK		Interna	l node	1.6 % E-3		0		
16	ATGGAATGGAATACTTTTTDK-4p101	Whiting	Sea water	DK		Internal	Inode	3.295-4	4 1	0		
17	ATGGAATGGAATACTTTTTDK-4p168	Atlantic Herring	Sea water	DK		Internal	Inode	143E\5	52	0		E
18	ATGGAATGGAATACTTTTT1DK-4p37	Blue whiting	Sea water	DK		Internal	Loode	6 705-1	10			
19	ATGGAATGGAATACGTTTT1DK-5123	Rainbow trout	Fresh water	DK		Interna	rnoue	0.72E-4	100	0		
20	ATGGAATGGAATACGCCCTTDK-5131	Rainbow trout	Fresh water	DK		Interna	Inode	5.28E-3	100	0		
1	•					Interna	Inode	4.41E-3	100	0		
						Interna	l node	4.05E-4	18	0		
						Interna	l node	4.81E-3	92	. 0		
						Interna	l node	0.03	100	0		
						Interna	Inode	4 85E-3	65	. 0		
						Internal	Loode	0.07	100	0		
						Internal	Thoue	0.07	100	1504	ATOO AATOO AATAO TTOO TOO TO	ATTOTOAT
					US-GODy1-5	Leat		0.01	1	1524	AIGGAAIGGAAIACTITITUUIGGIG	ATTOTGAT
	•				US-Makah	Leaf		0.02		1524	ATGGAATGTAATACTTTTTTCTTGGTG	ATTOTGAT
						Interna	l node	0.07	100	0		
		1 N			DK-1p55	Leaf		-5.10E-5		1524	ATGGAATGGAATACTTTTTTCTTAGTG	ATTTTGAT
					DK-1p53	Leaf		7.07E-4		1524	ATGGAATGGAATACTTTTTCTTAGTG	ATTTTGAT
			`			Interna	Lnode	0.04	100			
					NP-080-04	Loof	mode	0.01	100	1514	ATCOMATCOMATACTITUTCTCCTC	ATCTTOAT
					UN-000-94	Lear		0.02	0.5	024		AIOTIGAT
				1 N.		Interna	Inode	0.73E-3	97	0		
				1	FR-L59X	Leaf		0.01		1584	ATGGAATGGAATACTTTTTTCTTGGTG	ATCTTGAT
						Interna	l node	2.43E-3	82	! 🐧		
\checkmark	サンプルの屋性悟費	品を主とめた	Fycel		DK-4p101	Leaf		8.66E-3		1524		ATCTTGAT
•	ノノノルの両1上1月+					Internal	l node	2.55E-3	89	0	N Contraction of the second se	
	ファイル かどをマタデ	<u>_ カレー てょ</u>	へ ポー		IR-E13.02.97	Leaf		5.87E-3		1524	ATGGAATGGAATACTTTTTCTTGGTG	ATCTTGAT
	ノバールなしでグノノ	JCUCI	ノハ		1101012.01	Internal	Londo	7945-4	79	1024	1	
	トレーズ 公本時に明ら	もんけけもんに	ニマレガ			Interna	I nose	7.046-4				
	「し、木利」団に関い	モリリグイリ	ノービル	.		Internal	i node	4.40E-4	52			
					UK-H17-2-95	Leaf		2.03E-3		1524	ATGGAATGGAATACTTTTTCTTGGTG	ATCCTGA
	り形					Interna	l node	1.26E-3	92	. 0		*
										Import Mat	adata	
										unport Met		
					🔁 🖽 🔁							

9

メタデータインポート



✓ メタデータのインポートを行うと、系統樹の各サンプルに対して、Excelファイルの属性情報が関連 付けられ、マウスカーソルを合わせることで、情報が表示される



Country: JP

ACCNo: AB672614.1

Year: 1996

メタデータ表示設定



▶ Tree Settings	
Minimap	-
Tree layout	
Node settings	
Label settings	
Background settings	
Branch layout	
Bootstrap settings	
Metadata	
 Hide Unknown Legend font settings Node shape Node symbol size Node color Label text Label text color Label background color Branch color 	
Use default 🔹 👻	
Use default	
Water ACCNo Year H Country Host	

- ✓ メタデータのインポートを行うと、画面右側の「Metadata」項目 にて、メタデータの属性情報に基づいた、系統樹のノードやラベ ル表示の設定を変更できるようになる
- ✓ 複数の設定項目を、系統樹に同時に反映させることができ、 系統樹とサンプル属性情報を一つの画面に統合させての表示 が可能となる

メタデータ表示



✓ Node symbol sizeとNode colorにメタデータの情報をそれぞれ反映



メタデータ表示



✓ 続いてレイヤーにもメタデータの情報を反映



メタデータ表示



✓ 系統樹の表示レイアウトを変更した場合も、同様にメタデータの表示が可能





お問い合わせ先:フィルジェン株式会社 TEL 052-624-4388 (9:00~18:00) FAX 052-624-4389 E-mail: biosupport@filgen.jp