

マルチプルアライメントの目的

```

1nshA      SRPTEATERCIESLIAVFQKYAGKDGHSVTLTKTEFLSFMNTELAFTKNQKDPGVLD RMM
1j55A      --MTELETAMGMIIDVFSRYSYSGEGSTQTLTKGELKVLMEKELPGFLD-----AVDKLL
1ig5A      -----KSPEELKGI FEKYAAKEGDPNQLSKEELKLLQLQTEFFPSLLKG---PSTLDEL F
1qx2A      -----MKSPEEIKGAFEVFAAKEGDPNQISKEELKLVMTLGPSSLKG---MSTLDEMI
           .   :   * . : : : * . : : * * : : : . . . . . : * : :

```

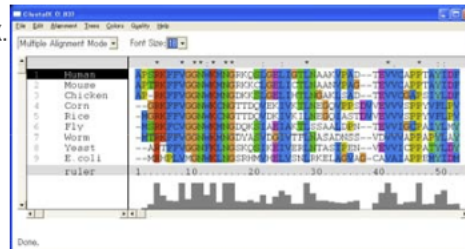
- ファミリ内の機能的重要な部位の検出
- ファミリを特徴付けるモチーフの発見
- プロフィール法による遠縁のホモログ発見
- 分子系統解析の第一ステップとして不可欠
- 進化的追跡法(evolutionary trace method)

ClustalW / ClustalX

UNIX/Windows/Mac版 : <ftp://ftp.ebi.ac.uk/pub/software/clustalw2>

WEBサーバ : <http://www.ebi.ac.uk/Tools/clustalw2>

- 現在、最も一般的な多重整列のプログラム
- アルゴリズムは累進法。ペアワイズアライメントはグローバルアライメントを用い、ガイド木はNJ法で作成。スコアは配列の重みを導入したSum-of-pairs。置換スコア行列の選択、ギャップペナルティ等に様々な経験的な工夫が見られる。
- CUI版はClustalW, GUI版はClustalX。UNIX, Windows, MACでも動作する。
- NJ法による系統樹計算機能付き。



Thompson, J.D., Higgins, D.G., Gibson T.J. "CLUSTALW : improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice". Nucleic Acids Research, 1994, 22, 4673-4680.

主要なマルチプルアライメントのプログラム

	WEBサイト	アルゴリズム	特徴
ClustalW・ClustalX	http://www.ebi.ac.uk/Tools/clustalw2	累進法。重み付きSPスコアを使用。置換スコア行列の選択、ギャップペナルティ等に様々な工夫	もっとも広く使われている標準的なプログラム
T-COFFEE	http://www.ebi.ac.uk/t-coffee/	ペアワイスアライメントをローカル、グローバル、進展を用いて多数生成。それらの集合から、位置特異的スコアを作成し、累進法を実行する。	計算時間がかかるが精度は高い。 配列の本数が100本以下 の場合に向いている。
MAFFT	http://align.bmr.kyushu-u.ac.jp/mafft/online/server/	高速フーリエ変換(FFT)を用いて、高速にペアワイスアライメントを実装、それを利用して、累進法、あるいは反復改善法を実行する。	計算時間は高速なので、 配列の本数が100~500本程度 でも、計算可能。

サイトごとに保存の度合いに差がある

よく保存しているサイト → そのファミリーにとって重要なサイト
→ 機能上重要なサイトである可能性が高い

```

5p21-  MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSY
1ctqA  MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSY
1clyA  MREYKLVVLGSGGVGKSALTVQFVQGI FVEKYDPTIEDSY
1kao-  MREYKVVVLGSGGVGKSALTVQFVTGTF IEKYDPTIEDFY
1huqA  --QFKLVLLGESAVGKSSLVLRVFKGQFHEYQESTIGAAF
1g16A  ----KILLIGDSGVGKSCLLVRFVE----DKFNPI--DFK
1ek0A  VTSIKLVLLGEAAVGKSSIVLRFVSNDFAEENKEPTIGAAF
3rabA  ---FKILII GNSSVGKTSFLFRYADDSETPAFVSTVGIDF
1mh1-  ----KCVVVDGAVGKTCLLI SYTTNAFPGEYIPTVFDNY
2ngrA  MQTIKCVVVDGAVGKTCLLI SYTTNKFPSEYVPTVFDNY
1tx4B  ----KLVIVGDGACGKTCLLIVNSKQF---YVPTVFENY

