

# 2016年度「バイオインフォマティクス」 講義予定表

1. 配列アライメント (五斗)
2. データベースサーチ (五斗)
3. 機能・構造予測 (五斗)
4. 配列モチーフ・隠れマルコフモデル (五斗)
5. アセンブリとマッピング (Blanc-Mathieu)
6. 分子進化 (緒方)
7. 分子系統解析 (緒方)
8. 分子系統解析【演習】 (Blanc-Mathieu)
9. 遺伝子予測 (緒方)
10. 機能アノテーション・比較ゲノム (緒方)
11. システムズバイオロジー (五斗)
12. メタゲノム解析 (緒方)
13. メタゲノム解析 (緒方)

<http://goto.kuicr.kyoto-u.ac.jp/lecture/bioinfo.html>

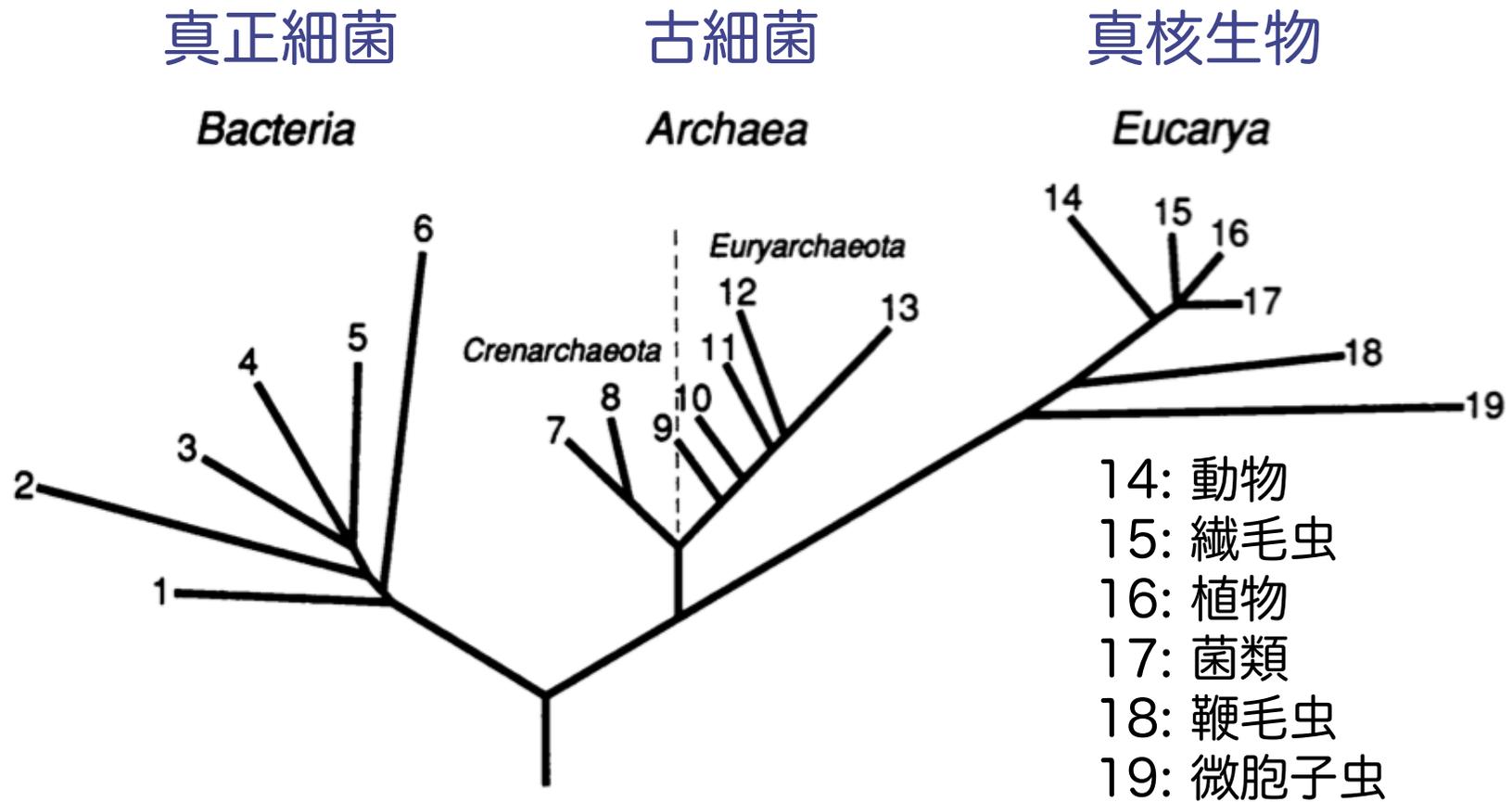
## 配列アライメント - 今日のトピック

- ◆ ドットマトリックス (Dot Matrix)
- ◆ ダイナミックプログラミング (Dynamic Programming)
- ◆ スコアマトリックス・ギャップペナルティ (Score Matrix, Gap Penalty)

# 配列アライメント

- ◆ 配列アライメント (sequence alignment)
  - 2つのタンパク質または遺伝子の配列を並べて、**進化的な**関連があるかどうかを調べること
- ◆ 2つの遺伝子が進化的に関連があるか？
  - 異なる生物種間で同じ**機能**を持つ遺伝子
  - 一つの生物種内で類似した**機能**を持つ遺伝子
- ◆ 異なる生物種で同じ機能の遺伝子を比較する
  - 生物種間の進化的な関係

# 生物種の系統関係



ウーズらによる進化系統樹  
分子レベル（配列レベル）の情報：16S rRNA

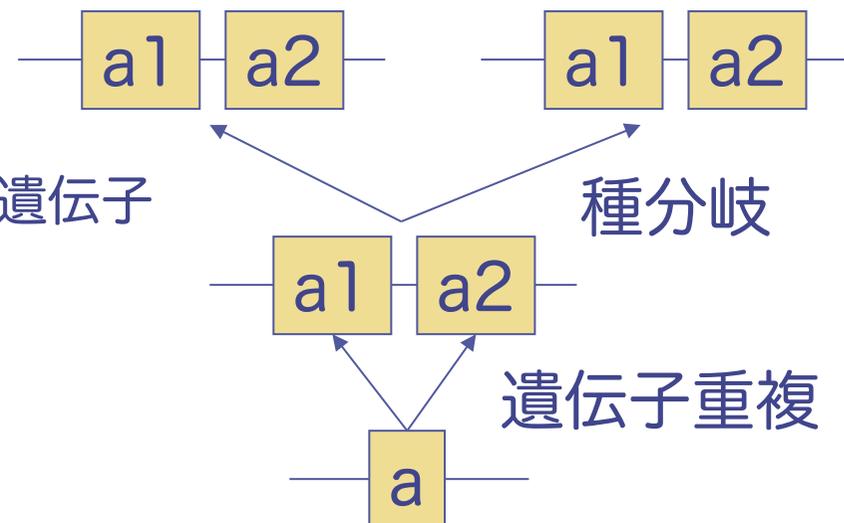
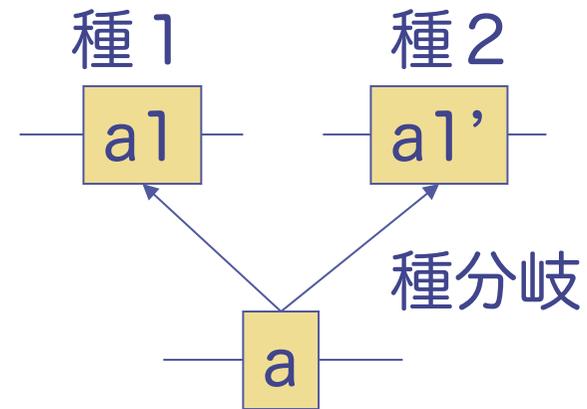
# オーソログとパラログ

