

## 多次元系統プロファイリングによる植物遺伝子機能の大規模推定

●佐藤 直樹

東京大学大学院総合文化研究科

### <研究の目的と進め方>

ポストゲノム研究の重要な課題の一つは、ゲノム上に推定された全タンパク質の約半数にも及ぶ機能未知タンパク質の機能確定である。既知タンパク質や既知ドメインとの類似性による従来からの解析に加え、多数の生物のゲノム情報を比較することによって、これまで個別のゲノムの解析ではわからなかった情報を得ることができる。こうしたゲノムならではの解析として系統プロファイリングがある。これは、類似形質をもつ生物群に共通に存在する遺伝子はその共通形質に関連している可能性があることに基づく方法論である。この問題は一般的なものであるが、最適なターゲットとして、水平移動によって遺伝子が同時に大量に導入されたという場合が考えられる。具体的な例の一つは、根粒菌における窒素固定共生に関わる遺伝子群の大量水平移動である。これは原核生物界における事例であるが、界を越えた現象の例として、シアノバクテリアの細胞内共生による葉緑体の誕生及びそれを持つ真核光合成生物の創成がある。細胞内共生説は一般に確実と考えられているが、どんなシアノバクテリアが共生体となったのかなど、実際には肝心な点がよくわかっておらず、また、既知の光合成反応関連タンパク質遺伝子の他にどれだけの遺伝子が共生体から真核宿主に導入されたのかもわかっていない。

系統プロファイリングの前提として、相同性によるタンパク質の分類が必要である。本研究開始以前にも、米国生物工学情報センター (NCBI) では、Clusters of Orthologous Groups of Proteins (COG) が作られていて、多くのゲノム研究では、これをタンパク質分類の基準としていた。ところが、COGの原理は、単細胞生物 66 種を 14 の系統に分け、そのうちの少なくとも 3 系統以上に共通に保存されたタンパク質に対して COG 番号を付けるというもので、シアノバクテリアにしかない光合成関連タンパク質や、根粒菌などにしかない共生関連タンパク質は、含まれていなかった。それにもかかわらず、ほとんどのゲノム解読論文では、COGの機能分類にあわせたタンパク質分類が慣習的に踏襲されてきた。

COG 自体はすでに古くなり、その代替データベースも作られているが、根本的な問題として、タンパク質のクラスタリングは、研究目的とするゲノムにコードされたタンパク質も加えたクラスタリングを行わなければ、正しい結果にはならないということがある。すなわち、多くの研究者が行っているように、研究対象ゲノムにコードされたタンパク質の配列を query として、COG (または同等品) の配列に検索をかけるというやり方では、正しいクラスタ同定はできない。現在あるいは本研究がスタートした 5 年前でもすでに層であったが、次々と新規ゲノム塩基配列が決める状況では、新規ゲノムの情報も加えていちいち新たなクラスタリングをすべきであるが、適当なソフトウェアが存在しなかった。KEGG (Kyoto Encyclopedia of Genes and Genomes) などでは、双方向ベストヒット (BBH) 法により相同タンパク質をコンパイルしたものが公開されているが、実際にはパラログが多数存在するケースも多く、BBH 法だけではタンパク質を分類すること

はできない。

タンパク質を、その類似度により分類する際の最大の問題点は、タンパク質ファミリーによって、配列類似度がきわめて異なることである。たとえば光合成反応中心を構成する PsaA, PsaB というペアのタンパク質は、80 kDa というサイズにもかかわらず、生物種間での類似度は極めて高く、BLASTP の E 値は  $10^{-150}$  程度である。それに対して、同じ光合成に関連するタンパク質でも酸素発生系の PsaO の場合、E 値は  $10^{-45}$  程度である。さらに光合成に関わるタンパク質の中には、30 残基程度のきわめて小さなものもあり、こうしたタンパク質の種間相同性はきわめて低く、E 値が  $10^{-10}$  程度の場合も多い。従来のバイオインフォマティクスでは、こうした現実の問題を考慮せず、計算の便宜だけでたとえば E 値を  $10^{-6}$  で切るといふようなことが行われてきた。