```

サイトごとに保存の度合いに差がある。

サイトごとにアミノ酸の出現傾向に差がある

[AG]-x(4)-G-K-[ST]

```

1         2         3         4         5         6         7         8         9
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
PLAS_ORYSI : VFEPNDFVKSGETITFFKNNAGFPHNVFDEDA-VPSGVD--VSKI SQ--EYLNAPGETFSVTLT---VPGTYGFYCEPHAGAGMVGKV
PLAS_HORVU : VFEPNDFSVKAGETITFFKNNAGFPHNVFDEDA-VPSGVD--VSKI SQ--EYLNAPGETFSVTLT---VPGTYGFYCEPHAGAGMVGKV
PLAS_ENTFR : AFVFNITVVGAGESIEFVNNAGFPHNVFDEDA-VPAGVD--ADAISA--EDYLNKSGQTVVRKLT---TPGTYGVYCDPHSAGAGMVTI
PLAS_ULVPE : AFVPSKISVAAGEAIEFVNNAGFPHNVFDEDA-VPAGVD--ADAISY--DDYLNKSGQTVVRKLT---TPGTYGVYCEPHAGAGMVTI
PLAS_CHLRE : EFVFKTLTIKSGETVNFVNNAGFPHNVFDEDA-IPSGVN--ADAISR--DDYLNAPGETYSVKLT---AAGEYGYCEPHQAGMVGKI
PLAS_DAUCA : VFSPSSFSVAKGEGISFKNNAGFPHNVFDEDE-VPAGVD--VSKI SQ--EDYLDGAGESFTVTLT---EKGTYKFCYCEPHAGAGMKGVE
PLAS_CAPBU : AFVPNDFSIAGKEKIVFKNNAGFPHNVFDEDE-IPSGVD--ASKI SMDENDLLNAGETYEVALT---EAGTYSFYCAPHQAGMVGKV
PLAS_1_ARATH: AFVPEFTVAKGEKIVFKNNAGFPHNVFDEDE-IPSGVD--ASKI SMDETALLNAGETYEVTLT---EPGSYGFYCAPHQAGMVGKL
PLAS_MERPE : AFVFNNSVSPSGEKITFKNNAGFPHNVFDEDE-IPSGVD--ASKI SMDDEADLLNAPGETYAVTLT---EKGSYFYCSFHQAGMVGKV
PLAS_PHAVU : VFVPEFSVSPSGEKIVFKNNAGFPHNVFDEDE-IPAGVD--AVKI SMPEEELLNAPGETYVVTLD---TKGTYSFYCSFHQAGMVGKV
PLAS_1_POPNI: AFVPEFSISPGEKIVFKNNAGFPHNVFDEDS-IPSGVD--ASKI SMSEEDLLNAPGETYFEVALS---NKGEYSFYCSFHQAGMVGKV
PLAS_1_SILPR: AFVPSDLISIASGEKIVFKNNAGFPHNVFDEDE-VPAGVD--VTKI SMPEEDLLNAPGEEYSVTLT---EKGTYKFCYCEPHAGAGMVGKV
PLAS_SOLCR : AFVPGNFSISAGEKIVFKNNAGFPHNVFDEDE-IPAGVD--ASKI SMPEEDLLNAPGETYSVTLT---EKGTYSFYCSFHQAGMVGKV
PLAS_SAMNI : AFIPSNFSVSPSGEKITFKNNAGFPHNVFDEDE-VPSGVD--SAKI SMSEEDLLNAPGETYSVTLT---ESGTYKFCYCSFHQAGMVGKV
PLAS_VICFA : AFVNSFEVSAGDTIVFKNNAGFPHNVFDEDE-IPSGVD--AAKI SMPEEDLLNAPGETYSVKLD---AKGTYKFCYCSFHQAGMVGQV
PLAS_PEA : AFVPSLEVSAGETIVFKNNAGFPHNVFDEDE-IPAGVD--ASKI SMPEEDLLNAPGETYSVKLD---AKGTYKFCYCSFHQAGMVGQV
PLAS_FRIAG : AFVPSNIEVAAGETIVFKNNAGFPHNVFDEDE-VPKGVD--AGAI SMKEEDLLNAPGETFSVTLK---EKGTYSIYCSFHQAGMAGKI
PLAS_GINBI : AFIPNELQVNAEQIVFKNNAGFPHNVFDEDA-VPAGVD--VSSI SMSEEDLLNAPGETYVVKLD---KGTYRFFCAPHQGIGMSGIV
PLAS_PHYPA : GFYPKDISVAAGESVTFVNNKGFPHNVFDEDA-VPAGVK--TEDINH--EDYLNPNESFSTIFK---TPGTYEFYCEPHQAGMKGV
PLAS_DRYCA : KFYPDSTITVSAAGEVEFTLVGETGHHIVFDIPAGAPGTVASELKAA SMDENDLLSEDEPSFKAKVS---TPGTYTFYCTPHKSNMKGTL
PLAS_PHOLA : QFEPANVTVHPGDTVVKWNNKLPNNILFDDKQ--VPGASKELADKLSHS--QLMFS PGESYEITFSDFPAGTYTYCAPHRGAGMVGKI
PLAS_SYNP6 : AFEPTIEIQAGDTVQWNNKLPNNIVVEGQ--P-----ELSHK--DLAFSPGETFEATFS---EPGTYYTYCEPHRAGMVGKI
PLAS_AMASO : VFEPAKLTIKPGDTVEFLNNKVPNNVFDAAAL-NPAKSADLAKLSLHK--QLMSPGQSTSTTFPADAPAGETTYCEPHRAGMVGKI
PLAS_SYNY3 : VFEPTVTIKAGEEVKWNKLSPHNVFADG-VDADT--AAKLSHK--GLAFAAGESFTSTFT---EPGTYYTYCEPHRAGMVGKI
PLAS_PROHO : LYEPAKALISAGDTVEFVNMKVPNNVFDKVP-AGES---APALSNT--KLA IAPGSFYSVTLG---TPGTYSFYCTPHRAGMVGTI
AZUP_ACHCY : VFEPAKLVAPGDTVTFIPTRDKG--HNVEIKGM-IPDGAEE-----AFKSKINENYKVTFT---APGVYGVKCTPHYGMGMVGIV
AZUP_PARDE : VFEPAFIRAEVPGDVIPTDKG--HNVEAIKEI-LPEGVE-----TFKSKINEAYALTVT---EPGLYGVKCTPHYGMGMVGLV
AZUP_ALCFA : VFEPAYIKANPGDVIPTDKG--HNVESIKDM-IPEGAE-----KFKSKINENYVLTVT---QPGAYLVKCTPHYAMGMIALI
AZUP_RHILV : VFEPAKLVAPGDTVTFIPTRDKG--HNVEIKGM-IPDGV-----DFKSKPNEQYVQVFD---IPGAYLVKCTPHYGMGMVALI
AZUP_METEX : VFEPAKLVAPGDTVTFIPTRDKG--HNVEIKGM-APDGAD-----YKVTYVQGEAVVVKFD---KEGVYGFKCAPHYMMGMVALI
AMCY_PARDE : KYTEPELVHVKVGDIVTWINREAMPNNVHFVAGVLGEAALK-----GPMKKEQAYSILFTT---EAGTYDYHCTPHP---FMRGKV
AMCY_PARVE : KYLTPVETIKAGETVYVWNGEVMPHNVAFKKGIVGEDAFR-----GEMMKTQAYAITFN---EAGSYDYFCTPHP---FMRGKV
AMCY_METEX : KFQTPVRIKAGSAVWTNTEALPHNVHFKSGPGVEKDVE-----GMLRSNQTYSVKFN---APGTYDYICTPHP---FMMGKV

