

## 1. 目的

Gclust サーバーは、比較ゲノムにより、系統特異的に保存されたタンパク質を効率よく検索することを目的としている。いわゆるオルソログを検索するサイトはいろいろ存在するが、これらは、2種のゲノムの間での比較によって決めたオルソログペアをもとにして、多種の生物におけるオルソログとしている。しかし、非常によく似たパラログが存在する場合があるほか、2種間のオルソログをたどっていくとつじつまが合わなくなることもあり得る。また、既存のサイトの場合、既知のタンパク質については、よく整理されたオルソログが提供されていることが多いが、機能がわからないタンパク質については、十分な整備がなされていない。Gclust は、はじめに選んだゲノムのセット（データセット）に含まれる推定タンパク質全部について、総当たりの BLASTP を行った結果に基づいて類似タンパク質のグループ（相同グループ）を構築するもので、類似度の高いパラログも含まれる。すべて完全自動で計算するため、既知のタンパク質も未知のタンパク質も同じ基準でクラスター化されている。このため、系統特異的に保存された機能未知タンパク質を検索するのに適している。また、既知のタンパク質のホモログを他の生物で探すのもきわめて容易であり、シロイヌナズナなどよく研究されている植物のホモログを、最近ゲノムが発表されたコケやクラミドモナスで取得し、機能解析をするなどの目的に適している。Gclust の相同グループでは、BLASTP の E 値とタンパク質全体の相同領域率を使ってホモログを仕分けている。タンパク質群ごとに、もともと類似度のきわめて高いものと、比較的保存性の低いものがあるので、各生物種あたりのホモログ数も判断材料とし、あまり多くならないところで切り分けている。また、BLASTP の結果に基づいてドメイン構成を自動判定し、ドメイン構成が著しく異なるタンパク質はマルチドメインタンパク質として単独扱いとしている。ドメイン構成やタンパク質全体の類似度を重視するが、真核と原核に共通に保存されたタンパク質を判定する場合には、N 末の輸送配列の存在を許容している。この

ように、生物学者の立場に立ったタンパク質のクラスタリングを実現している点が、Gclust の最大の特徴であり、自然なタンパク質の分類を自動化した点は、他の情報系のクラスタリングとは異なる点である。

## 2. アクセス

以下の URL からアクセスできる。

<http://gclust.c.u-tokyo.ac.jp/>

現在、アクセス制限はないので、どこか



らでも利用できる。しかし、基本的に、利用はアカデミック目的に限定している。さらに、Gclust のソフトウェアやデータも完全に公開していて、同じサイトからダウンロードできる。ソフトウェアを使って自分でデータベースを作るには、非常に大きなメモリをもった計算機が必要なので、おすすめしないが、バクテリア 20 種程度であれば、8G 程度のメモリを積んだ Mac や Linux で計算可能である。

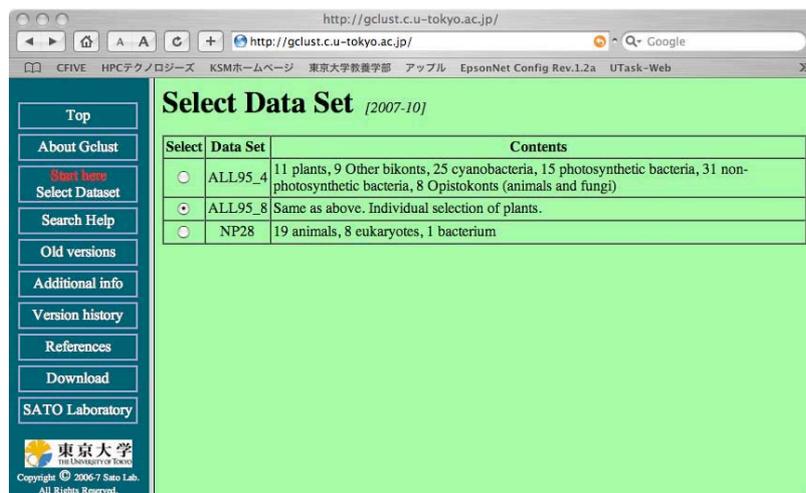
ウェブサイトは、図の通り、シンプルなトップページで始まる。左にボタンが並んでいて、通常は上から 3 番目の **Select Dataset** から始める。

