

## 発生生物学・神経科学のモデル生物 シー・エレガンスの高精度ゲノムを再構築

### 1. 発表者：

森下 真一 (東京大学大学院新領域創成科学研究科 メディカル情報生命専攻 教授)  
Andrew Z. Fire (スタンフォード大学 遺伝学専攻 教授)  
Erich M. Schwarz (コーネル大学 分子生物学・遺伝学専攻 教授)

### 2. 発表のポイント：

- ◆シー・エレガンス (*C. elegans*、代表的な線虫) のゲノムを高精度に再構築することに成功した
- ◆1998年に最初に解読されたゲノムで見落とされていた多数の領域を補うため、20年ぶりに、配布可能な線虫の個体を決め、ゲノムを近代的な長鎖 DNA シーケンシング技術(注1)で解読した
- ◆新たに50個の候補遺伝子と多様な繰り返しパターンが発見され、今後の線虫を使った発生生物学・神経科学の精度を高めるのに寄与する

### 3. 発表概要：

線形動物(線虫)の代表例であるシー・エレガンス(図1)は、成体(雌雄同体)でも細胞数は959個と少なく、受精卵から成虫に至るまで細胞が分裂してゆく様子(細胞系譜)が解明された唯一の動物です。さらに302個の神経細胞のつながりも分かっており、発生生物学および神経科学では長年多用されているモデル生物です。骨や心臓がない簡素な構造をしていますが、ヒトと共通の遺伝子ネットワークを数多く持っており、遺伝子機能・老化・ヒト疾患遺伝子の研究にも使われ、アポトーシス、RNA干渉などの現象が発見されています。

ゲノムは、このような研究の精度を高める基盤となります。シー・エレガンスのゲノムは多細胞生物では最初に1998年に解読されました。ただ20年前の黎明期の技術(注2)で解読されたため、さまざまなゲノム領域が見落とされていることが分かってきました。また解読した個体は数百の研究室に配られ培養された結果、多様化してしまい共通性が乏しくなりつつあります。そこで20年ぶりに、配布可能な代表的な線虫サンプルを再度決めて、そのゲノムを近代的な長鎖 DNA シーケンシング技術で解読することになりました。東京大学森下教授のグループ、米国スタンフォード大学ファイヤー教授のグループ、米国コーネル大学シュワルツ教授らは共同してこの問題に取り組み、従来のゲノムの不具合を補う高精度ゲノムを再構築することに成功しました。

### 4. 発表内容：

本研究の技術的核心は約1億塩基対(注3)のシー・エレガンスのゲノムをどのように解読したか、という点です。記号列を情報処理するアルゴリズムを工夫することで困難を乗り越えてきました。ゲノムを解読する情報分析作業を理解してもらうために、壺の修復作業を紹介します。

オーストリアの古城に大量の古伊万里が発見されたというニュースが、2018年12月3日に流れました。記事をよく見てみると古伊万里の大きな壺がそのままの形で発見されたわけではなく、複数の壺が粉々に砕けてしまった断片が混ざって見つかったようです。復元するには、

粉々の断片を張り合わせてゆく必要があります。まず断片に残る柄などから断片をグループに分け、次に各グループの破片同士を睨んで、形が合うものを見つけては少しずつ繋ぎ合わせてゆく根気のいる作業を繰り返してゆくことになります。

ゲノムを解読する情報分析作業は、この壺の復元作業に似ています。たとえば線虫ゲノムの場合、6本の染色体全体で、約1億塩基対の長大な文字列になります。それを黎明期の技術では500塩基程度のDNA断片へと粉々に砕き、各々の断片の記号列を解読します。互いに共通した部分を持つ記号列どうしをつなげて伸ばして、元の染色体の記号列復元を試みます。共通部分が多いほど解読しやすいので、元のDNAを異なるやり方で粉々にし、数千万個の記号列を生成します。ピースの数が膨大なジグソーパズルになります。