```

- (1) 完全に保存しているサイト番号は: 12G
- (2) そのうち銅イオンの結合に関与するサイト番号は: _____

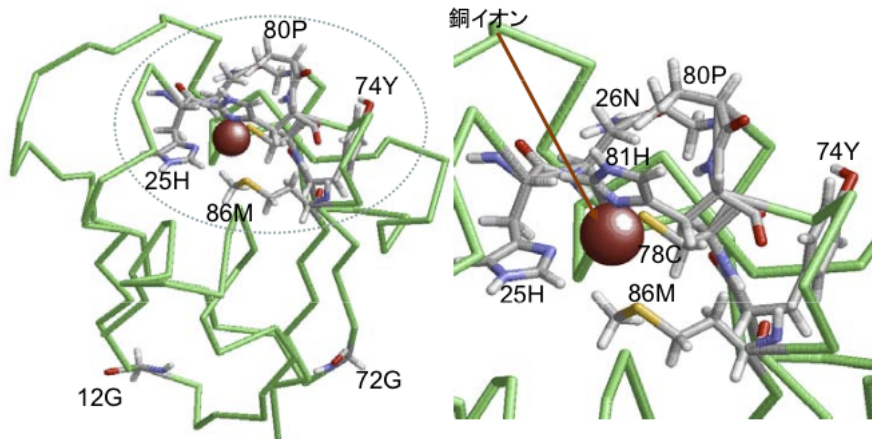
```

1         2         3         4         5         6         7         8         9
12345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
PLAS_ORYSI : VFEPNDFVKSGETITFFKNNAGFPHNVFDEDA-VPSGVD--VSKI SQ--EYLNAPGETFSVTLT---VPGTYGFYCEPHAGAGMVGKV
PLAS_HORVU : VFEPNDFSVKAGETITFFKNNAGFPHNVFDEDA-VPSGVD--VSKI SQ--EYLNAPGETFSVTLT---VPGTYGFYCEPHAGAGMVGKV
PLAS_ENTFR : AFVFNITVVGAGESIEFVNNAGFPHNVFDEDA-VPAGVD--ADAISA--EDYLNKSGQTVVRKLT---TPGTYGVYCDPHSAGAGMVTI
PLAS_ULVPE : AFVPSKISVAAGEAIEFVNNAGFPHNVFDEDA-VPAGVD--ADAISY--DDYLNKSGQTVVRKLT---TPGTYGVYCEPHAGAGMVTI
PLAS_CHLRE : EFVFKTLTIKSGETVNFVNNAGFPHNVFDEDA-IPSGVN--ADAISR--DDYLNAPGETYSVKLT---AAGEYGYCEPHQAGMVGKI
PLAS_DAUCA : VFSPSSFSVAKGEGISFKNNAGFPHNVFDEDE-VPAGVD--VSKI SQ--EDYLDGAGESFTVTLT---EKGTYKFCYCEPHAGAGMKGVE
PLAS_CAPBU : AFVPNDFSIAGKEKIVFKNNAGFPHNVFDEDE-IPSGVD--ASKI SMDENDLLNAGETYEVALT---EAGTYSFYCAPHQAGMVGKV
PLAS_1_ARATH: AFVPEFTVAKGEKIVFKNNAGFPHNVFDEDE-IPSGVD--ASKI SMDETALLNAGETYEVTLT---EPGSYGFYCAPHQAGMVGKL
PLAS_MERPE : AFVFNNSVSPSGEKITFKNNAGFPHNVFDEDE-IPSGVD--ASKI SMDDEADLLNAPGETYAVTLT---EKGSYFYCSFHQAGMVGKV
PLAS_PHAVU : VFVPEFSVSPSGEKIVFKNNAGFPHNVFDEDE-IPAGVD--AVKI SMPEEELLNAPGETYVVTLD---TKGTYSFYCSFHQAGMVGKV
PLAS_1_POPNI: AFVPEFSISPGEKIVFKNNAGFPHNVFDEDS-IPSGVD--ASKI SMSEEDLLNAPGETYFEVALS---NKGEYSFYCSFHQAGMVGKV
PLAS_1_SILPR: AFVPSDLISIASGEKIVFKNNAGFPHNVFDEDE-VPAGVD--VTKI SMPEEDLLNAPGEEYSVTLT---EKGTYKFCYCEPHAGAGMVGKV
PLAS_SOLCR : AFVPGNFSISAGEKIVFKNNAGFPHNVFDEDE-IPAGVD--ASKI SMPEEDLLNAPGETYSVTLT---EKGTYSFYCSFHQAGMVGKV
PLAS_SAMNI : AFIPSNFSVSPSGEKITFKNNAGFPHNVFDEDE-VPSGVD--SAKI SMSEEDLLNAPGETYSVTLT---ESGTYKFCYCSFHQAGMVGKV
PLAS_VICFA : AFVNSFEVSAGDTIVFKNNAGFPHNVFDEDE-IPSGVD--AAKI SMPEEDLLNAPGETYSVKLD---AKGTYKFCYCSFHQAGMVGQV