ホモログ (Homolog)  
進化的な起源を同じくする遺伝子

オーソログ (Ortholog)  
種分岐の際に同じ遺伝子だったもの  
通常同じ機能を持つ

パラログ (Paralog)  
遺伝子重複によってできた類似遺伝子  
通常異なる機能を持つ

ゼノログ (Xenolog)  
水平移動によって得られた類似遺伝子



# 配列アライメント

## ◆ 配列アライメント (alignment)

- 2つのタンパク質または遺伝子の配列を並べて、**ホモログ (相同)** かどうかを調べること
- 実際には類似性を調べる
- 文字の一致 (マッチ)、不一致 (ミスマッチ)、挿入、欠失を考慮する

## ◆ アライメントのキーポイントは

- アライメントの種類
- アライメントの方法・アルゴリズム
- アライメントを評価するためのスコア
- スコアの重要性を評価するための統計的基準

## 配列アライメントの種類

### ◆ グローバルアライメント

- 配列全体を並べる

LG	PSSKQ	TG	KGW	-	SRI	WDN	
	+				+		
LN	-	IT	SAG	KG	AI	MRL	GDA

マッチ (+, |)  
ミスマッチ  
ギャップ・挿入 (-)

### ◆ ローカルアライメント

- 局所的によく似た部分を探す

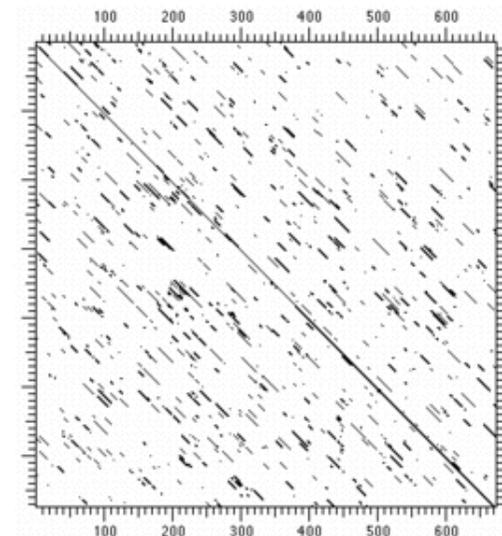
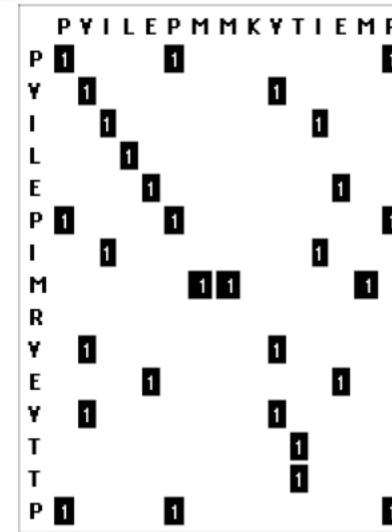
-----	TG	KG	-----
-----	AG	KG	-----

## アライメント方法の種類

- ◆ 2本の配列のアライメント
  - ペアワイズアライメント
    - ◆ ドットマトリックス
    - ◆ ダイナミックプログラミング
    - ◆ ワードを使った方法
  
- ◆ 3本以上の配列のアライメント
  - マルチプルアライメント
    - ◆ 累進法
    - ◆ 反復改善法

# ドットマトリックス

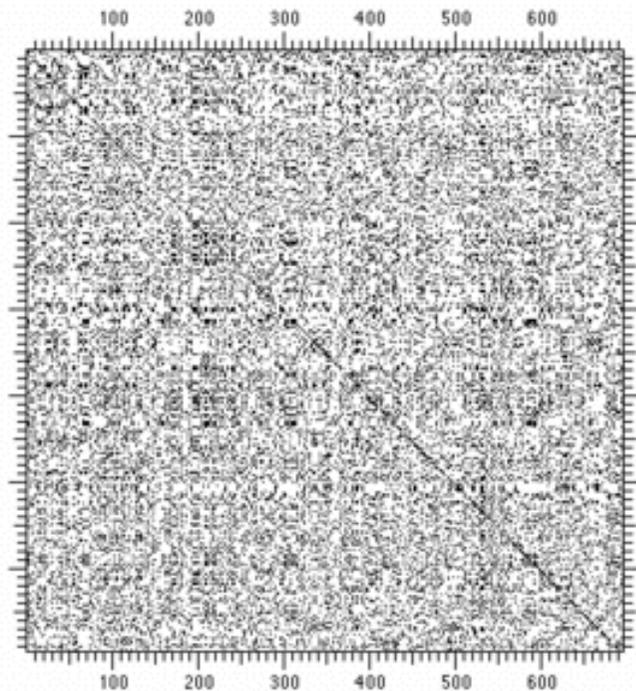
- ◆ 比較する配列を行列の行と列に並べて一致する文字に対応する行列成分にドットをうつ
- ◆ 類似部分をマトリックスの対角成分として表示
- ◆ 同一配列をマトリックス表示するとリピート部分も抽出できる



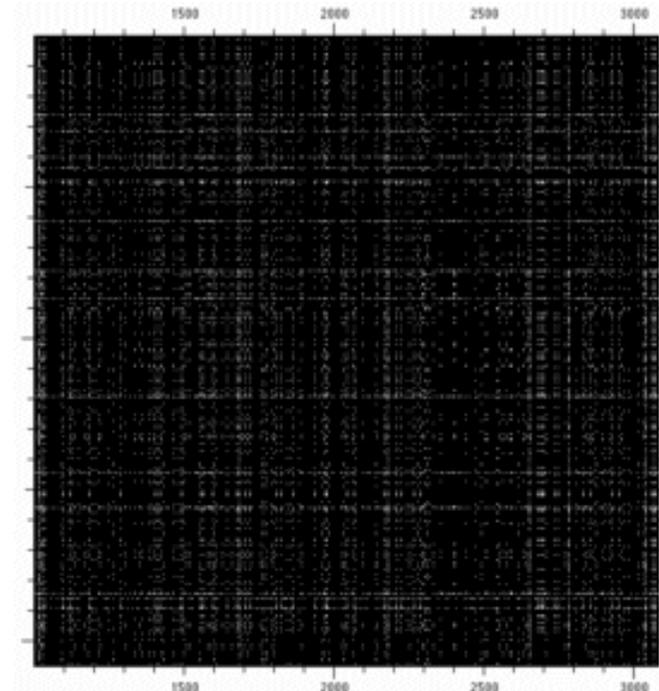
<http://www.tulane.edu/~biochem/lecture/723/seqalign.html>

# ドットマトリックス

二つのバクテリアの EF-G (転写伸長因子) の例  
*S. aureus* (黄色ブドウ球菌) と *T. thermophilus* (好熱菌)



