特別 WEB コラム

DNA シーケンサ

武田　健一
日立製作所　研究開発グループ

1．　まえがき

DNA は生命の設計図と呼ばれることもあるように，DNA に記録された情報に基づいて，タンパク質が合成され，細胞，そして最終的にはその総体としての生命体が形作られる．この DNA に記録された情報，具体的には塩基配列を分析する装置が DNA シーケンサである．以下，DNA の特徴，配列解析の意味，具体的な計測方法について解説する．

2．　DNA とその解析の意義

DNA はデオキシリボ核酸の英語表記（Deoxyribonucleic Acid）の省略形であり，リン酸骨格（主鎖）に対して，4 種類の塩基（アデニン(A)，グアニン(G)，シトシン(C)，チミン(T)）が側鎖として結合した構造を持つ．これが一本鎖 DNA であり，2 本の一本鎖 DNA が水素結合によって相補的に結合することで，教科書などでよく目にする梯子（はしご）状の二重らせん構造を形成する（図 1 (a)，(b)）．ここで相補的という意味は，先ほどの 4 種類の塩基はそれぞれ相手が決まっており，A と T，G と C が水素結合でつながるということで，梯子の踏み桟に相当する部分を構成する．踏み桟（リン酸骨格間）の幅は 2nm 程度，踏み桟（塩基対）同士の間隔は 0.3nm 程度という極めて微細な構造を持つ（図 1 (a)）．この二重らせん構造は，1953 年に J. Watson と F. Crick によって発見され [1]，1962 年にノーベル生理学・医学賞を受賞した．

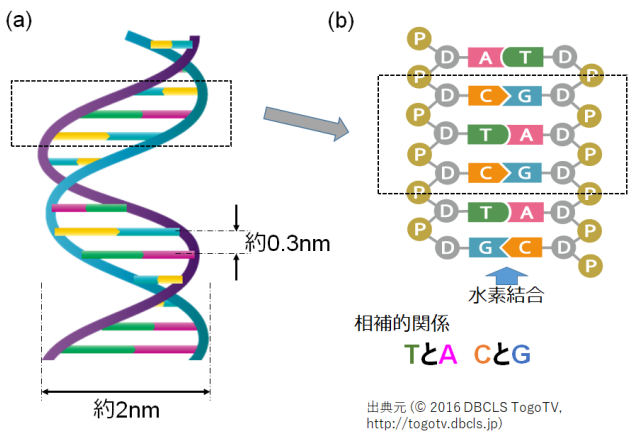


図 1　DNA の構造概略

地球上の多くの生物・ウイルスでは，この DNA に遺伝情報が保存されている．今回の COVID-19 の原因ウイルスである SARS-CoV-2 などはリボ核酸（RNA）に遺伝情報が保存されている．RNA は DNA とよく似た構造をもつが，塩基が結合する糖の一部が異なる（H か OH か）ことと，RNA は二重らせん構造をとらず一本鎖構造をとる点が主な違いである．この DNA や RNA に保存された遺伝情報の総体をゲノムと呼んでいる．このゲノムは DNA(RNA)における塩基配列（AGCT の並び）としてコードされている．