PLAS_PEA : AFVPSLEVSAGETIVFKNNAGFPHNVFDEDE-IPAGVD--ASKI SMPEEDLLNAPGETYSVKLD---AKGTYKFCYCSFHQAGMVGQV
PLAS_FRIAG : AFVPSNIEVAAGETIVFKNNAGFPHNVFDEDE-VPKGVD--AGAI SMKEEDLLNAPGETFSVTLK---EKGTYSIYCSFHQAGMAGKI
PLAS_GINBI : AFIPNELQVNAEQIVFKNNAGFPHNVFDEDA-VPAGVD--VSSI SMSEEDLLNAPGETYVVKLD---KGTYRFFCAPHQGIGMSGIV
PLAS_PHYPA : GFYPKDISVAAGESVTFVNNKGFPHNVFDEDA-VPAGVK--TEDINH--EDYLNPNESFSTIFK---TPGTYEFYCEPHQAGMKGV
PLAS_DRYCA : KFYPDSTITVSAAGEVEFTLVGETGHHIVFDIPAGAPGTVASELKAA SMDENDLLSEDEPSFKAKVS---TPGTYTFYCTPHKSNMKGTL
PLAS_PHOLA : QFEPANVTVHPGDTVVKWNNKLPNNILFDDKQ--VPGASKELADKLSHS--QLMFS PGESYEITFSDFPAGTYTYCAPHRGAGMVGKI
PLAS_SYNP6 : AFEPTIEIQAGDTVQWNNKLPNNIVVEGQ--P-----ELSHK--DLAFSPGETFEATFS---EPGTYYTYCEPHRAGMVGKI
PLAS_AMASO : VFEPAKLTIKPGDTVEFLNNKVPNNVFDAAAL-NPAKSADLAKLSLHK--QLMSPGQSTSTTFPADAPAGETTYCEPHRAGMVGKI
PLAS_SYNY3 : VFEPTVTIKAGEEVKWNKLSPHNVFADG-VDADT--AAKLSHK--GLAFAAGESFTSTFT---EPGTYYTYCEPHRAGMVGKI
PLAS_PROHO : LYEPAKALISAGDTVEFVNMKVPNNVFDKVP-AGES---APALSNT--KLA IAPGSFYSVTLG---TPGTYSFYCTPHRAGMVGTI
AZUP_ACHCY : VFEPAKLVAPGDTVTFIPTRDKG--HNVEIKGM-IPDGAEE-----AFKSKINENYKVTFT---APGVYGVKCTPHYGMGMVGIV
AZUP_PARDE : VFEPAFIRAEVPGDVIPTDKG--HNVEAIKEI-LPEGVE-----TFKSKINEAYALTVT---EPGLYGVKCTPHYGMGMVGLV
AZUP_ALCFA : VFEPAYIKANPGDVIPTDKG--HNVESIKDM-IPEGAE-----KFKSKINENYVLTVT---QPGAYLVKCTPHYAMGMIALI
AZUP_RHILV : VFEPAKLVAPGDTVTFIPTRDKG--HNVEIKGM-IPDGV-----DFKSKPNEQYVQVFD---IPGAYLVKCTPHYGMGMVALI
AZUP_METEX : VFEPAKLVAPGDTVTFIPTRDKG--HNVEIKGM-APDGAD-----YKVTYVQGEAVVVKFD---KEGVYGFKCAPHYMMGMVALI
AMCY_PARDE : KYTEPELVHVKVGDIVTWINREAMPNNVHFVAGVLGEAALK-----GPMKKEQAYSILFTT---EAGTYDYHCTPHP---FMRGKV
AMCY_PARVE : KYLTPVETIKAGETVYVWNGEVMPHNVAFKKGIVGEDAFR-----GEMMKTQAYAITFN---EAGSYDYFCTPHP---FMRGKV
AMCY_METEX : KFQTPVRIKAGSAVWTNTEALPHNVHFKSGPGVEKDVE-----GMLRSNQTYSVKFN---APGTYDYICTPHP---FMMGKV