タンパク質アミノ酸配列



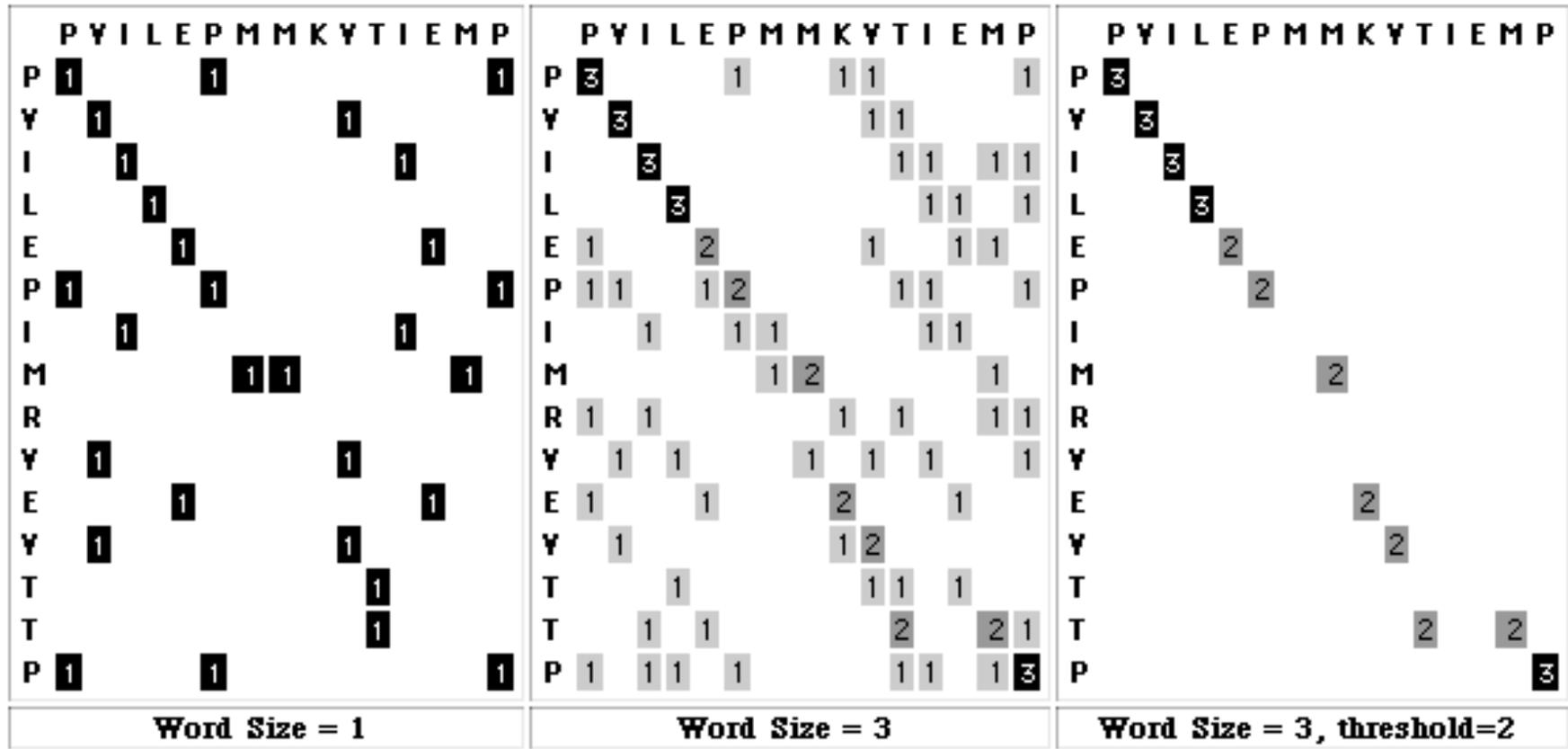
DNA塩基配列

<http://www.tulane.edu/~biochem/lecture/723/seqalign.html>

<b>Entry</b>	SA0505	<b>Entry</b>	tlr1749	CDS	T.elongatus
<b>Gene name</b>	fus	<b>Gene name</b>	fus		
<b>Definition</b>	elongation fa	<b>Definition</b>	translation elongation factor EF-G [EC:3.6.5.3]		
<b>Orthology</b>	KO: K02355 e	<b>Orthology</b>	KO: K02355 elongation factor EF-G		
<b>Class</b>	<a href="#">BRITE hierarchy</a>	<b>Class</b>	<a href="#">BRITE hierarchy</a>		
<b>SSDB</b>	<a href="#">Ortholog</a> <a href="#">Para</a>	<b>SSDB</b>	<a href="#">Ortholog</a> <a href="#">Paralog</a> <a href="#">Gene cluster</a>		
<b>Motif</b>	Pfam: <a href="#">GTP_EFT</a> PROSITE: <a href="#">EFAC</a> <a href="#">Motif</a>	<b>Motif</b>	Pfam: <a href="#">GTP_EFTU</a> <a href="#">MMR_HSR1</a> <a href="#">Miro</a> <a href="#">GTP_EFTU_D2</a> <a href="#">EFG_IV</a> <a href="#">EFG_C</a> PROSITE: <a href="#">EFACTOR_GTP</a> <a href="#">Motif</a>		
<b>Other DBs</b>	NITE: <a href="#">SA0505</a> NCBI-GI: <a href="#">1592</a> NCBI-GeneID: <a href="#">UniProt: P687</a>	<b>Other DBs</b>	CYORF: <a href="#">tlr1749</a> Kazusa: <a href="#">tlr1749</a> NCBI-GI: <a href="#">22299292</a> NCBI-GeneID: <a href="#">1011972</a> <a href="#">UniProt: Q8DI43</a>		
<b>LinkDB</b>	<a href="#">PDB</a> <a href="#">All D</a>	<b>LinkDB</b>	<a href="#">PDB</a> <a href="#">All DBs</a>		
<b>Position</b>	588492..59057	<b>Position</b>	1829639..1831714	<a href="#">Genome map</a>	
<b>AA seq</b>	693 aa <a href="#">AA se</a> MAREFSLEKTRNI ITITSAATTAWE RQATTYGVPRIVF MKCFKYTNDLGTE KEAIRQATTNVEF AKADDSAEFAALA NSRQEIDTVYSGD DKMTQALVKLQEE SYRETFKSSAQVQ VEAGLKDAMENGV LEPMMKVTIEMPE NTQGRGTYTMYFD	<b>AA seq</b>	691 aa <a href="#">AA seq</a> <a href="#">DB search</a> MARTTPLERVRNIGIAAHIDAGKTTTTTERILFYSGVVHKIGEVHEGTTVTDWMEQERERG ITITAAAISTSWRDHQINIIDTPGHVDFTIEVERSMRVLGVIIVFCSVGGVQPQSETVW RQADRYSVPRIVFVNKMDRTGANFYKVHDQIRDLRANAVPIQLPIGAEDQFKGIVDLVR MRAKIYKDDLKGEIEDTEIPAEMTELAQEYRTKLEAVALAETDDALMEKYFEHEELTEEEI RAALRKGTIAGTIVPMLCGSAFKNKGVQLLLDAVVDYLPAPIDIPAIGRLPDGTEVERA ADDDQPLAALAFKIMSDPYGRLTFVRVYSGVLKKGSYVLNATKGKKERISRLIVLKADER IEVDEL RAGDLGAALGLKETFTGDTLCESSPVILESLYIPEPVISVAVEPKTKQDMEKL SKALQALSEEDPTFRVSVDPETNQTVIAGMGELHLEILVDRMQREYKVEANIGQPQVAYR ETIRKPVRAEGKFIHQSGGKQYGHVVIEVEPAEPTGFEFVSKIIVGGVVPKEYIPPAAEQ GMKEACESGILAGYPVIDLKVTLVDGSYHEVDSSEMAFKIAGSIAIKEAVMKANPVLLEP MMKVEVEVPEEFLGTVMGDLIARRGQIEGQTVENGIKAVTAKVPLERMFGYATDIRSNTQ GRGIFSMESHYEEVPRNVAEAI IAKNKGNA		