まずは人間のゲノム（ヒトゲノム）の特徴について述べた後，DNA シーケンシング（DNA 配列解析）の歴史を簡単に紹介する．ヒトゲノムは約 30 億塩基対からなり，体細胞（20μm 程度）内の 46 本の染色体（ゲノム 2 セット）に分散して存在している．ヒトゲノムを一本につなげると，0.3nm×30 億×2(2 倍体)=1.8m 程度の長さと計算される．人間を構成する細胞は 60 兆個程度と言われており，ほぼ全ての細胞にこのヒトゲノムが格納されていることになる．これらの細胞も 1 つの受精卵にさかのぼることができるわけであり，細胞分裂のたびに 30 億塩基対のゲノムが正確に複製されていることは生体機構の精密さを物語っている．このゲノム複製機構は，DNA シーケンシングにおいて重要な役割を有しているため，次章にて説明する．数多くの科学者の努力により，DNA 塩基配列にゲノムがコードされていることが明らかとなった後も，ヒトゲノムの全貌は長い間未知のままであった．その理由は DNA シーケンシングが極めて手間がかかる作業だったためである．そのような中，1990 年に期間 15 年間の米国主導ビッグプロジェクトとしてヒトゲノム計画が始動し，DNA の二重らせん構造発見から半世紀の節目の年に，ヒトゲノム解読を完了した．日本もこのプロジェクトに参画し，大きな貢献を果たしている [2]．ヒトゲノムの全体像が示されたことで，ゲノム研究は次の段階へと進み，応用研究が急速に拡大していった．この拡大を支えたのが，さまざまなバイオ技術や DNA シーケンシング技術である．技術発展により，DNA シーケンシング費用が急速に低下し，当初ヒトゲノム解読に 3000 億円かかっていたのが，2019 年には 10 万円程度になっている [3]．

ヒトゲノム計画の完遂，DNA シーケンシング技術の発達は，数多くの科学的，技術的進歩をもたらした．さまざまな生物種のゲノム配列決定に加え，ゲノム配列の差違（変

異）認識も大きく進展した．例えば，ヒトとチンパンジーのゲノム配列の差違は 1%強といわれている[4]．また，ヒトゲノムの個人差は 0.1%程度といわれており，この差違が形質的な違いや病気のかかりやすさと関係することが分かってきている．また，日本では 2019 年にゲノム医療が保険適用され，がん遺伝子の変異（ゲノム配列の変化）情報を活用した治療方針の策定が始まっている．このゲノム配列の変化に着目すると，また異なる応用分野が開かれている．ゲノムには一定割合で変異が発生し，その一部が次世代以降に受け継がれていく．したがって，この遺伝子変異の類似性を解析することで，類縁関係を推定することが可能となる．このゲノム配列の類似性解析は種の遠近を調べる手法としてよく用いられており，ウイルス解析にも適用されている．なお，SARS-CoV-2 など一部ウイルスは RNA をゲノムとして持つため，RNA を相補的 DNA（complementary DNA，cDNA）に転写した後，DNA シーケンスを行って配列解析している．

3．DNA シーケンシングの原理

DNA シーケンシングを理解する上で，ゲノム複製機構および PCR（Polymerase Chain Reaction，ポリメラーゼ連鎖反応）を理解しておくことが重要である．ゲノム複製時には，二重らせん構造の一部がほどけ（水素結合している塩基同士が解離し），DNA ポリメラーゼと呼ばれる酵素が中心的役割を果たして，一本鎖化した DNA の各塩基に対して，相補的な塩基（ヌクレオチド）をつないでいく．より単純化した言い方をすれば，二本鎖 DNA が 2 つに分かれ，それぞれ相補的な塩基が結合することで，2 本の二本鎖 DNA が形成（複製）される．このプロセスは，細胞分裂時に行われているが，これを試験管内で，かつ，サイクル的に行えるようにしたのが PCR 法である（図２）．二本鎖 DNA は高温（90℃以上）にすると一本鎖化する．その後，複製開始点を定義する DNA プライマ（数十塩基長）存在下で温度を下げる（50℃前後）と DNA プライマが鋳型 DNA と結合する．その後，DNA ポリメラーゼ存在下で 70℃程度まで昇温すると，DNA プライマを起点に相補的な遊離ヌクレオチドが結合し DNA が伸長され，二本鎖 DNA が合成される．これを 1 サイクルとし，このサイクルを繰り返すことで，2 の n 乗のペースで DNA が複製されていく．この手法の開発者である K. Mullis は 1993 年にノーベル化学賞を受賞している．