```

- (1) 完全に保存しているサイト番号は: 12G, 25H, 26N, 72G, 74Y, 78C, 80P, 81H, 86M
- (2) そのうち銅イオンの結合に関与するサイト番号は: _____

実際の金属イオン結合サイト



PLAS_ORYSIを1plsAを鋳型にモデリングした構造

より細やかにサイトの保存性を抽出するには？

```

PLAS_ORYSI  PNDFTVKSGETITFKNNAGFPHNVVFEDEDA
PLAS_MERPE  PNNFSVPSGEEKITFKNNAGFPHNVVFEDEDE
PLAS_DAUCA  PSSFSVAKGEGISFKNNAGFPHNIVFEDEDE
PLAS_SAMNI  PSNFSVPSGEEKITFKNNAGFPHNVVFEDEDE
PLAS_VICFA  PNSFEVSAGDTIVFKNNAGFPHNVVFEDEDE
PLAS_CUCPE  PNDFSVAAGEKIVFKNNAGFPHNVVFEDEDE
PLAS1_ARATH PSEFTVAKGEKIVFKNNAGFPHNVVFEDEDE
PLAS_PEA    PSSLEVSAGETIVFKNNAGFPHNVVFEDEDE
PLAS_FRIAG  PSNIEVAAGETVVFKNAGFPHNVLFDEDE
PLAS_PHYPA  PKDISVAAGESVTFVNNKGFPHNVVFEDEDA
PLAS_ULVPE  PSKISVAAGEAIEFVNNAGFPHNIVFEDEDA
PLAS_ANASO  PAKLTIKPGDTVEFLNNKVPVPHNVVDAAL
PLAS_SYN6   PSTIEIQAGDTVQWVNNKLAPHNVVVEGQP
PLAS_DRYCA  PDSITVSAGEAVEFTLVGETGHNIVFDIPA
AZUP_RHILV  PGFLKIAPGDTVTFIPTDK-SHNVETFKGL
AMCY_METEX  TPEVRIKAGSAVTWTNTEALPHNVHFK---
AMCY_PARDE  TPELVKVGDTVTWINREAMPVHVF----
AMCY_PARVE  --EVTIKAGETVYVWNGEVMPHNVAFKKGII
              PNDFSVKSGETVWVKNAGFPHNVVFEDEDE
              TSNLEIAA EKITFVLVKAPG  IHTAAA
              -KSIT PK DA S IPTEVAS  E FGQL
              AKVK  P  G  E  T  GLT  A KIPP
              DT   V   Q   D   -   -KGI
    
```