<p><b>NT seq</b></p>	<p>2082 nt <span>NT seq</span> +upstream <input type="text" value="0"/> nt +downstream <input type="text" value="0"/> nt</p>	<p><b>NT seq</b></p>	<p>2076 nt <span>NT seq</span> +upstream <input type="text" value="0"/> nt +downstream <input type="text" value="0"/> nt</p>
	<p>atggctagagaatt gctggtaaaacgac ggtgaaacacacga attactatcacatc gatacacctggaca ggagcagttacagt cgtcaggctacaac ggtgctaacttcga ccaatccaattacc atgaaatgtttcaa gaagaccacttaga actagcgacgaatt aaagaagctatccg gctttcaaaaacaa ccactagacgtaa gcaaaagcagacga tatgttggtaaatt gttaagaactctac aactcacgtcaaga aaagatacaggtac atggaattcccaga gataaaatgactca actgacgaagaac ttagtagaccgtat tcataatcgtgaaac ggtggctcgtggtca ggtttcgaattcga gtagaagctggtct gatgttaaagctaa ttcaaaattgctgc ttagaaccaatgat ggtgacgtaacatc ggtgtaaatgctta aacactcaaggtcg tcaatcgtgaaga</p>		<p>gtggcacggacgaccccgctagagcgagtgcgaaatatcgggattgccgctcacattgat gcgggtaaaacaaccacaactgaacgtattctcttctattccgggtggtgacaaaatt ggtgaagtgcacgaggggaccacggttaccgactggatggaacaggagcgagagcggggc atcaccattacagcagccgccattagcacctcttggagagatcaccaaatacaacattatt gacactcccggccacgtggacttcacgattgaggtggagcgcctccatgccgggttctcgat gggggtgattgccggtattctgctccggttgggtggtgagcagcccaatcagaaaccgtgtgg cggcaagccgatcgctacagcgttccccgcacggtttttgtcaacaagatggatcgacg ggggcaaaacttctacaaagtccatgaccagattcgcgatcgctacggggcaacgccgtg ccgattcagttgccattggtgctgaagatcaattcaaaggcatcgttgatctggtgagc atgctgcgcaaaatctacaaagacgaccttggcaaggaaatcgaagacactgaaattcct gccgagatgactgaactggcacaggagatcgcaccaagctcatcgaggctgttgccgaa accgacgacgccctcatggaaaaatacttcgagggggaagaactcacagaagaggaaatc cgtgcccgtctacgcaaaaggacgatcgctggcaccattgtgcccattgctttgcccgtct gccttcaaaaataaaaggcgtgcaattactcctggatgccgtggtggactacctaccggcg cccattgacattcctgccattaaaggcgtctgcccgatggcacagaagtggagcgggct gccgatgacgatcaaccctggcggcactcgttttaagatcatgctggatccctatggc cgcttacctttgtgctggtttattccgggtgtctgaaaaaggtagctatgtgctcaat gccaccaaaaggcaagaagagcgcatttctcgtcctaatacgtgctgaaggccgatgaacgg attgaggtggatgaattgagcagctggcgatctaggagctgccttaggtctcaaggaaacc ttcacaggggataaccctctgagatgaaagctcgcggatgattcttgagtcgctgtatatt ccagagccagttatttccgtggccgtagaacccaaaactaaacaggacatggaaaagctc tcaaggctctgcaagccctctctgaggaggacccaccttccgcgctcagcgttgatcct gaaacgaaccaaacctgattgccgggaatgggggaactacacctcgagattcttgtggat cggatgcagcgggagtagaaagtggaaagcaacatcggccaaccccaagtggcttaccgc gaaaccattcgcaaacccgtccgggctgaaggcaagtttatccgtcaaagtggtggtaaa ggtcagataggtcagcttgttattgaagtgaacctgctgaacccggtactgggtttgaa ttgtctccaaaattgtcgggtggtgctgacccaagagtagattccaccgctgagcag gggatgaaggaaagcctgtgaatcggggattttggcagggtagccagtgattgacctcaa gtcacactagtggtatggctcctaccacgaggtggactcctcagagatggctttcaaaatt gctggctccatcgccattaaagaggcagtcagaaagccaatccagtagctgctggagcca atgatgaaggtcgaggtcgaggtccagaagaattccttgggacagtgatgggagatcta attgcccgctcgcggtcaaattgagggacaaaccgtagaaaatggcattgccaaagtgcg gcaaaagttcctctcgaacggatgtttgggtatgccactgacattcgctcgaatacccaa ggtcgggggaatttctcgtatggaatttagccattacgaggaagtccccgtaatgtggct gaggcgatcattgccaaaaataaagggaaacgcatag</p>