### 3. データセットの選択

まずデータセットを選択する。現在のバージョンでは以下の 2 種類が提供されている。

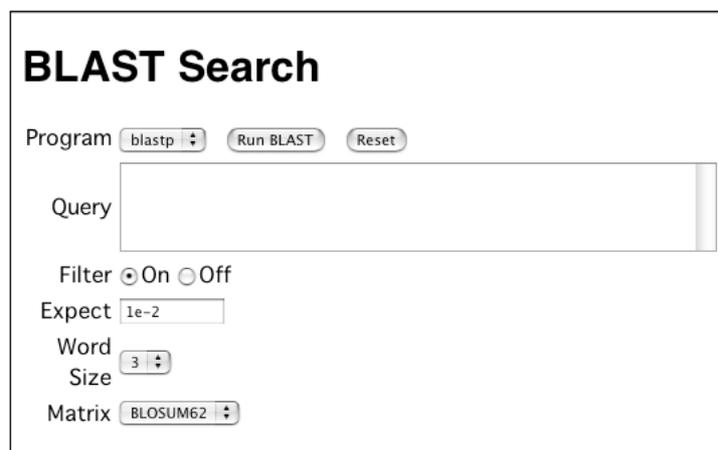
**ALL95**: 植物、シアノバクテリア、動物、菌類、バクテリア、原生動物など全部で 95 種の生物のすべてのゲノムにコードされたタンパク質のクラスタリングによって得られたデータ。動物はヒト、線虫、ショウジョウバエだけに絞っている。

**NP28**: 非植物中心のセット。動物 19 種を中心として、原生動物、菌類に加え、植物、バクテリアなども代表を入れて 28 種にしている。



なお、系統プロファイルを調べる際には生物群のまとめ方が何通りかあると便利なので、**ALL95\_4** と **ALL95\_8** がある。前者では植物がまとまっており、後者では 4 種の陸上植物を別々に選択できる。それ以外のまとめ方が必要な場合、メールなどで連絡してもらえば適宜対応するつもりである。

### 4. 検索の方法



主に、配列や配列名がわかっているとそのホモログを探す場合と、系統を指定してそれらに共通に存在する相同グループを探す場合がある。

#### (1) 配列をもとに探す場合

アノテーションのキーワードを **Annotations** フィールドに書き込んで、左のボタンを押した上で、**submit** ボタンを押す。

配列名は、内部では、ゲノム名と配列名、さらにタンパク質識別番号（高等生物は選択スプライシングによる複数の推定タンパク質があるので、それらはみな別のタンパク質として扱われる。その際、gi 番号を追加して区別している）からできているので、非常に複雑であったり、単なる番号だったりするので、そのまま検索には使えない。しかし、ワールドカード文字としてアスタリスク(\*)が使えるので、わかっている部分の両側を\*で囲んで、Sequence ID フィールドに書き込み、左のボタンを押した上で、上か下の端にある submit ボタンを押す。(例：正式 ID：'Syn\_slr2007=ndhD'の代わりに\*slr2007\* または \*ndhD\*,'ATHc\_ndhB'の代わりに\*ndhB\*,'ATH\_AT1G12340\_145335427'の代わりに\*AT1G12340\*) もしも塩基配列かアミノ酸配列だけがわかっているときには、ページ上部の BLAST というリンクをクリックし、BLAST サーチを行う。これにより、対応するデータベース内の配列がわかり、それをもとに検索が可能になる。