このつなげて伸ばす作業を阻む問題があります。ここで再び、壺を復元する問題の比喻を使いましょう。壺の模様の一部が、同じようなパターンが繰り返す「市松模様」をしていたとしましょう。その市松模様を細かく粉々に断片化してしまうと、復元するのは骨が折れます。特に巨大な市松模様は、復元することを諦めるしか無くなるでしょう。実はゲノムでもこのような市松模様（反復配列と呼ばれます）が多数あることが分かってきました。20年前の解読技術では500塩基ぐらいまで粉々に断片化しないと読めなかったため、ゲノムの市松模様部分を解明することはできませんでした。

ところが近代的なDNA解読技術では2桁長い1万~10万塩基の長鎖DNA断片（注4）も解読可能になりました。例えると、巨大な市松模様部分がすっぽり入る大きな破片を得られるようになりました。問題がとても簡単に解けることが期待されたのですが、実は1つ大きな課題が残りました。それは断片が大きくなるのと引き換えに、模様がたくさん汚れが入ってしまうという良くない性質です。詳しくはDNA断片を解読したとき、読み取りミスが約20%入ります。そのため、この読み取りミスを丁寧に取り除くアルゴリズムが必要になり、本研究で特に苦勞しました。

このような努力の結果、高精度なゲノムを再構築できましたが、さて、20年前に解読したゲノムと、今回の研究で解読したゲノム、どこに大きな違いがあったのでしょうか。約200万塩基対の新しい部分が追加で見つかりました。その中には50個以上の新規遺伝子の候補もみつかりましたが、驚いたことに、殆どは市松模様のように基本単位の記号列が繰り返すパターンだったことです（図2）。これらの市松模様がどのような生物学的な機能を持つのでしょうか。この問題を解明することが今後の課題として残っています。

## 5. 発表雑誌：

雑誌名：「*Genome Research*」（オンライン版：5月24日）

論文タイトル： *Recompleting the *Caenorhabditis elegans* genome*

著者： Jun Yoshimura, Kazuki Ichikawa, Massa J. Shoura, Karen L. Artiles, Idan Gabdank, Lamia Wahba, Cheryl L. Smith, Mark L. Edgley, Ann E. Rougvie, Andrew Z. Fire\*, Shinichi Morishita\*, and Erich M. Schwarz\*

## 6. 用語解説：

（注1）DNAシーケンシング技術：DNAの各塩基を解読する技術

（注2）黎明期の技術：500塩基程度のDNA断片を解読するサンガー法

(注3) 塩基対 (えんきつい) : デオキシリボ核酸の2本のポリヌクレオチド分子が、アデニン (A) とチミン (T)、グアニン (G) とシトシン (C) という決まった組を作り、水素結合で繋がったもの。

(注4) DNA 断片 : DNA シーケンシング技術が解読できる長さに DNA を切断した結果の断片。様々な方法がある。

7. 添付資料：



図1 線形動物（線虫）の代表的生物種 シー・エレガンス（*C. elegans*）の写真。上中央は成虫（雌雄同体、約1mm）、下中央は幼虫、そして11個の卵。

Copyright : Marie-Anne Félix 博士（フランス高等師範学校、École normale supérieure）  
本人から掲載の許諾を得ていますので copyright の掲示をお願いします。