・配列の本数が多い場合、完全保存サイトは置きにくくなる。

・完全保存サイトではなくてもより相対的に保存が良いサイトはある。

⇒より定量的に保存性を評価する必要がある

⇒サイトごとのアミノ酸頻度を計算する必要性

サイトごとのアミノ酸頻度 $f_i(a)$ の計算

12345678
 seq1 **ASHSGADK**
 seq2 **ASHSADDK**
 seq3 **LLHDLEDK**
 seq4 **LLHSSEEK**
 seq5 **ASHSEGKH**

$C_i(a)$: i 番目のサイトのアミノ酸 a の数
 A : アミノ酸の集合 {A, C, D, E, F, ... Y}

$$p_i(a) = \frac{C_i(a)}{\sum_{x \in A} C_i(x)}$$

i	配列	度数 $C_i(a)$								頻度 $f_i(a)$							
		A	D	E	G	H	K	L	S	A	D	E	G	H	K	L	S
1	AALLA	3						2		0.6						0.4	
2	SSLLS							2	3							0.4	0.6
3	HHHHH					5								1.0			
4	SSDSS	1						4		0.2							0.8
5	GALSE	1		1	1			1	1	0.2		0.2				0.2	0.2
6	ADEEG	1	1	2	1					0.2	0.2	0.4	0.2				
7	DDDEK	3	1				1			0.6	0.2				0.2		
8	KKKKH					1	4							0.2	0.8		

モチーフ・プロフィールを用いた類似性

より大きなグループ(スーパーファミリー)にまとめようとした場合、弱い相同性をより正確に認識できる類似性を採用する必要

→近縁の配列群のマルチプルアライメントから、このファミリーの本質的な特徴を見出したい

```

5p21-  MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSY
1ctqA  MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSY
1clyA  MREYKLVVLGSGGVGKSALTVQFVQGFVVEKYDPTIEDSY
1kao-  MREYKVVVLGSGGVGKSALTVQFVTGTFIEKYDPTIEDFY
1huqA  --QFKLVLLGESAVGKSSLVLRVFKGQFHEYQESTIGAAF
1g16A  ----KILLIGDSGVGKSCLLVRFVE----DKFNPI--DFK
1ek0A  VTSIKLVLLGEAAVGKSSIVLRFVSNDFAEENKEPTIGAAF
3rabA  ---FKILIIIGNSSVGKTSFLFRYADDSFTPAFVSTVGIDF
1mh1-  ----KCVVVGDGAVGKTCLLISYTTNAFPGEYIPTVFDNY
2ngrA  MQTIKCVVVGDGAVGKTCLLISYTTNKFPSEYVPTVFDNY
1tx4B  ----KLIVVGDGACGKTCLLIVNSKDQF---YVPTVFENY
    
```

サイトごとに保存の度合いに差がある。
 サイトごとにアミノ酸の出現傾向に差がある [AG]-x(4)-G-K-[ST]

モチーフ解析

- 正規表現風のパターンで、局所的な配列のパターンを表現。

PROSITE(<http://www.expasy.ch/prosite/>)が有名

1.進化的に保存している局所配列パターン

- ・マルチプルアライメント由来
- ・保存しているサイト→機能的に重要なサイト→活性部位

2.機能的な局所配列パターン

- ・リン酸化サイト、N-ミリスチル化サイトなど

PROSITEのモチーフの記述法

(例) ATP_GTP_A : [AG]-x(4)-G-K-[ST]

2FE2S FERREDOXIN:

C-{C}-{C}-[GA]-{C}-C-[GAST]-{CPDEKRHFYW}-C

ZINC_FINGER_C2H2_1:

C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H

x	: 任意のアミノ酸
x(n)	: n個の任意のアミノ酸
x(n, m)	: nからm個の任意のアミノ酸
[ACD]	: AかCかDのいずれかのアミノ酸
{ACD}	: AでもCでもDでもないアミノ酸

x : 任意のアミノ酸
x (n) : n個の任意のアミノ酸
x (n,m) : nからm個の任意のアミノ酸
[ACD] : AかCかDのいずれかのアミノ酸
{ACD} : AでもCでもDでもないアミノ酸