本研究では、タンパク質の相同性に基づく自動分類を行い、それを元にして、系統プロファイリングによって、系統特異的タンパク質の抽出を行う。さらにこうして得られたものを、実際に実験によって検証することにより、情報的な方法の評価を行う。さらに、従来の系統プロファイリングからワンランクアップした新しい手法として、生物種、配列情報、生物形質情報の 3 者の間の教師なしプロファイリング (多次元系統プロファイリング) を開発し、事前の知識なしに、生物形質と配列情報の連関を浮かび上がらせる手法の開発にチャレンジしたい。全体を通じて、計算と実験を融合した新しい研究手法を確立することを目指した。本研究で提案している方法は、原理的にはどんな生物に関しても適用できる手法であるので、動物、微生物も含めた大きなデータベースを作り上げ、一般に利用できるようにもしたい。

### <研究開始時の研究計画>

#### 1. 計算機による研究

(1) **Gclust ソフトウェアの開発と改良**：20-100 種程度のゲノムにコードされたすべてのタンパク質配列の自動分類を行うため、これらタンパク質のすべての組み合わせによる相同性検索 (BLASTP 解析) の結果を元にして、クラスタリングを行うためのソフトウェア Gclust を開発する。本研究以前にもある程度の開発を行っていて、一定の閾値で単純にクラスタリングを行うことまではできるようになっていた。本研究では、タンパク質グループごとに異なる閾値を自動設定して、生物学的に最も自然な相同グループを形成する方法を開発する。本研究以前に Perl でつくった単純クラスタリングを行うソフトウェア Isort があったが、アルゴリズムを根本的に開発し直し、Gclust と名称を変えて、C 言語で作る。さらに、並列化により、計算速度を速くする。

(2) **系統プロファイリングのためのソフトウェアの開発**：クラスタリングに用いたゲノムを、生物の系統によって分類し、それに従って、相同タンパク質群の保存関係を検索するソフトウェア tbsort (バージョン 2) を開発する。本研究以前に Perl で作った簡易版があったが、それを汎用性のあるものにして C 言語で作り直す。

(3) **Gelust サーバーの公開**：Gelust ソフトウェアを用いて得られた代表的なクラスタリング結果を使って、系統プロファイリングによる検索を行うツールとして、ウェブサーバーを構築して公開する。

(4) **さまざまなデータセットの作成とそれに基づく機能推定解析**：主に光合成生物を中心としてゲノムを集めたデータセットを作るが、動物ゲノムに関しても、別のデータセットを作る。さらに、シアノバクテリアを中心とした別のデータセットや根粒菌中心のデータセットも作り、それぞれのデータセットを活用して系統プロファイリングを行う。

(5) **相同グループを利用した微生物ゲノムの距離構造の解析**：これまでの遺伝子の並び方を利用したゲノム比較を拡張するため、Gelust 相同グループを単位としたタンパク質遺伝子の距離関係に基づく新たなゲノム比較の手法を開発する。オルソログの距離関係のゲノム間での保存について、isoapostatic (等距離) 関係という観点を導入し、これを多次元尺度法により推定することにより、バクテリアゲノムの構造的進化を解析する。

(6) **教師なしの多次元プロファイリングの基本的技術開発**：これまでの系統プロファイリングでは、系統に関する情報をユーザーが与えていたが、配列情報と生物形質情報データを系統情報と組み合わせることで、教師なし多次元プロファイリングができるはずである。このための基本的な技術開発を行う。一つの方法は、マイクロアレイ発現データを利用して、共発現プロファイルを生体種横断的に組み合わせる。しかし、全生物についての発現データがあるわけではないので、系統情報を組み合わせながら可能な範囲で利用する。

(7) **Gelust を活用した進化系統解析**：Gelust を利用すれば、相同タンパク質が簡単に取得でき、すでに遺伝子につけられているアノテーションに頼る必要がないため、多数のゲノムにわたって統一的に相同タンパク質の取得が可能になる。これを利用して、多くの遺伝子の進化系統解析を行う。