## (2) 系統プロファイルから探す場合

通常は生物群ごとに、yes, no, any を選ぶ。存在することを積極的に指定するのが yes, 積極的に否定するのが no, 選択基準としないのが any である。どんなに系統特異的なタンパク質でも、予想外のバクテリアなどにホモログが存在することがあるので、あまり基準を厳しくしない方がよい。旧バージョンの多くのデータセットや NP28 では、各生物群について、50%を境に yes, no を決めているので、no を選んでも、かなりの数のホモログがあることがある。また、この場合の選択は、配列の数ではなく、配列をもつ生物種の数に

submit

Annotations		
Organism Group	Algae <input type="checkbox"/> any <input type="checkbox"/> Cyanobacteria <input type="checkbox"/> any <input type="checkbox"/> Bacteria <input type="checkbox"/> any <input type="checkbox"/> Eukaryotes <input type="checkbox"/> any <input type="checkbox"/> Moss <input type="checkbox"/> any <input type="checkbox"/> ATH <input type="checkbox"/> any <input type="checkbox"/> OSA <input type="checkbox"/> any <input type="checkbox"/> PoTR <input type="checkbox"/> any <input type="checkbox"/>	
Organisms Unchecked species are regarded as <input type="checkbox"/> any <input type="checkbox"/>	<b>Algae (<i>Algae and bikonts</i>)</b> <input type="checkbox"/> CME ( <i>Cyanidioschyzon merolae</i> ) <input type="checkbox"/> CRE ( <i>Chlamydomonas reinhardtii</i> ) <input type="checkbox"/> DPTM ( <i>Paramecium tetraurelia</i> ) <input type="checkbox"/> GTH ( <i>Guillardia theta</i> ) <input type="checkbox"/> NGR ( <i>Naegleria gruberi</i> ) <input type="checkbox"/> PFA ( <i>Plasmodium falciparum</i> ) <input type="checkbox"/> PHRA ( <i>Phytophthora ranorum</i> ) <input type="checkbox"/> PHSO ( <i>Phytophthora sojae</i> ) <input type="checkbox"/> PTR ( <i>Phaeodactylum tricorutum</i> ) <input type="checkbox"/> TET ( <i>Tetrahymena</i> )	
	<b>Cyanobacteria</b> <input type="checkbox"/> Ana ( <i>Anabaena</i> sp. PCC 7120) <input type="checkbox"/> Ava ( <i>Anabaena variabilis</i> ATCC 29413) <input type="checkbox"/> Glv ( <i>Gloeobacter violaceus</i> ) <input type="checkbox"/> Npun ( <i>Nostoc punctiforme</i> sp. PCC73102) <input type="checkbox"/> Pm1 ( <i>Prochlorococcus marinus</i> MIT9313) <input type="checkbox"/> Pm2 ( <i>Prochlorococcus marinus</i> SS120) <input type="checkbox"/> Pm3 ( <i>Prochlorococcus marinus</i> SS120) <input type="checkbox"/> Pm4 ( <i>Prochlorococcus marinus</i> MIT9312) <input type="checkbox"/> Pm5 ( <i>Prochlorococcus marinus</i> MIT9303) <input type="checkbox"/> Pm7 ( <i>Prochlorococcus marinus</i> MIT9303) <input type="checkbox"/> Pm8 ( <i>Prochlorococcus marinus</i> MIT9315) <input type="checkbox"/> Pm9 ( <i>Prochlorococcus marinus</i> NATLA) <input type="checkbox"/> PmA ( <i>Prochlorococcus marinus</i> MIT9315) <input type="checkbox"/> S79 ( <i>Synechococcus</i> sp. PCC 7942) <input type="checkbox"/> S81 ( <i>Synechococcus</i> sp. WH8102) <input type="checkbox"/> S93 ( <i>Synechococcus</i> sp. CC9311) <input type="checkbox"/> S96 ( <i>Synechococcus</i> sp. CC9605) <input type="checkbox"/> S99 ( <i>Synechococcus</i> sp. CC9605) <input type="checkbox"/> Tel ( <i>Thermosynechococcus elongatus</i> ) <input type="checkbox"/> Ter ( <i>Trichodesmium erythraeum</i> 405 I) <input type="checkbox"/> YelA ( <i>Cyanobacterium Yellowstone A-prime</i> ) <input type="checkbox"/> YelB ( <i>Cyanobacterium Yellowstone B-prime</i> )	