# ワードサイズによる評価

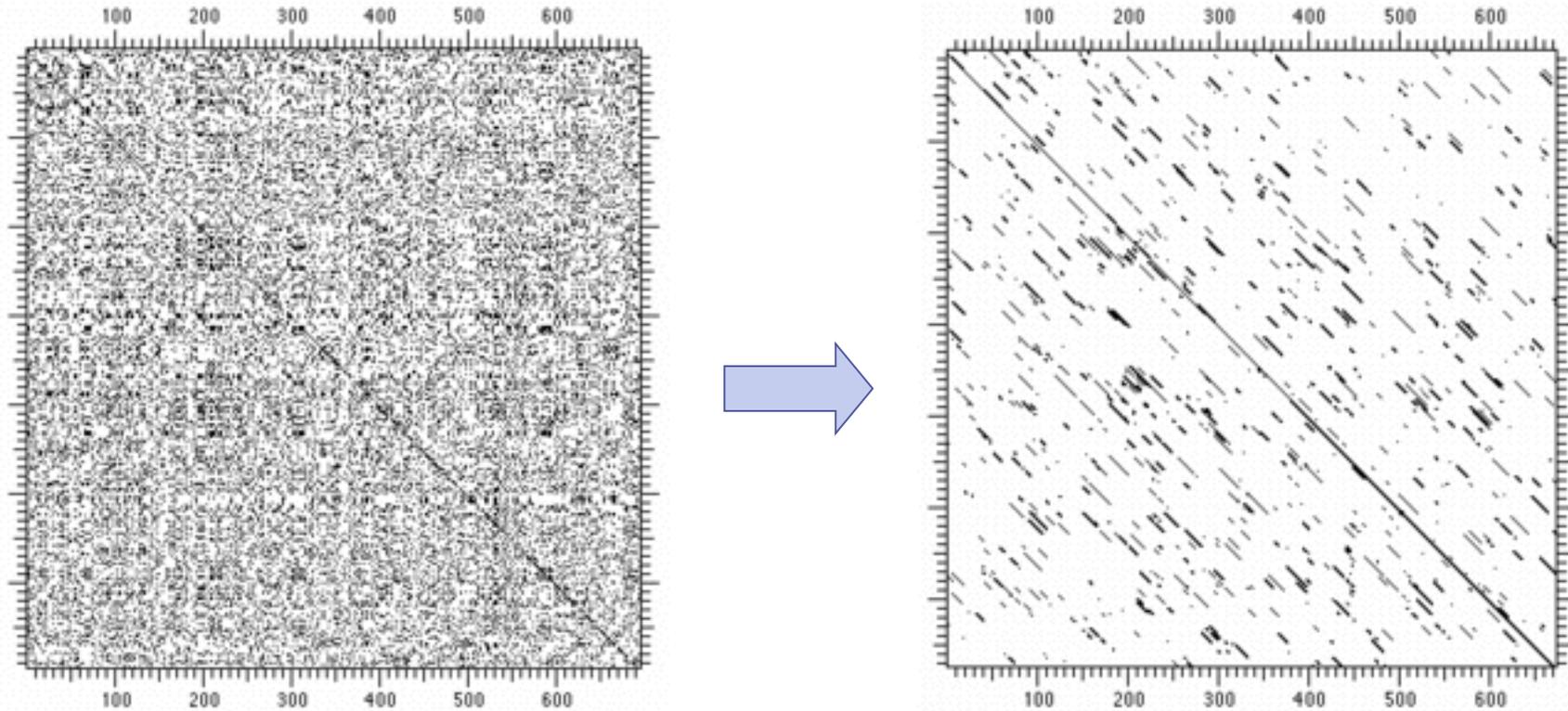


1文字ずつ比較して、一致するところにドット

3文字ずつ比較して、完全一致するところにドット

3文字ずつ比較して、そのうち2文字一致するところにドット

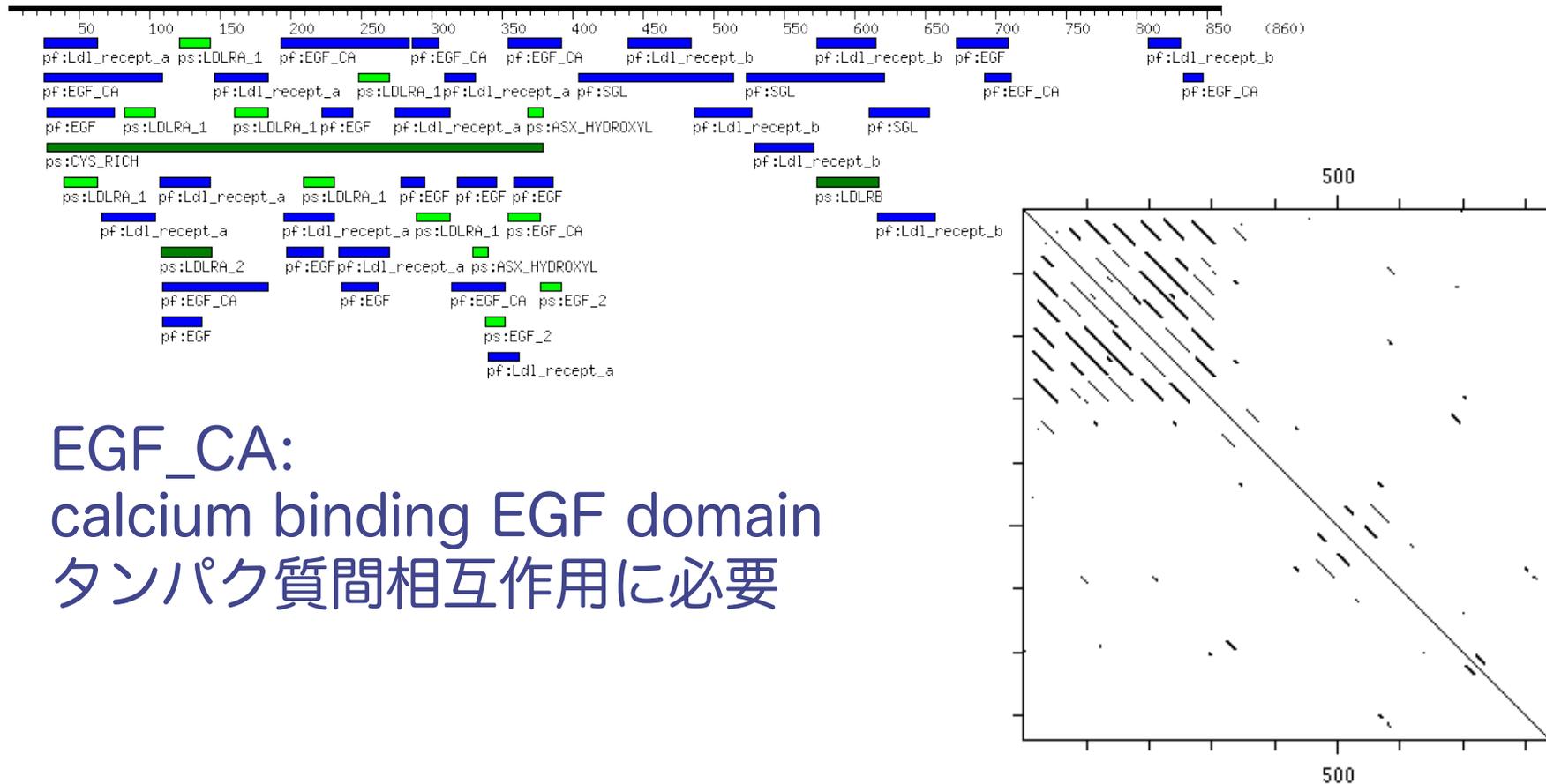
## ワードサイズによる評価



ワードサイズ 23、閾値 5（23文字のうち5文字）の  
マッチでドット

# ドットマトリックスによるリピート同定

## human LDL (low density lipoprotein) receptor



EGF\_CA:  
calcium binding EGF domain  
タンパク質間相互作用に必要

<http://www.blc.arizona.edu/courses/bioinformatics/matrix-lect.html>

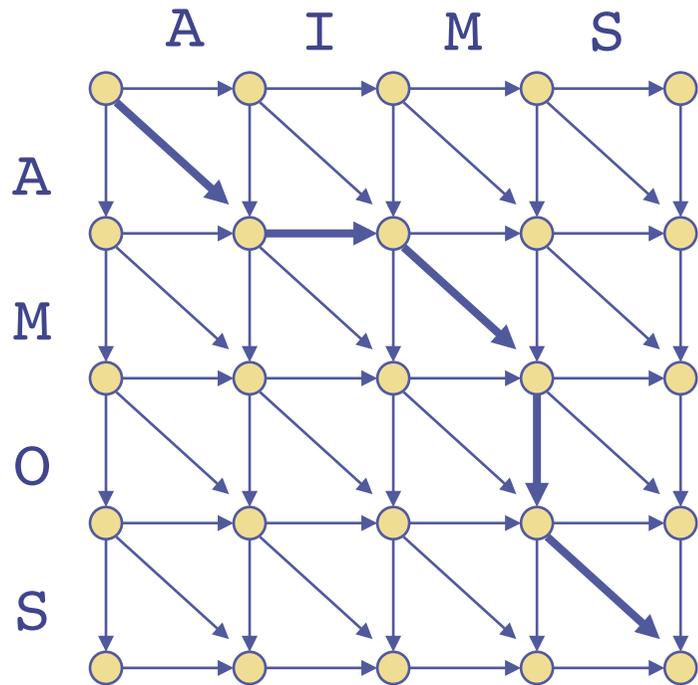
## アライメント方法の種類

- ◆ 2本の配列のアライメント
  - ペアワイズアライメント
    - ◆ ドットマトリックス
    - ◆ ダイナミックプログラミング
    - ◆ ワードを使った方法
  
- ◆ 3本以上の配列のアライメント
  - マルチプルアライメント
    - ◆ 累進法
    - ◆ 反復改善法

# ダイナミックプログラミング (DP)

ドットマトリックスでは対角線成分だけ

→ ギャップも考慮したい (塩基の挿入・欠失)