## 2. 実験による解析

主として光合成生物を中心として、機能未知タンパク質の新規機能の解析を進める。

(1) **シアノバクテリア起源の葉緑体タンパク質候補**：シアノバクテリアと植物のゲノム比較については、前ゲノム特定において決めた紅藻ゲノムのデータを加えたデータセット CZ16 に基づいて、8種のシアノバクテリアと植物・藻類に共通し、非光合成生物に含まれない73個の相同グループを同定し、これを共生体起源の葉緑体タンパク質 (CPREND0: chloroplast proteins of endosymbiont origin) の候補とする。そのうち光合成や葉緑体形成に関わる既知タンパク質を含む36クラスターを除いた機能未知タンパク質からなる残りの37個の相同グループを実験の対象とした。これらに含まれる56個のシロイヌナズナタンパク質と43個の *Synechocystis* (シアノバクテリア) タンパク質について、機能解析実験を行う。

(2) **植物関連の系統特異的な酵素の同定**：系統特異的に保存されていることがわかっていて遺伝子の同定がされていない酵素について遺伝子を探し、機能を特定する。特に私がもともと専門としていた脂質代謝系の酵素を中心として、これまで欠けていた遺伝子を見つけ出す。

## <研究期間の成果>

### 1. 計算機による研究

(1) **Gelust ソフトウェアの開発と改良**：A. 本研究の開始から開発を始めた Gelust ソフトウェアは、当初、単一の E 値を閾値と

してシングルリンケージクラスタリングを行う形で開発し、E 値をさまざまに変えながら最適なクラスターを作ることを目指した。しかし、この場合、閾値を低くする (条件を厳しくする) と相同性の低いタンパク質ファミリーが分裂してしまい、閾値を高くすると、巨大クラスターができることがわかった。この理由としては、一番よく似たもの以外のタンパク質もクラスターに入ることと、一部分だけが似ているだけでもクラスターに合流することが考えられた。そこで、タンパク質ファミリーごとに最適な閾値を設定するため、タンパク質の相同性を判断する指標として、E 値のほかに、相同領域率 (2つのタンパク質の全長のうちで、互いに相同な部分の長さの割合) も考慮した2次元ヒストグラムを作り、この上で、分布の最初の山を抽出する方法を考案した (0601261302)。その際、分布のエントロピーを用い、また、選択する山にはできるだけ多数の生物種のそれぞれ少数のタンパク質が含まれるようにした。これを Entropy-optimized organism count (EOOC) 法と命名した。しかし、ゲノムの数が多くなると、クラスターが分割してしまう場合や類似クラスターが大きなクラスターを形成する問題が生じた。類似したタンパク質グループを合併するか分離するかの判断を、Jensen-Shannon distance measure を用いて処理することその他により、これらの問題を改善した。B. Gelust に入力する総当たり BLASTP の結果を取り込むファイル形式として、blastall の -m 8 オプションで出力される表を利用することにした。これは blastall の通常の出力形式が、バージョンによりたびたび変更されることに対応したものである。C. その他、細かいパラメータの調整や評価関数の変更を行うことで、実際の大型データの処理が適切になるよう、微調整を行った。こうして、現在の最終版バージョン 3.5.5z を作った (0903102249)。D. Gelust データベースの特徴を詳しく考察するため、NCBI で公開されている COG(Cluster of Orthologous Groups of proteins) との比較を行った。タンパク質ファミリーを与えるという点では、両者とも類似の性質をもつと考えられたため、Gelust と COG がどの程度似ているのか検討した。Gelust クラスターは COG のサブクラスターとなっていることが多いことがわかった。このほかの Gelust の大きな特徴として、真核原核にまたがったデータベースを作ることができることがあげられる。E. さらに計算時間を短くするため、並列化を行った。これは、複数コアをもつ計算機が最近急速に普及していることをうけて、利便性を高めることをねらったものである。もともと Gelust ソフトウェアでは、最初に BLASTP の結果を中間ファイルに書き出すようにして、以後の計算は中間ファイルを読み込むことで、パラメータをさまざまに試すのを効率よくできるようにしていた。この後半部分の約半分を占める EOOC の計算は、タンパク質ごとに別々にできるため、並列化ができ、この部分だけでいえば、ほぼコアの数に比例した処理速度を実現できた。これをバージョン 3.6.1 とした。