	<b>Bacteria (<i>Bacteria</i>)</b> <input type="checkbox"/> Afu ( <i>Archaeoglobus fulgidus</i> DSM 4304) <input type="checkbox"/> Ape ( <i>Aeropyrum pernix</i> KI) <input type="checkbox"/> Atu ( <i>Agrobacterium tumefaciens</i> str. C58) <input type="checkbox"/> Bja ( <i>Bradyrhizobium japonicum</i> USDA 119) <input type="checkbox"/> Bms ( <i>Brucella suis</i> 1330) <input type="checkbox"/> Bpe ( <i>Bordetella pertussis</i> Tohama I) <input type="checkbox"/> Bsu ( <i>Bacillus subtilis</i> Marburg 168) <input type="checkbox"/> Caur ( <i>Chloroflexus aurantiacus</i> ) <input type="checkbox"/> Cch ( <i>Chlorobium chlorobium</i> DSM 245) <input type="checkbox"/> Clim ( <i>Chlorobium limicola</i> DSM 245) <input type="checkbox"/> Cph ( <i>Chlorobium phaeobacteroides</i> DSM 266) <input type="checkbox"/> Ctep ( <i>Chlorobium tepidum</i> ) <input type="checkbox"/> Cvi ( <i>Chromobacterium violaceum</i> ATCC 35061) <input type="checkbox"/> Eco ( <i>Escherichia coli</i> K-12) <input type="checkbox"/> Fal ( <i>Frankia alni</i> ACN14a) <input type="checkbox"/> Fra ( <i>Frankia</i> sp. CcI3) <input type="checkbox"/> Gox ( <i>Gluconobacter oxydans</i> 621H) <input type="checkbox"/> Hal ( <i>Halobacterium</i> sp. NRC-1) <input type="checkbox"/> Mes ( <i>Mesorhizobium</i> sp. BNC1) <input type="checkbox"/> Mlo ( <i>Mesorhizobium loti</i> MAFF303099) <input type="checkbox"/> Mtu ( <i>Mycobacterium tuberculosis</i> H37Rv) <input type="checkbox"/> Neq ( <i>Nanoarchaeum equitans</i> Kin-4) <input type="checkbox"/> Pst ( <i>Pseudomonas syringae</i> pv. tomato str. DC3000) <input type="checkbox"/> Pvi ( <i>Prosthecochloris vibrioformis</i> DSM 265) <input type="checkbox"/> Rde ( <i>Roseobacter denitrificans</i> Och 114) <input type="checkbox"/> Rhe ( <i>Rhizobium lotii</i> 84) <input type="checkbox"/> Rpa1 ( <i>Rhodospseudomonas plustris</i> BisA53) <input type="checkbox"/> Rpa2 ( <i>Rhodospseudomonas plustris</i> BisB4) <input type="checkbox"/> Rpa3 ( <i>Rhodospseudomonas plustris</i> BisB18) <input type="checkbox"/> Rpa4 ( <i>Rhodospseudomonas plustris</i> BisB18) <input type="checkbox"/> Rrub ( <i>Rhodospirillum rubrum</i> ATCC 11170) <input type="checkbox"/> Rsh ( <i>Rhodobacter sphaeroides</i> ATCC 17029) <input type="checkbox"/> Rso ( <i>Ralstonia solanacearum</i> GMI1000) <input type="checkbox"/> Rsp ( <i>Rhodobacter sphaeroides</i> ATCC 17029) <input type="checkbox"/> Sep ( <i>Staphylococcus epidermidis</i> ATCC 12228) <input type="checkbox"/> Sme ( <i>Sinorhizobium meliloti</i> 1021) <input type="checkbox"/> Sto ( <i>Sulfolobus tokodaii</i> str. 7) <input type="checkbox"/> Vvy ( <i>Vibrio vulnificus</i> YJ016)	
	<b>Eukaryotes (<i>Non-photosynthetic eukaryotes</i>)</b> <input type="checkbox"/> CEL ( <i>Caenorhabditis elegans</i> ) <input type="checkbox"/> DCGR ( <i>Candida glabrata</i> CBS138) <input type="checkbox"/> DKLA ( <i>Kluyveromyces lactis</i> NRRL Y-1140) <input type="checkbox"/> DME ( <i>Drosophila melanogaster</i> ) <input type="checkbox"/> HSE ( <i>Hansenula sermo</i> ) <input type="checkbox"/> SCE ( <i>Saccharomyces cerevisiae</i> ) <input type="checkbox"/> SPO ( <i>Schyzosaccharomyces pombe</i> )	
	<b>Moss (<i>Moss</i>)</b> <input type="checkbox"/> PPT ( <i>Physcomitrella patens</i> )	
	<b>ATH (<i>Arabidopsis</i>)</b> <input type="checkbox"/> ATH ( <i>Arabidopsis thaliana</i> )	
	<b>OSA (<i>Rice</i>)</b> <input type="checkbox"/> OSA ( <i>Oryza sativa</i> )	
	<b>PoTR (<i>Poplar</i>)</b> <input type="checkbox"/> PoTR ( <i>Populus trichocarpa</i> )	
	Sequence ID	Advise for input: Use an asterisk as a wild card, such as *AT1G12345* or *slr1234*, where * is used to represent any length of characters. Gclust internal sequence ID
	Cluster No.	