(3)以下のPROSITEのモチーフに適合する箇所を口で囲め

1) [AG]-x(4)-G-K-[ST]
 >5p21-
 M T E Y K L V V V G A G G V G K S A L T I Q L I Q N H F V D E Y D P T I
 E D S Y R K Q V V I D G E T C L L D I L D T A G Q E E Y S A M R D Q Y M
 R T G E G F L C V F A I N N T K S F E D I H Q Y R E Q I K R V K D S D D
 V P M V L V G N K C D L A A R T V E S R Q A Q D L A R S Y G I P Y I E T
 S A K T R Q G V E D A F Y T L V R E I R Q H

2) C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H
 >ZN428_HUMAN
 R G G P S R R A P R A A Q P P A Q P C Q L C G R S P L G E A P P G T P P
 C R L C C P A T A P Q E A P A P E G R A L G E E E E E P P R A G E G R P
 A G R E E E E E E E E E G T Y H C T E C E D S F D N L G E L H G H F M L
 H A R G E V

3) [GA]-x(0,2)-[YSA]-x(0,1)-[VFY]-x-C-x(1,2)-[PG]-x(0,1)-H-x(2,4)-[MQ]
 >PLAS_ORYSI
 V F E P N D F T V K S G E T I T F K N N A G F P H N V V F D E D A V P S
 G V D V S K I S Q E E Y L N A P G E T F S V T L T V P G T Y G F Y C E P
 H A G A G M V G K V T V N

x : 任意のアミノ酸
x (n) : n個の任意のアミノ酸
x (n,m) : nからm個の任意のアミノ酸
[ACD] : AかCかDのいずれかのアミノ酸
{ACD} : AでもCでもDでもないアミノ酸

(3)以下のPROSITEのモチーフに適合する箇所を口で囲め

1) [AG]-x(4)-G-K-[ST]
 >5p21-
 M T E Y K L V V V G A G G V G K S A L T I Q L I Q N H F V D E Y D P T I
 E D S Y R K Q V V I D G E T C L L D I L D T A G Q E E Y S A M R D Q Y M
 R T G E G F L C V F A I N N T K S F E D I H Q Y R E Q I K R V K D S D D
 V P M V L V G N K C D L A A R T V E S R Q A Q D L A R S Y G I P Y I E T
 S A K T R Q G V E D A F Y T L V R E I R Q H

2) C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H
 >ZN428_HUMAN
 R G G P S R R A P R A A Q P P A Q P C Q L C G R S P L G E A P P G T P P
 C R L C C P A T A P Q E A P A P E G R A L G E E E E E P P R A G E G R P
 A G R E E E E E E E E E G T Y H C T E C E D S F D N L G E L H G H F M L
 H A R G E V

3) [GA]-x(0,2)-[YSA]-x(0,1)-[VFY]-x-C-x(1,2)-[PG]-x(0,1)-H-x(2,4)-[MQ]
 >PLAS_ORYSI
 V F E P N D F T V K S G E T I T F K N N A G F P H N V V F D E D A V P S
 G V D V S K I S Q E E Y L N A P G E T F S V T L T V P G T Y G F Y C E P
 H A G A G M V G K V T V N

x : 任意のアミノ酸
x (n) : n個の任意のアミノ酸
x (n, m) : nからm個の任意のアミノ酸
[ACD] : AかCかDのいずれかのアミノ酸
{ACD} : AでもCでもDでもないアミノ酸

(3)以下のPROSITEのモチーフに適合する箇所を口で囲め

1) **[AG]-x(4)-G-K-[ST]**
 >5p21-
 MTEYKLVVV GAGGVGKSALTIQLIQNHFVDEYDPTI
 EDSYRKQVVIDGETCLLDILDTAGQEEYSAMRDQYM
 RTGEGFLCVFAINNTKSFEDIHQYREQIKRVKDSDD
 VPMVLVGNKCDLAARTVESRQAQDLARSYGIPYIET
 SAKTRQGVEDAFTLVREIRQH

2) **C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H**
 >ZN428_HUMAN
 RGGPSRRAPRAAQPPAQP CQLCGRSPLGEAPPPTPP
CRLCCPATAPQEAPAPEGRALGEEEEEPPRAGEGRP
 AGREEEEEEEEEEGTYH CTECEDSF DN LGELHGHFML
HARGEV

3) **[GA]-x(0,2)-[YSA]-x(0,1)-[VFY]-x-C-x(1,2)-[PG]-x(0,1)-H-x(2,4)-[MQ]**
 >PLAS_ORYSI
 VFEPNDFTVKSGETITFKNNAGFP HNVVFD EDAVPS
 GVDVSKISQEEYL N A PGETFSVTLTVPGTYGFYCEP
HAGAGMVGKVTVN

x : 任意のアミノ酸
x (n) : n個の任意のアミノ酸
x (n, m) : nからm個の任意のアミノ酸
[ACD] : AかCかDのいずれかのアミノ酸
{ACD} : AでもCでもDでもないアミノ酸