(2) **系統プロファイリングのためのソフトウェアの開発**：データセットに含まれる各生物ごとに、それが属する系統を定義し、各系統に含まれるか含まれないかに基づいて、すべての相同クラスターを分類するソフトウェア tbsort2 を開発した。これを用いて得られる代表的なデータを、次に述べる Gelust サーバーより公開した。また、ソフトウェアもサーバー上で公開した。

(3) **Gelust サーバーの公開**：A. 系統プロファイルに基づく宇保存タンパク質群の発見のツールとして、tbsort2 による結果をインタラクティブに検索できるサーバーを構築し、Gelust サーバーとして公開した。その後、文部科学省の統合データベースプロジェクトにもアーカイブとして加えられ、より細かい検索の設定ができる形で公開された。ただし、Gelust サーバーでできる系統解析などは含まれていない。B. Gelust により推定されるタンパク質

クラスターごとに、シロイヌナズナ、イネ、ヒメツリガネゴケの遺伝子上流配列に存在する共通モチーフを検索するサーバを、名古屋大学小保方・山本氏との共同研究により構築した。これは PPDB (Plant promoter DB) として公開されている。C. CyanoClust については、(4) に述べる。

(4) **さまざまなデータセットの作成とそれに基づく機能推定解析**：植物を中心に光合成生物ゲノムを集めた「ALL95 データセット」を構築、公開した。これには、真核光合成生物として、被子植物 3 種 (シロイヌナズナ、ポプラ、イネ)、コケ植物 1 種 (ヒメツリガネゴケ)、緑藻 2 種 (*Chlamydomonas*, *Ostreococcus*)、珪藻 2 種 (*Thalassiosira*, *Phaeodactylum*)、紅藻 1 種 (*Cyanidioschyzon*) が含まれ、シアノバクテリア 25 種、光合成細菌 15 種と合わせ、光合成生物の間でのゲノム比較の性能が飛躍的に向上した。この他、光合成をしない細菌 31 種、光合成をしない Bikont 系統の単細胞真核生物 9 種、Opisthokont 系統の真核生物 (菌類と動物) 9 種が含まれ、光合成をするものとししないものの比較も強化された。さらに Bikont に属する非光合成生物である *Naegleria*、テトラヒメナ、ゾウリムシ、卵菌類などの詳細な系統解析のツールとしても機能することがわかった。さらに、この他に、ヒト、霊長類、マウス、4 種の魚類などの脊椎動物を含む 19 種の動物を中心とした「NP28 データセット」も構築し、公開した。さらに、2009 年秋の時点での最新データに基づく、38 種のシアノバクテリアを中心として多数の色素体ゲノムも含めた「Cyanoclust データセット」を構築し、これについては別のサーバを作って、CyanoClust サーバとして公開した。現在バージョン 3 である。

(5) **相同グループを利用した微生物ゲノムの距離構造の解析**：Gclust ソフトウェアを用いて、2007 年時点で利用できた 25 種のシアノバクテリアからなるデータセット Cyano25 を構築した。このうち *Nostoc punctiforme* については、完全なゲノム配列が出ていなかったため、残る 24 種のデータを利用して、相同なタンパク質のゲノム上での存在位置を比較する新たな手法を開発した。まず、個別のタンパク質遺伝子ではなく、それらが所属する相同クラスタを単位とし、24 種のゲノム上での並び方を、metric multidimensional scaling により正規化する。この際、距離の指標として、次のようなものを用いた。まず、ゲノム上での遺伝子の位置を、遺伝子の個数を単位として円周上にマップし、基準点からの角度で表した。一つのゲノム上での 2 個の遺伝子間の距離は、この角度の差として表した。さらに、この 2 個の遺伝子の相同遺伝子の各ゲノム上での位置を調べ、それぞれ角度の差を求めた。こうして、全部のゲノムにわたる 2 個の遺伝子の距離 (角度の差) の分散を計算し、この値を 2 個の遺伝子の距離と定義した。こうして、すべてのゲノム上で相同遺伝子が定義できる 900 個あまりの遺伝子について、それらの間の距離を定義した。この分散を測度とする multidimensional scaling によって、仮想空間上に仮想ゲノムを描いた。すると、シンテニー群に属する遺伝子はひとまとまりになり、これをクラスタリングすることにより、生物種を超えた大域的シンテニーを示す遺伝子をいくつかのグループに分類することができた。これは種を超えた遺伝子の並び方の保存関係であるので、真核生物のリンケージグループとの類似性ととらえ、virtual linkage group (VLG) と名付けた。VLG は、従来の隣接関係を積み上げていく synteny 概念ではとらえられない、大域的な遺伝子近接関係を発見するのに効果的であった。