submit

基づいている。ALL95 では、植物に中心をおいているため、それ以外の生物の選択基準を厳しく押さえていて、no を選んだ場合、1～2種で存在するものだけを許容している。

生物群をもとにして選択する場合、Organism group の左をクリックした上で、項目をセットし、上の submit ボタンをおす。そのほか、個別の生物種を選ぶことも可能である。その場合、Organisms にチェックを入れ、各生物にチェックを入れる。チェックしない生物の扱いは、存在しない(no)か選択基準にしない(any)の2通りを画一的に決めるしかできない。従って、生物群を自分で定義してやり直したい場合には、Gclust のテーブルファイル\*\*\*.tbl と tbsort2 プログラムをダウンロードして、自分で計算することになる。その場合でも、クラスター番号は利用できるもので、一度選択したものをグラフィカルに表示させるときに、クラスター番号を Cluster No.の欄に書き入れればよい。

## 5. クラスターリストの表示

上記のいずれかの方法で検索を行うと、該当するクラスターのリストが表示される。生物群ごとに何種の生物でその配列のホモログが見つかるかが表になっている。一番右には、アノテーションが着いているが、クラスターの1番目の配列にもとのデータベース（多く

Basic Search BLAST Data Set: ALL95\_8 [2007-10]

**Search results** 13576 hits(s) Key:specGrp Val:Bacteria\_no Eukaryotes\_no Moss\_yes ATH\_no OSA\_no PoTR\_no

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 Bottom [Help]

No.	Cluster Number	Sequence ID	Length	seqs	Algae	Cyanobacteria	Bacteria	Eukaryotes	Moss	ATH	OSA	PoTR	Annotations	
1	2	PPT_94467	140	1822	0	0	0	0	1	0	0	0	fgenes1 pg.scaf261000031;	
2	11	PPT_94452	117	764	0	0	0	0	1	0	0	0	fgenes1 pg.scaf261000016;	
3	160	PPT_147688	60	173	1	0	0	0	1	0	0	0	e.gw1.264.50.1;	
4	193	PPT_91141	231	158	0	0	0	0	1	0	0	0	fgenes1 pg.scaf206000052;	
5	227	PPT_91488	118	147	0	0	0	0	1	0	0	0	fgenes1 pg.scaf211000020;	
6	414	PPT_87024	206	110	0	0	0	0	1	0	0	0	fgenes1 pg.scaf153000003;	
7	440	PPT_68854	163	107	0	0	0	0	1	0	0	0	fgenes1 pg.scaf210000062;	
8	485	PPT_81458	333	103	0	0	0	0	1	0	0	0	fgenes1 pg.scaf95000046;	
9	835	PPT_28787	51	82	2	0	0	0	0	1	0	0	0	gw1.225.55.1;
10	940	PPT_94904	142	76	0	0	0	0	0	1	0	0	0	fgenes1 pg.scaf270000002;
11	1740	PPT_86120	134	48	0	0	0	0	1	0	0	0	0	fgenes1 pg.scaf142000042;
12	1810	PPT_93867	159	46	0	0	0	0	1	0	0	0	0	fgenes1 pg.scaf249000053;
13	1895	PPT_159831	580	45	1	25	0	0	1	0	0	0	0	estExt fgenes1 f150267;
14	2182	Aoa_al7528	545	40	3	25	2	0	1	0	0	0	0	hypothetical prot