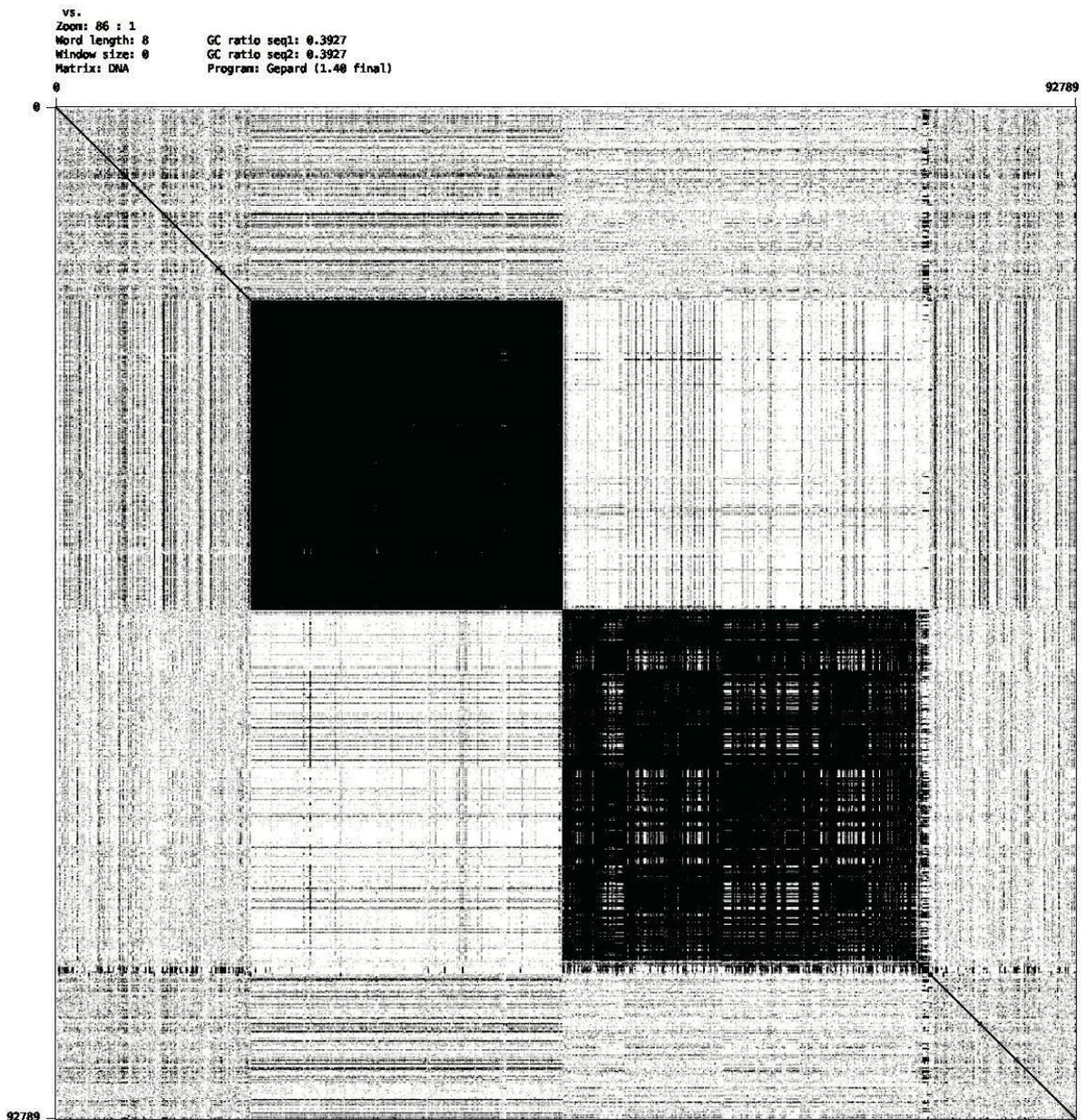


図2 新しく再構築したシー・エレガンスのゲノムで発見された市松模様（同じ単位の配列が千回以上出現するパターン）。

この図は次のようにして生成しています。まず長さ約9万塩基のDNA断片解読配列を複製し、縦軸と横軸に配置します。つぎに横軸座標  $x$  の縦軸座標  $y$  から始まる8塩基が同じときに、2次元座標  $(x, y)$  に点を打ちます。すると対角線には必ず点が打たれるわけです。加えてこの図には、稠密に点が打たれた巨大な正方形が2つ、中央左上と右下に現れています。この巨大な正方形の1辺は各々約3万塩基あり、各々約25塩基の異なる配列が千回以上も繰り返しています。そのため、8塩基の点も繰り返し打たれ、真っ黒な正四角形が出現します。