(6) **教師なしの多次元プロファイリングの基本的技術開発**：まず多次元化を行うため、さまざまな生物形質情報を検討したが、すべての生物について整えられる情報の種類はきわめて限られているため、単純に多次元化するのは難しいことがわかった。それに代わる手段として、いくつかの代表的な生物についての発現

データを利用することを考えた。光合成生物では、シロイヌナズナのマイクロアレイデータの蓄積があり、共発現データが東北大の大林武氏により ATTED という形でまとめられている。これを利用することを念頭に、ほかの生物でも同様な形のデータを作ることにした。シアノバクテリアのうちで、*Synechocystis* と *Anabaena* については、KEGG に集められた生データが利用できたので、これをそれぞれ ATTED と同様の形式の共発現データにまとめるソフトウェア ArrayStat を開発した。現在までの進行状況はここまでであるが、これら 3 種の生物の共発現データを Gclust を介してつなぐことによって、相同クラスタを機能的関連に基づいてまとめることが可能になると考えている。実際の実装は年度末までかかる見通しである。

こうした方針変更のため、教師なしを実現する方法は新たに考え直す必要があるが、発現情報のセットをもう少し増やすことにより、可能になる見通しである。

(7) **Gclust を活用した進化系統解析**：シアノバクテリアは大きく 2 つのグループに分かれ、系統 1 は *Prochlorococcus* や *Synechococcus* を含み、系統 2 は、*Anabaena* などを含んでいた。色素体は単系統で系統 2 から分岐することがわかった (著書 0612221115)。ゲノムのアノテーションへの利用としては、ヒメツリガネゴケのミトコンドリアゲノム (0612221038) や、ゲノム配列を 100% 完全に解読したシアニジオシゾン (0801251157) の脂質代謝系酵素のアノテーション (0707231610) に活用した。光化学系 2 の酸素発生系を構成する PsbP タンパク質ホモログの複雑な系統関係の推定と分類にも成功した (0912031425)。

## 2. 実験による解析

(1) **シアノバクテリア起源の葉緑体タンパク質候補**：シアノバクテリアと植物・藻類に共通に存在する約 50 個のタンパク質に関して、シアノバクテリアと植物の両方における機能解析により、これらが確かに植物の葉緑体に局在し、光合成に関連する機能を持つ chloroplast proteins of endosymbiont origin (CPRE) であることを証明した (0601261302, 0903102238)。遺伝子破壊株の形質を、蛍光測定 (PAM 法) により詳しく調べた結果、シロイヌナズナと *Synechocystis* の双方で表現型が見られるものを多いものの、実際の蛍光による表現型は異なる場合が多いことがわかった。すなわち、細胞内共生によりシアノバクテリアから葉緑体が成立した際に持ち込まれたタンパク質は、光合成の基本反応や遺伝子発現などに関わるものを除き、機能が厳密には保存されていないと推定される。