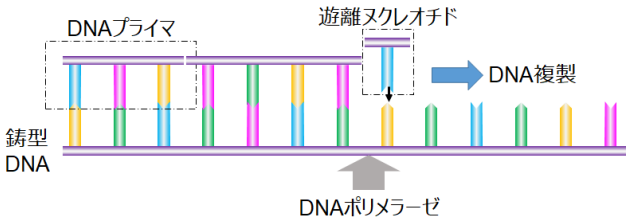


図２　PCR 反応の原理を示す模式図

次に，第 1 世代 DNA シーケンサにおける代表的な計測原理であるサンガー法について説明する．図２で説明したように，DNA ポリメラーゼ存在下では相補的なヌクレオチド（dNTP，デオキシヌクレオチド）を取り込んで DNA の合成が行われる．ところが，このヌクレオチドの構造の一部を変更すると（ddNTP，2′,3′-ジデオキシヌクレオチド），その分子を取り込んだ時点で次の分子をつなぐことができなくなり，DNA 合成が停止する．この dNTP と ddNTP を混合した状態で DNA 合成を行うと，さまざまな位置で停止した DNA 断片を入手することができる．4 種塩基（AGCT）に対応する ddNTP をそれぞれ別の蛍光色素で標識して，電気泳動を行うことで，DNA 配列を決定することができる（図３）．この手法の原理を開発した功績により F. Sanger は 1980 年にノーベル化学賞を受賞している．

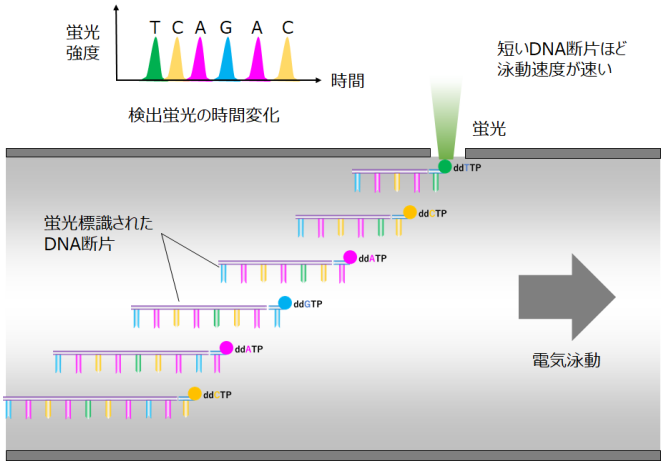


図３　サンガー法を用いた DNA 配列解析の模式図

4．各世代シーケンサの特徴

ここでは代表的な DNA シーケンサの原理と特徴を説明する．DNA シーケンサの分類方法はいくつか考えられるが，ここでは計測時に PCR 増幅を伴うか（単分子 DNA シーケンスが可能か否か）と，蛍光色素での標識が必要か（直接計測が可能か否か）の 2 つの観点で分類する．表 1 には各世代の測定原理，特徴，代表的な機種を記載している．各世代の特徴は，計測原理に直接由来する性能（例えば読取塩基長）と装置実装・運用で変わる性能があるため，あくまで参考として見ていただきたい．また，現在，第四世代のシーケンサが実用化フェーズにあるが，旧世代（例えば第一，第二世代）が単純に置き換わるわけではないことには注意が必要である．現在でも DNA シーケンサの主力は第二世代であり，第一世代もシーケンス精度や使い勝手が評価されて依然として根強い需要があり，全世代が共存しているのが実態である．それでは以下，各世代の特徴を簡単に述べる．

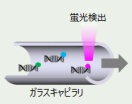
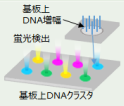
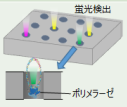
世代	第一世代	第二世代	第三世代	第四世代
PCR増幅	必要	必要	不要（単分子計測）	不要（単分子計測）
検出手法	蛍光	蛍光	蛍光	蛍光色素不要 （イオン電流など）
代表手法	サンガー法＋キャピラリ電気泳動	ブリッジPCR法＋シーケンスBySynthesis	微小ウェル内DNA伸長反応	バイオポア，ソリッドナノポア，ナノギャップなど
模式図				
特徴	・読取誤差が小さい ・読取塩基長が比較的長い（～1000塩基）	・超並列処理による高スループット ・読取塩基長が短い（～200塩基）	・塩基毎の識別が可能 ・読取塩基長が長い（平均10k塩基） ・読取誤差が大きい	・塩基毎の識別可能 ・読取塩基長が長い（平均10k塩基） ・読取誤差が大きい
代表機種	SeqStudio (Thermo Fisher Scientific)	NovaSeq 6000 (Illumina)	Sequel II (Pacific Biosciences)	PromethION (Oxford Nanopore Technology)