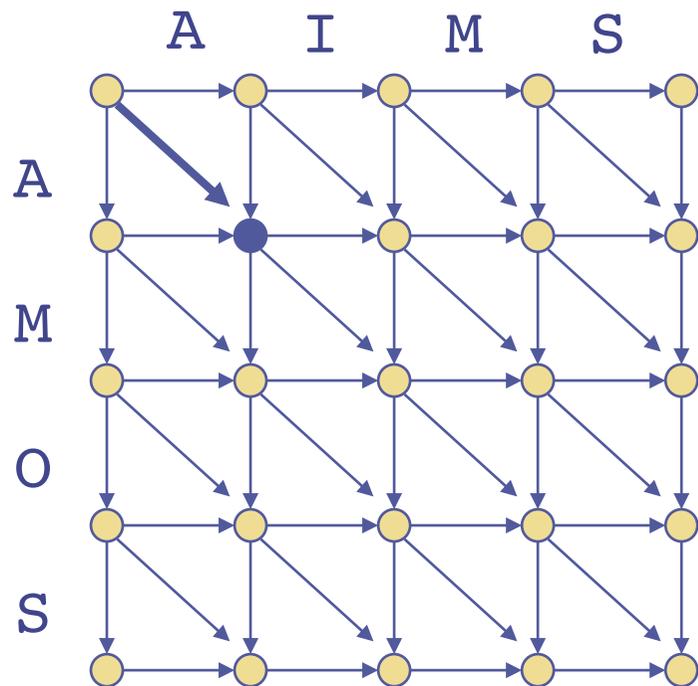
全パターンの並べ方から最適なものを探し出す問題

全パターン

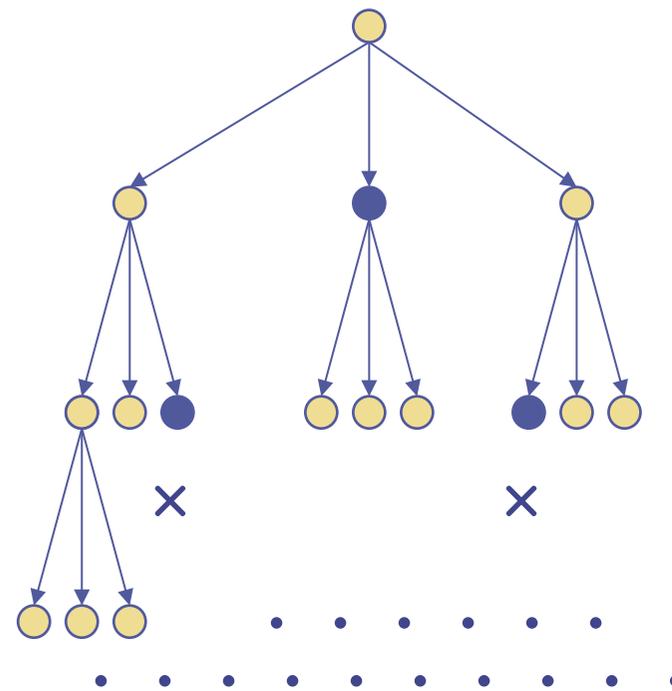
左上から右下まで辿る経路  
縦と横の矢印がギャップ

アライメント AIM-S  
A-MOS

# ダイナミックプログラミング (DP)

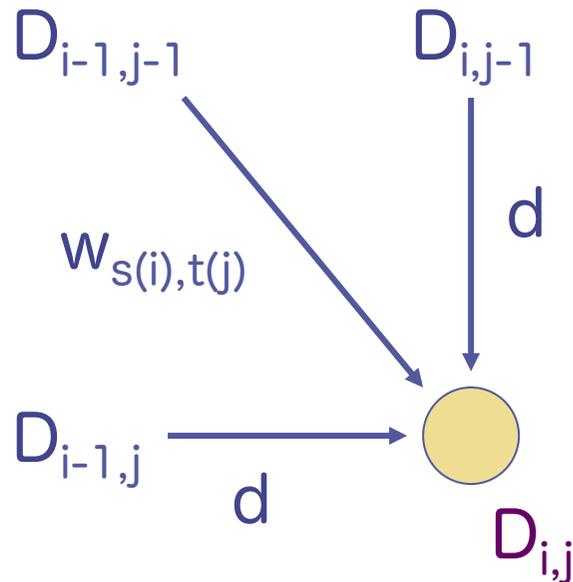


DP計算用の行列D



評価関数による枝刈り

# 評価関数（スコア）の計算方法



$$D_{i,j} = \max \left( \begin{array}{l} D_{i-1,j-1} + W_{s(i),t(j)}, \\ D_{i-1,j} + d, \\ D_{i,j-1} + d \end{array} \right)$$

	1	2	3	4
文字列 s	A	I	M	S
文字列 t	A	M	O	S

$W_{s(i),t(j)}$ :

アミノ酸の場合、 $s(i)$  と  $t(j)$  の類似度

塩基配列の場合、通常マッチ 1， mismatch 0

$d$ : ギャップペナルティ

# グローバルアライメント

ニードルマンとブンシュによる評価関数の計算方法

$$D_{0,0} = 0$$

$$D_{i,0} = i d$$

(for  $i = 1$  to  $n$ )

$$D_{0,j} = j d$$

(for  $j = 1$  to  $m$ )

$$D_{i,j} = \max ($$

$$D_{i-1,j-1} + W_{s(i),t(j)},$$

$$D_{i-1,j} + d,$$

$$D_{i,j-1} + d)$$

		A	T	A	T	G	C	C	T	T
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	1	0	-1	-2	-3	-4	-5	-6	-7
A	-2	0	1	1	0	-1	-2	-3	-4	-5
G	-3	-1	0	1	1	1	0	-1	-2	-3
C	-4	-2	-1	0	1	1	2	1	0	-1
C	-5	-3	-2	-1	0	1	2	3	2	1
T	-6	-4	-2	-2	0	0	1	2	4	3
G	-7	-5	-4	-2	-1	1	0	1	3	4
T	-8	-6	-4	-3	-1	0	1	0	2	4

ここでは  $d = -1$  で計算

max の計算で使った経路を覚えておく

# グローバルアライメント

ニードルマンとブンシュによる評価関数の計算方法

$$D_{0,0} = 0$$

$$D_{i,0} = i d$$

(for  $i = 1$  to  $n$ )

$$D_{0,j} = j d$$

(for  $j = 1$  to  $m$ )

$$D_{i,j} = \max ($$

$$D_{i-1,j-1} + W_{s(i),t(j)},$$

$$D_{i-1,j} + d,$$

$$D_{i,j-1} + d)$$

		A	T	A	T	G	C	C	T	T
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	1	0	-1	-2	-3	-4	-5	-6	-7
A	-2	0	1	1	0	-1	-2	-3	-4	-5
G	-3	-1	0	1	1	1	0	-1	-2	-3
C	-4	-2	-1	0	1	1	2	1	0	-1
C	-5	-3	-2	-1	0	1	2	3	2	1
T	-6	-4	-2	-2	0	0	1	2	4	3
G	-7	-5	-4	-2	-1	1	0	1	3	4
T	-8	-6	-4	-3	-1	0	1	0	2	4

ATATGCCT-T

A-A-GCCTGT

右下 (右端) から覚えていた経路を辿って戻る (バックトラック)

# ローカルアライメント

スミスとウォーターマンによる評価関数の計算方法

$$D_{0,0} = 0$$

$$D_{i,0} = 0$$

(for  $i = 1$  to  $n$ )

$$D_{0,j} = 0$$

(for  $j = 1$  to  $m$ )

$$D_{i,j} = \max ($$

$$D_{i-1,j-1} + W_{s(i),t(j)},$$

$$D_{i-1,j} + d,$$

$$D_{i,j-1} + d,$$

$$0)$$

		A	T	A	T	G	C	C	T	T
		0	0	0	0	0	0	0	0	0
A	0	1	0	1	0	0	0	0	0	0
A	0	0	1	1	1	0	0	0	0	0
G	0	0	0	1	1	2	1	0	0	0
C	0	0	0	0	1	1	3	2	1	0
C	0	0	0	0	0	1	2	4	3	2
T	0	0	1	0	1	0	1	2	5	4
G	0	0	0	1	0	2	1	1	4	5
T	0	0	1	0	2	1	2	1	3	5

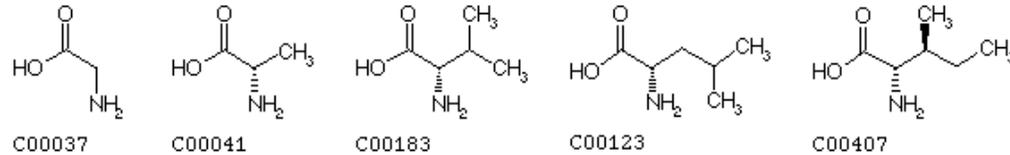
ATGCCT-T

AAGCCTGT

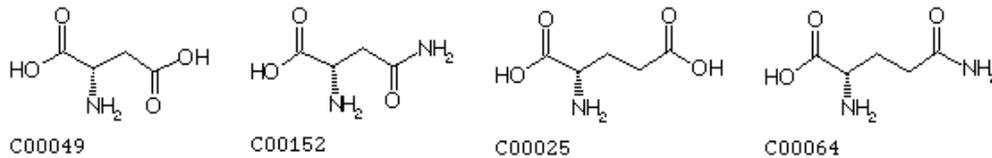
スコア最大の箇所から覚えていた経路を辿って戻る (バックトラック)