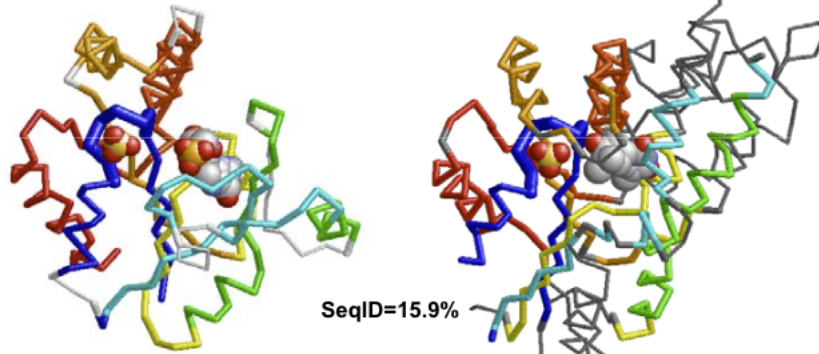
(3)以下のPROSITEのモチーフに適合する箇所を口で囲め

1) **[AG]-x(4)-G-K-[ST]**
 >5p21-
 MTEYKLVVV GAGGVGKSALTIQLIQNHFVDEYDPTI
 EDSYRKQVVIDGETCLLDILDTAGQEEYSAMRDQYM
 RTGEGFLCVFAINNTKSFEDIHQYREQIKRVKDSDD
 VPMVLVGNKCDLAARTVESRQAQDLARSYGIPYIET
 SAKTRQGVEDAFTLVREIRQH

2) **C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H**
 >ZN428_HUMAN
 RGGPSRRAPRAAQPPAQP CQLCGRSPLGEAPPPTPP
CRLCCPATAPQEAPAPEGRALGEEEEEPPRAGEGRP
 AGREEEEEEEEEEGTYH CTECEDSF DN LGELHGHFML
HARGEV

3) **[GA]-x(0,2)-[YSA]-x(0,1)-[VFY]-x-C-x(1,2)-[PG]-x(0,1)-H-x(2,4)-[MQ]**
 >PLAS_ORYSI
 VFEPNDFTVKSGETITFKNNAGFP HNVVFD EDAVPS
 GVDVSKISQEEYL N A PGETFSVTLTVPGTYGFYCEP
HAGAGMVGKVTVN

P-loopモチーフ: [AG]-x(4)-G-K-[ST] の立体構造



1gky:Guanilate Kinase
(8-15:GPSGTGKS)

1e2kA:Thymidine Kinase
(56-63:GPHGMGKT)

- ・ P-loopモチーフは、ヌクレオチドのリン酸基結合サイトに対応
- ・ モチーフ以外の領域も、立体構造は似ている

プロフィール法

マルチプルアライメントからサイトごとのスコア行列を作成。
これに対して動的計画法等を用いて配列をアライメント。

サイトごとのスコア行列



プロフィール(Profile)

位置特異的スコア行列

(PSSM; Position Specific Score Matrix)

	1	2	3	4	5	6	..
A	3	-1	-3	-4	6	-4	..
Q	0	3	-1	-2	-4	0	..
H	-3	-3	-4	11	-4	4	..
:	:	:	:	:	:	:	:
V	-4	-2	-1	-6	-2	-4	..

HMMer

マルチプルアライメントを入力とする。隠れマルコフモデル(HMM)を使用しているため、表現力はPSI-BLASTより高いはずだが、計算速度は遅い。PfamはHMMerを採用している。

PSI-BLAST

BLASTの拡張版。反復的にデータベース検索を行うことで、厚いマルチプルアライメントを生成する。

マルチプルアライメント

	1	2	3	4	5	6	7	8	9	..
query	A	Q	S	H	A	T	K	H	K	..
homolog1	A	N	S	H	A	T	K	H	K	..
homolog2	S	G	K	H	A	K	S	F	Q	..
homolog3	A	R	K	H	G	E	-	L	L	..
homolog4	S	D	L	H	A	H	-	L	R	..

良質のマルチプルアライメントを作るには淡い相同性の配列を集め、アラインする必要がある。それには、よいプロフィールが不可欠

プロフィール

	1	2	3	4	5	6	7	8	..
A	Q	S	H	A	T	K	H

A	3	-1	-3	-4	6	-4	-3	-4	..
G	-2	-1	-5	-5	-1	-4	-2	-6	..
H	-3	-3	-4	11	-4	4	-3	6	..
:	:	:	:	:	:	:	:	:	..
V	-4	-2	-1	-6	-2	-4	-4	-2	..

良質のプロフィールを作るにはできるだけ多くの配列を集めたマルチプルアライメントが必要

堂々巡りの関係