(2) **植物関連の系統特異的な酵素の同定**：光合成膜の主要脂質であるジガラクトシルジアシルグリセロール (DGDG) を合成する酵素は、緑色植物では知られていたが、そのホモログが紅藻やシアノバクテリアには存在しないため、紅藻やシアノバクテリアにおける DGDG 生合成酵素は未知のままだった。Gclust を用いて、「紅藻とシアノバクテリアにあって緑色植物など他の生物にはない相同グループ」を選択し、さらに、糖転移酵素モチーフを持つものを探すと 2 個に絞られた。このうちの 1 つは紅藻の葉緑体ゲノムにコードされていたので、これに着目した。*Synechocystis* におけるホモログである *slr1508* を破壊したところ、DGDG を含まない変異体を得られた (0707231610, 0801251334)。このことから、通常条件下では DGDG が必須ではないこと、しかし、光化学系の構築に異常があって光合成活性が低いことなどがわかった。このほかの酵素についてもいろいろ試しているが、まだ明確な同定には至っていない。

<国内外での成果の位置づけ>

当初, Gclust を紹介しても, 「COG とどこが違うのか」などの無理解に基づく批判があり, アメリカの大きな研究機関が作ったものに絶対の信頼を置く一般の研究者からなかなか理解してもらえなかった。このため, 論文の発表もかなり遅くなった。しかし, 論文が出版された 2009 年始め以降, Gclust に基づくデータを使った論文が次々と公表できるようになり, 次第に認知度が高まってきた。また, ウェブサーバーとして, Gclust サーバーのほかに, シアノバクテリアに特化した CyanoClust も公開し, これを用いた新規ゲノムのアノテーションにも応用され始めている。Gclust サーバーへのアクセス数は非常に多いとはいえないが, 検索実行などすべて含めて毎日数百件のアクセスがあるようになったので, この点でも認知度は高まったといえる。

#### <達成できなかったこと、予想外の困難、その理由>

(6) 多次元系統プロファイリングについては, 生物種を越えた発現プロファイルを集約する方法の開発に時間がかかっている, まだ実装するに至っていない。Gclust を用いて得られる新規情報がきわめて多く, それらをまとめて発表することに時間が割かれたことも一因である。Gclust そのものはだいぶ時間が経っているが, これを研究に活かすことができるとことを一般の研究者に理解してもらうのに困難があったため, まず, 実際にこういう使い方ができることをやってみせる必要があったことが大きな理由である。この点については, だいぶ理解が得られるようになったので, 次の段階として「多次元化」に加え「教師なし」に取り組みたい。

#### <今後の課題、展望>

教師なし多次元プロファイリングについての課題は, 上にまとめた。別の課題としては, データ更新がある。ゲノム解読は次々と進められるため, 新たなゲノムを取り込んだデータセットを更新することが必要である。本年度中に陸上植物を追加したデータセットを構築する計画で, 現在作業中である。今後, Gclust を利用して得られる比較ゲノム情報を多数公表していくことにより, 有効性を実証し, さらにソフトウェアも含めた開発を進めたい。

#### <研究期間の全成果公表リスト>

##### 1) 論文/プロシーディング

- 0601261302:** Sato, N., Ishikawa, M., Fujiwara, M., Sonoike, K. (2005) Mass identification of chloroplast proteins of endosymbiont origin by phylogenetic profiling based on organism-optimized homologous protein groups. *Genome Inform*, 16, 56-68.
- 0612221038:** Terasawa, K., et al., The mitochondrial genome of the moss *Physcomitrella patens* sheds new light on mitochondrial evolution in land plants, *Mol. Biol. Evol.*, 24, 699-709 (2007).
- 0707231610:** Sato, N. and Moriyama, T. Genomic and biochemical analysis of lipid biosynthesis in the unicellular rhodophyte *Cyanidioschyzon merolae*: lack of plastidic desaturation pathway results in mixed pathway of galactolipid synthesis. *Eukaryotic Cell* 6, 1006-1017 (2007).
- 0801251157:** Nozaki, H. et al., A 100%-complete sequence reveals unusually simple genomic features in the hot spring red alga *Cyanidioschyzon merolae*. *BMC Biology* 5, 28 (2007).
- 0801251334:** Sakurai, I. et al., Digalactosyldiacylglycerol is required for stabilization of the oxygen-evolving complex in photosystem II. *Plant Physiol.*, 145, 1361-1370 (2007).
- 0903102249:** Sato, N. (2009) Gclust: *trans*-kingdom classification

of proteins using automatic individual threshold setting. *Bioinformatics*, 25, 599-605.