の場合、GenBank の RefSeq) に書かれていた内容を自動で写したものである。コケや JGI 由来の配列では、単なる記号しか書かれていないものも多く、その他でも適切なことが書かれていないこともあるので、あまり当てにせずに、左の方にあるクラスター番号をクリックして、類似度のマトリクスを表示させる。

左上の表は、コケだけに保存されたクラスターの表である。

**Search results** 1522 hit(s) Key:specGrp Val:Algae\_no Cyanobacteria\_no Bacteria\_no Eukaryotes\_no Moss\_yes ATH\_yes OSA\_yes PoTR\_yes

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 Bottom [Help]

No.	Cluster Number	Sequence ID	Length	seqs	Algae	Cyanobacteria	Bacteria	Eukaryotes	Moss	ATH	OSA	PoTR	Annotations
1	19	OSA_Os11g046700	191	590	1	0	0	0	1	1	1	1	
2	38	ATH_AT5G50820.145359098	193	376	0	0	0	0	1	1	1	1	ANAC097 (Arabidopsis NAC domain containing protein)
3	68	PoTR_200531	284	281	0	0	1	1	1	1	1	1	gw1.IX.1016.1
4	120	ATH_AT1G601390.15223392	489	206	0	0	0	0	1	1	1	1	UDP-glucuronosyl/UDP-glucosyl transferase family p
5	162	ATH_AT3G10080.15232820	227	172	0	0	0	0	1	1	1	1	germin-like protein, putative
6	187	PPT_60724	383	161	0	0	0	0	1	1	1	1	fgenes1 pm.scaffold 281000007;
7	215	ATH_AT1G05650.15220596	394	149	1	0	1	1	1	1	1	1	polygalacturonase, putative / pectinase, putative
8	236	PoTR_424421	116	143	0	0	0	0	1	1	1	1	gw1.87.35.1
9	244	ATH_AT3G14300.15231826	968	141	1	0	3	0	1	1	1	1	pectinesterase family protein
10	249	PoTR_171892	130	139	1	0	0	0	1	1	1	1	gw1.1492.1
11	252	ATH_AT2G32530.30685335	755	138	0	1	0	0	1	1	1	1	ATCSLB03 (Cellulose synthase-like B3); transferase
12	281	ATH_AT2G12100.15226484	86	131	0	0	0	0	1	1	1	1	auxin-responsive protein, putative
13	292	PPT_27856	113	129	0	0	0	0	1	1	1	1	gw1.12.172.1;
14	340	ATH_AT1G4020.42562051	499	118	0	0	0	0	1	1	1	1	unknown protein
15	406	ATH_AT1G02800.15218612	501	111	1	2	0	0	1	1	1	1	ATCEL2 (Arabidopsis thaliana Cellulase 2); hydrolase
16	418	PoTR_235465	229	110	0	0	0	0	1	1	1	1	gw1.XI.3445.1
17	431	ATH_AT1G30440.15220750	665	108	0	0	0	0	1	1	1	1	phototropic-responsive NPH3 family protein
18	593	ATH_AT2G37130.15228090	327	96	0	0	0	0	1	1	1	1	peroxidase 21 (PER21) (P21) (PRXR5)
19	662	PoTR_280312	255	92	0	0	0	0	1	1	1	1	gw1.2426.2.1
20	667	ATH_AT1G51770.145324905	379	92	0	0	0	0	1	1	1	1	unknown protein
21	765	ATH_AT1G47670.15220283	519	86	0	0	0	0	1	1	1	1	amino acid transporter family protein
22	784	ATH_AT1G00480.18390426	431	85	0	0	0	0	1	1	1	1	pectate lyase family protein
23	817	PoTR_786283	187	83	0	0	0	0	1	1	1	1	fgenes4 pg.C scaffold 166000020
24	906	ATH_AT2G46480.15227235	528	78	0	0	0	0	1	1	1	1	GAUT2/LGT2 (Galacturonosyltransferase 2); polygalacturonase
25	925	ATH_AT5G36130.15239309	140	77	0	1	0	1	1	1	1	1	cytochrome P450 family
26	942	OSA_Os04g0165000	276	76	0	0	0	1	1	1	1	1	
27	999	ATH_AT1G64750.15217739	481	73	0	0	0	0	1	1	1	1	glycosyl hydrolase family 17 protein
28	1021	ATH_AT1G29340.15218915	729	72	0	0	0	0	1	1	1	1	PUB17 (PLANT U-BOX17); ubiquitin-protein ligase
29	1083	OSA_Os03g075830080	449	70	0	0	0	0	1	1	1	1	
30	1175	ATH_AT1G68640.15221460	452	67	0	0	0	0	1	1	1	1	PAN (PERIANTHIA); DNA binding / transcription fact