# アミノ酸の構造と類似度

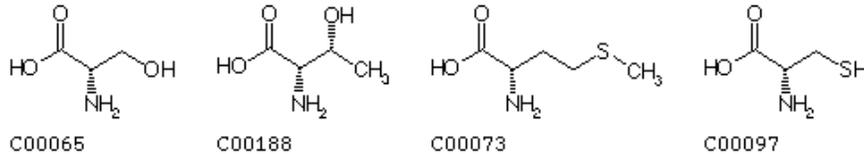
Glycine (Gly)   Alanine (Ala)   Valine (Val)   Leucine (Leu)   Isoleucine (Ile)



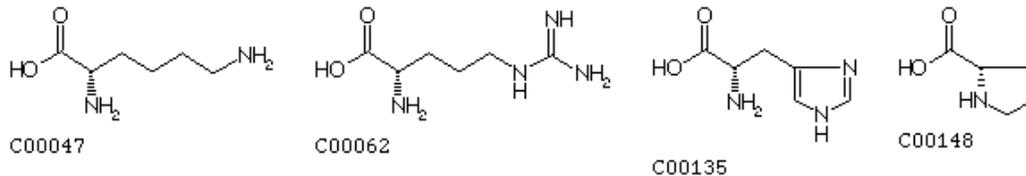
Aspartate (Asp)   Asparagine (Asn)   Glutamate (Glu)   Glutamine (Gln)



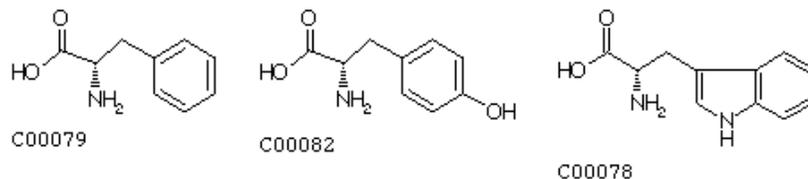
Serine (Ser)   Threonine (Thr)   Methionine (Met)   Cysteine (Cys)



Lysine (Lys)   Arginine (Arg)   Histidine (His)   Proline (Pro)



Phenylalanine (Phe)   Tyrosine (Tyr)   Tryptophan (Trp)



- 疎水性、親水性
- 電荷の有無
- 残基のサイズ
- コドン使用パターン

# アミノ酸置換行列

Ala	1																			
Arg	-5	7																		
Asn	-5	-1	7																	
Asp	-5	-2	2	8																
Cys	-5	-4	-2	-4	13															
Gln	-5	1	0	0	-3	7														
Glu	-5	0	0	2	-3	2	6													
Gly	-5	-3	0	-1	-3	-2	-3	8												
His	-5	0	1	-1	-3	1	0	-2	10											
Ile	-5	-4	-3	-4	-2	-3	-4	-4	-4	5										
Leu	-5	-3	-4	-4	-2	-2	-3	-4	-3	2	5									
Lys	-5	3	0	-1	-3	2	1	-2	0	-3	-3	6								
Met	-5	-2	-2	-4	-2	0	-2	-3	-1	2	3	-2	7							
Phe	-5	-3	-4	-5	-2	-4	-3	-4	-1	0	1	-4	0	8						
Pro	-5	-3	-2	-1	-4	-1	-1	-2	-2	-3	-4	-1	-3	-4	10					
Ser	-5	-1	1	0	-1	0	-1	0	-1	-3	-3	0	-2	-3	-1	5				
Thr	-5	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	2	5			
Trp	-5	-3	-4	-5	-5	-1	-3	-3	-3	-3	-2	-3	-1	1	-4	-4	-3	15		
Tyr	-5	-1	-2	-3	-3	-1	-2	-3	2	-1	-1	-2	0	4	-3	-2	-2	2	8	
Val	-5	-3	-3	-4	-1	-3	-3	-4	-4	4	1	-3	1	-1	-3	-2	0	-3	-1	5
	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V

BLOSUM50

*BLOck SUBstitution Matrices*

参考：AAindexデータベース

<http://www.genome.jp/dbget/aaindex.html>

# PAM (Accepted Point Mutations)

- ◆ デイホフらによる定式化 (1978年)
  - 71 **グループ**の配列：類似度が 85% 以上
  - 1,572 の置換をカウント
  - 各アミノ酸置換の確率を1,572 の置換の割合で求める
  - 対数尤度 (Log odds) 比で表現
    - ◆  $\log_{10}(q_{ij}/p_i p_j)$  に比例したスコア
    - ◆  $p_i$  はその文字の出現率、 $q_{ij}$  は同時出現率
  - PAM 1
  - PAM 250 は PAM 1 を250 回掛け合わせたもの
    - ◆ 20 % の類似度に相当する

# BLOSUM (*BLO*ck *SU*bstitution Matrices)

- ◆ ヘニコフらによる定式化 (1992年)
  - ギャップなしのローカルアライメントのデータベース (ブロック) から計算
  - すべての置換頻度と非置換頻度 (すべてのアミノ酸ペア  $20 \times 19 / 2 = 210$  パターン) をカウント
  - 対数尤度 (Log odds) 比を使うのはPAMと同じ
    - ◆  $\log_2(q_{ij}/p_i p_j)$  に比例したスコア
    - ◆  $p_i$  はその文字の出現率、 $q_{ij}$  は同時出現率
  - 70% 一致、60% 一致など配列一致度ごとに作成したブロックを用いてそれぞれの置換行列を計算