- 0903102238:** Ishikawa, M., Fujiwara, M., Sonoike, K., and Sato, N. (2009) Orthogenomics of photosynthetic organisms: Bioinformatic and experimental analysis of chloroplast proteins of endosymbiotic origin in *Arabidopsis* and their counterparts in *Synechocystis*. *Plant Cell Physiol.*, 50:773-788.
- 0912031425:** Sato, N. (2009) Phylogenomic and structural modeling analyses of the PsbP superfamily reveal multiple small segment additions in the evolution of photosystem II-associated PsbP protein in green plants. *Mol. Phylogenet. Evol.* doi:10.1016/j.ymp.2009.11.021

##### 2) 学会発表 (最近の主なもののみ)

- 佐藤直樹, 紅藻の比較ゲノム・比較生理学から学ぶこと, 日本植物学会, 2006 年 9 月 13-16 日, 熊本
- 佐藤直樹, ゲノムからみた拡張光合成遺伝子, 日本植物学会, 2006 年 9 月 13-16 日, 熊本
- Sato, N., Gclust server: phylogenetic profiling with pre-defined organism sets., GIW, 2006 年 12 月 19-21 日, 横浜
- 佐藤直樹, 比較ゲノムによる植物進化の解析, 日本植物学会, 2007 年 9 月 7-9 日, 野田
- 田島直幸・佐藤直樹, 比較ゲノムデータベース Gclust の紹介, 日本植物生理学会, 2008 年 3 月 20-22 日, 札幌
- Sato, N., Use of comparative genomics in finding unidentified enzymes in lipid metabolism, International Symposium on Plant Lipids, 2008 年 7 月 20-27 日, ボルドー (フランス)
- 佐々木直文・佐藤直樹, 位置プロファイル法に基づく海洋性シアノバクテリアのゲノム構造進化の可視化, 日本ゲノム微生物学会, 2009 年 3 月 5-7 日, 東京
- 佐藤直樹・佐々木直文, 光合成生物の比較ゲノムサーバー Gclust と CyanoClust, 日本光合成研究会, 2009 年 5 月 29-30 日, 東京
- Sato, N., Bioinformatic and functional comparative genomics of endosymbiosis., 2009 年 9 月 20-23 日, ベルリン (ドイツ)
- Sato, N., Lipid-related ortholog database based on Gclust, Asian Symposium on Plant Lipids, 2009 年 11 月 27-29 日, 横浜

##### 3) 図書

- 0612221115:** Sato, N., Springer Verlag, Origin and Evolution of Plastids: Genomic View on the Unification and Diversity of Plastids, The Structure and Function of Plastids, pp. 75-102 (2006), 全 575 ページ.
- 0707231629:** 東京大学光合成教育研究会 (佐藤直樹ほか), 東京大学出版会, 光合成の科学 (2007), 全 291 ページ.
- 0912031511, 0912031513:** 佐藤直樹, 北海道大学低温研究所刊, 光合成研究法, 低温科学 67 巻 (北海道大学低温科学研究所・日本光合成研究会共編 ISBN: 1880-7593) 第 5 章 1 b (PCR を使った *Synechocystis* の高速簡便遺伝子破壊法, pp. 583-586), 第 6 章 1 (比較ゲノムによる光合成関連遺伝子の推定, pp. 657-662) (2009) 全 678 ページ.

##### 4) データベース/ソフトウェア

- 0606210932:** Gclust サーバー: <http://gclust.c.u-tokyo.ac.jp/>  
これは統合 DB にもアーカイブされている。
- 0901152201:** CyanoClust サーバー: <http://cyanoclust.c.u-tokyo.ac.jp/>

0801292027: PPDB サーバー : <http://ppdb.gene.nagoya-u.ac.jp/>

5) 研究成果による産業財産権の出願・取得状況 :

なし

6) 新聞発表、その他顕著なもの :

なし