左下の表は植物4種に共通に存在し、他の生物にはほとんど存在しないタンパク質グループを示している。

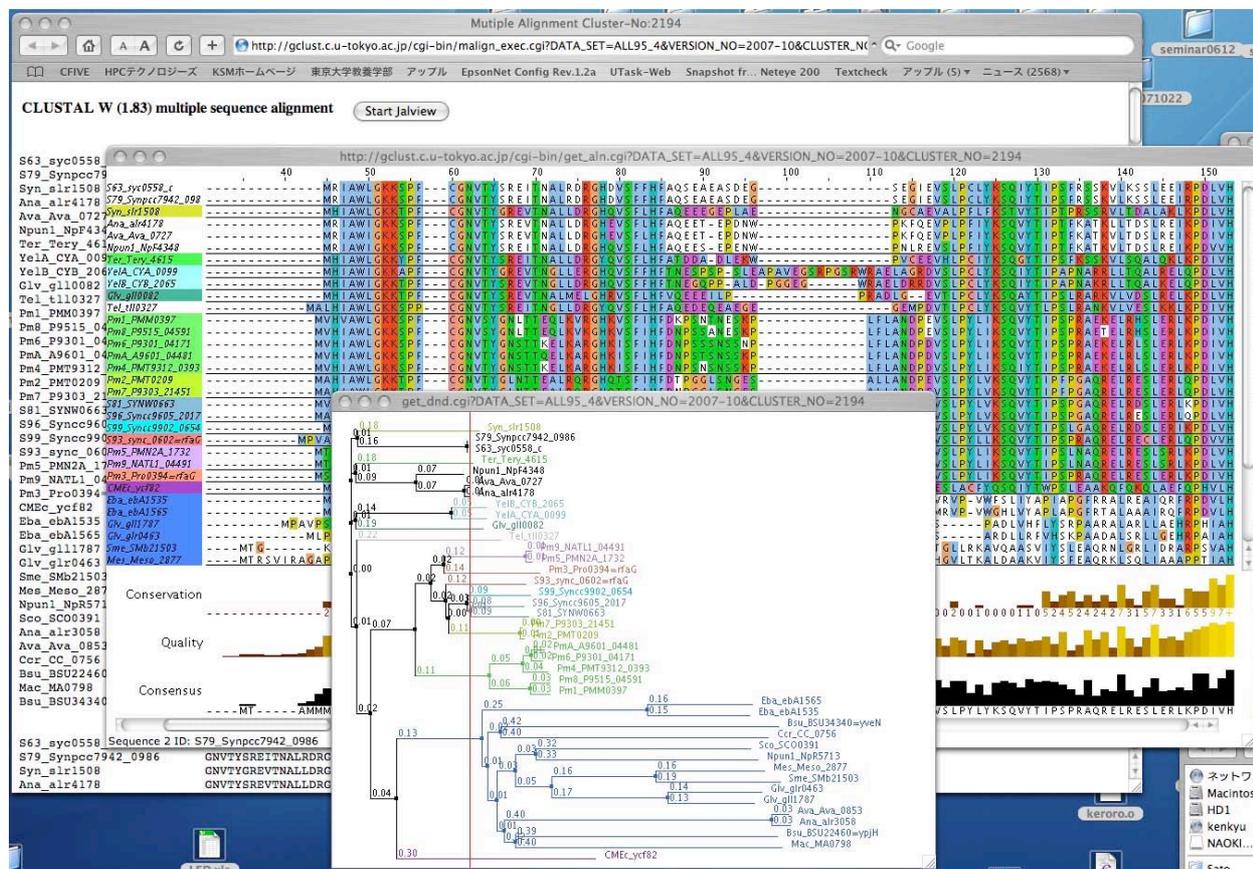
どの表でも、それぞれのクラスターごとに、1つめのメンバーの配列名(ゲノム名とタンパク質 ID)をつないだもので、必要に応じて識別のための gi 番号がついている)、