## ギャップペナルティの種類

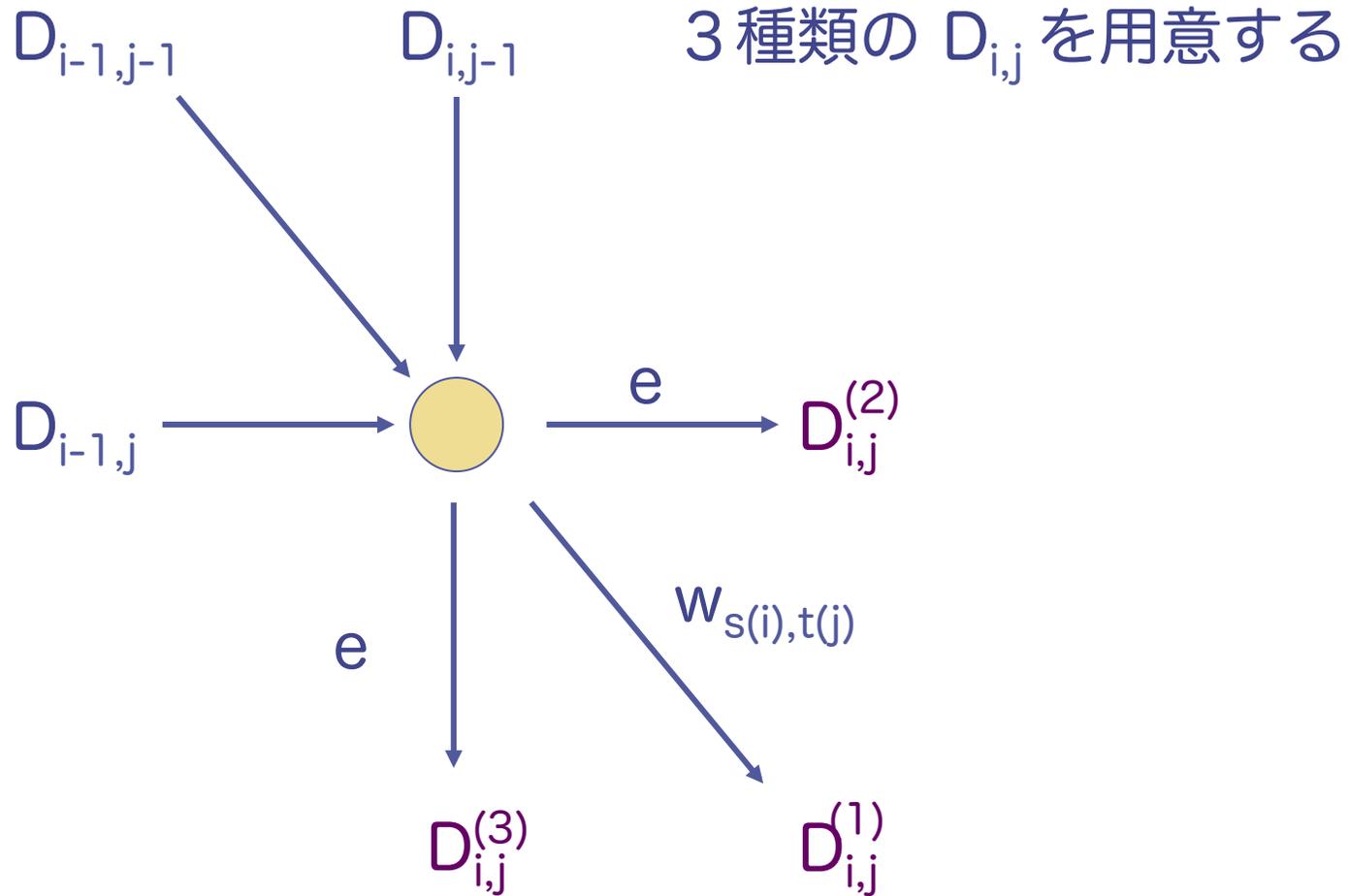
### ◆ 線形スコア

- 長さ  $g$  のギャップペナルティ
- $r(g) = -gd$

### ◆ Affine スコア

- 長さ  $g$  のギャップペナルティ
- $r(g) = -d - (g - 1)e$
- $d$ : gap open penalty
- $e$ : gap extension penalty
- 通常は  $d > e$

# アフィンスコアを使った DP



# 文献

- ◆ Woese, C. R., Kandler, O. and Wheelis, M. L. (1990) Towards a natural system of organisms: Proposal for the domains, Archaea, Bacteria and Eukarya. *Proc. Natl. Sci. USA*, 87, 4576-4579.
- ◆ Gibbs, A. J. and McIntyre, G. A. (1970) The diagram, a method for comparing sequences. Its use with amino acid and nucleotide sequences. *Eur. J. Biochem.*, 16, 1-11.
- ◆ Needleman, S. B. and Wunsch, C. D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, 48, 443-453.
- ◆ Smith, T. F. and Waterman, M. S. (1981) Identification of common molecular subsequences. *J. Mol. Biol.*, 147, 195-197.
- ◆ Gotoh, O. (1982) An improved algorithm for matching biological sequences. *J. Mol. Biol.*, 162, 705-708.
- ◆ Dayhoff, M. O., Schwartz, R. M. and Orcutt, B. C. (1978) A model of evolutionary change in proteins. In *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 3, pp. 345-352.
- ◆ Henikoff, S. and Henikoff, J. G. (1992) Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA*, 89, 10915-10919.

# オミックス解析技術とバイオインフォマティクス

## ◆ ゲノム・メタゲノム

- シーケンサー
- 配列アセンブリ、遺伝子領域予測、機能予測

## ◆ トランスクリプトーム

- DNAマイクロアレイ、DNAチップ、RNA-seq
- 類似発現パターンクラスタリング、転写ネットワーク予測

## ◆ プロテオーム

- 2次元電気泳動、酵母2ハイブリッド、質量分析
- クラスタリング

## ◆ メタボローム

- 代謝産物濃度測定、質量分析
- シミュレーション

遺伝子機能解析  
ネットワーク解析

# オミックスプロジェクト

- トランスクリプトーム (Transcript + -ome)
  - 転写産物 (RNAのこと) の総体
  - 細胞内で実際に RNA として発現している遺伝子
- プロテオーム (Protein + -ome)
  - タンパク質の総体
  - 細胞内で実際にタンパク質として働いている遺伝子
- メタボローム (Metabolite + -ome)
  - 代謝産物 (アミノ酸、グルコースなど) の総体
  - 代謝系で合成されている化合物
- インタラクトーム、フェノーム、グライコーム。。。

# ゲノム解析とバイオインフォマティクス

## ◆ 配列アセンブリ

- DNA 断片をつなぎあわせて一つのゲノムに
- 重なり情報を用いた組み合わせ最適化問題

## ◆ 遺伝子領域予測

- ゲノム中のどこがコーディング領域か？
- 開始コドン～終止コドン
- エクソン～イントロン境界
- 転写・翻訳制御領域

## ◆ 遺伝子機能予測

- 予測された遺伝子の機能は何か？
- 遺伝子個々の機能
- 分子ネットワークとしての機能

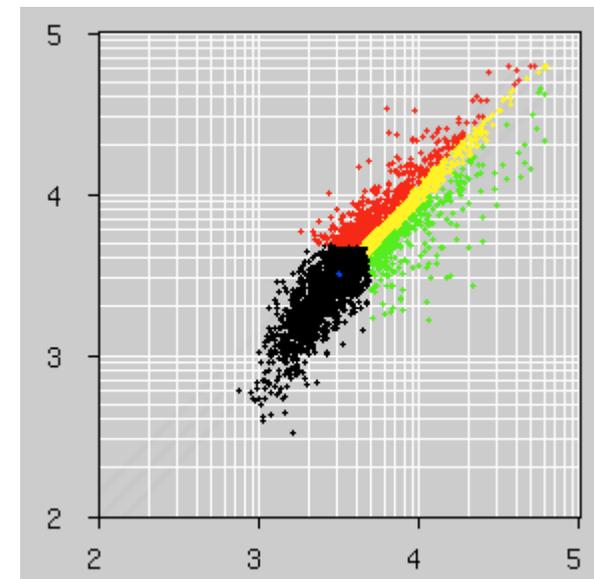
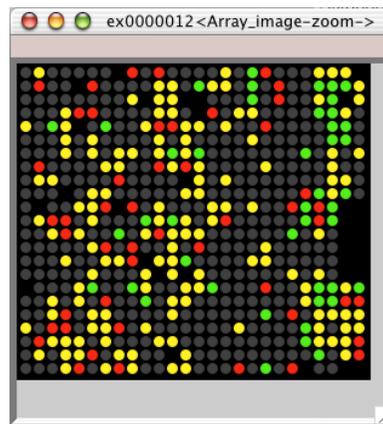
# トランスクリプトーム解析とバイオインフォマティクス

## ◆ 類似発現パターンクラスタリング

- 遺伝子 (mRNA) の発現をゲノムレベルで調べる
- 同時に発現している遺伝子のクラスタリング
- 細胞, 条件特異的に発現している遺伝子の同定
- 正常細胞と変異細胞との比較による病原遺伝子の探索

## ◆ 転写ネットワーク予測

- 時系列での発現データを使って、時間差で発現している遺伝子の情報から予測する





## メタボローム解析とバイオインフォマティクス

- ◆ 代謝産物などの低分子化合物の濃度を測定
- ◆ タンパク質と低分子化合物の結合を調べる
- ◆ Liquid/Gas Chromatography, Mass Spectrometry
- ◆ 遺伝（ゲノム）情報だけでなく化合物の情報も利用



- ◆ 必要なパスウェイの計算
- ◆ 動的なシミュレーションのためのデータ

# バイオインフォマティクス

- ◆ 分子の解析からシステムの解析へ
- ◆ 生命システム解析のためのツール
  - データベース
  - アルゴリズム
- ◆ 生命システム解析
  - 生命の進化解析
  - 医療・創薬への応用