クラスターを作る過程で適当な E 値をもちながら、クラスターのメンバー全体とは必ずしも似ていない場合に、別のクラスターとなっていることを示している。Related Groups を調べてみると、足りないメンバーがわかることもある。しかし、トランスポーターや糖転移酵素などでは、同じドメインをもつホモログが非常に多く、きわめて雑多なタンパク質が並んでくることがあり、その場合には、Related Groups は考慮しない方がよいかもしれない。

## 7. アラインメントと系統樹

Clustal W のボタンを使うと、アラインメントが表示される。さらに、JalView というソフトが入っているので、このボタンをクリックすると、色つきのアラインメントと系統樹が表示される。ただし、この系統樹は Clustal W が計算するアラインメントのガイドとなる簡便な系統樹であり、本格的に系統解析をする際には適していない。また、Clustal W は配列の数が 40 程度以上の時には Failed という表示が出て、処理してくれない。それらの場合には、Get All Sequences というボタンをつかって全部の配列をダウンロードし、自分でアラインメントを作り、計算を行う必要がある。





バクテリアと紅藻ではそのホモログが存在しなかった。このため、シアノバクテリアと紅藻に共通で他の生物に存在しないタンパク質を検索し、さらに、糖転移酵素と思われるもの一つに絞り込んだ上で変異株を作成したところ、ねらい通り DGDG を欠損した変異体が得られた。

I. Sakurai, N. Mizusawa, H. Wada and N. Sato (2007) Digalactosyldiacylglycerol is required for stabilization of the oxygen-evolving complex in photosystem II. *Plant Physiol.* 145: 1361-1370.

現在、同様の考え方で、これまでタンパク質も遺伝子もわかっていない遺伝子の同定を試みている。

系統プロファイルを利用する方法は、ここに示したようにタンパク質の存在・非存在を使う以外に、もっと多種の生物に存在するタンパク質について系統樹を描いて系統関係を確かめるやり方もある。西山さんたちはこのやり方で進めている。Gclust では後者のやり方は単純にはできないが、信頼度の高い系統樹を自動作成した上で、うまく系統関係を仕分ける方法を開発中である。

**謝辞**：本研究、特に Gclust サーバーの構築に関しては、ゲノム特定領域研究の支援をうけた。また、Gclust データを作成する計算では、東京大学ヒトゲノムセンターのスーパーコンピュータを利用した。

**お願い**：Gclust サーバーに関して、わからないこと、まごついたこと、こうしてほしいと思うことなど、是非お知らせください。少しいじってみただけでなんだかわからない、といって放り出さずに、どこがわかりにくいかわからない、是非教えてください。

**新刊案内**：「光合成の科学」東京大学光合成教育研究会著，東京大学出版会 2007 年  
ISBN 978-4-13-062214-1 3800 円（本体価格）

2008 年 3 月 17 日記 連絡先：naokisat@bio.c.u-tokyo.ac.jp

この文章と図の無断転載を禁じます。Copyright 2008 Naoki Sato