表１　各世代 DNA シーケンサの特徴

4.1　第一世代

前章で説明したサンガー法を用いており，当初はガラス板で挟んで電気泳動を行っていた．その後，ガラス毛細管（ガラスキャピラリ）を用いることで，電気泳動の高速化と並列計測（～100）が可能となった[5]．この方式が開発されたことでシーケンス速度（データ出力数）は大幅に向上し，ヒトゲノム計画の前倒し完了に貢献したと言われている．Thermo Fisher Scientific 社の最新機種である SeqStudio Genetic Analyzer（2017 年国内販売）では，4 本キャピラリ，最上位機種である 3730 シリーズでは最大 96 本キャピラリでの計測が可能となっている．シーケンス解析可能な DNA 長さ（塩基長）は 900 塩基程度であり，これより長い場合には DNA を分割して計測することになる．装置がコンパクトで，データ信頼性が高く，後段のデータ処理も含めてハンドリングが容易であることから，根強い需要がある．

4.2　第二世代

登場した当時，次世代シーケンサ（Next-generation Sequencer：NGS）と呼ばれたことから NGS と呼ばれることも多い．現在のトップメーカーは米国 Illumina 社であり，デスクトップサイズから超大型装置まで取りそろえている．基本的なシーケンス原理は共通で，平面基板（フローセル）上に大量に DNA クラスター（増幅された DNA 集団）を形成し，各クラスターで DNA 合成反応を行い，塩基種を蛍光色で判定する手法である[6]．DNA クラスター数を増やせばシーケンス速度（データ出力数）は増えるため，大規模化に向いている．Illumina 社は自社プロダクトである HiSeq X シリーズのリリースに合わせ，シーケンスコストにおける大きなマイルストンであった 1000 ドルゲノム（ヒトゲノム解読を 1000 ドル以下）を達成したと報告している[7]．この方式は１回で読み取り可能な塩基長が短い（100～200 塩基程度）という特徴がある．これは基板上で DNA 標識反応を行うことに起因している．第 2 章で述べたように，ヒトゲノムは約 30 億塩基対から構成されていることから，その読取長との差違は 7 桁程度ある．ゲノム全体を短い塩基長で解析する方法として，少しずつ読み取り位置を変えてシーケンスを行い，最後に１つにまとめるというやり方がある．ランダムに発生した断片をつなぎ合わせて正しい大規模構造を再構成するために，優れたアルゴリズムの開発と大規模計算機パワーの活用が行われている．しかしこの方法でも十分には再構成できないといわれている．例えば，読取塩基長より長い周期での繰り返し構造などがそれにあたる．そのような背景で生まれたのが，読取塩基長の長い（ロングリード型）第三，第四世代である．

4.3　第三世代

ロングリードを実現するためには，読み取りを進めても精度を大きく低下させない工夫が必要となる．第三世代に相当する米国 Pacific Biosciences 社の DNA シーケンサ Sequel II では，基板に形成されたマイクロウェル内にポリメラーゼを固定し，このウェル底部で生じる DNA 合成に伴う蛍光発光を検出する方法を採用している[8]．このマイクロウェルの数が並列度を決めることになる．１回当たりの読み取り精度はそれほど高くないため，複数回読み取りを行って精度を高めている．平均 10k（10000）塩基長の読み取りが可能ではあるが，第二世代よりシーケンスコストが高いため，第二世代と組み合わせて用いられる場合も多い．

4.4　第四世代

蛍光反応を使わずに，DNA 分子を物理的・化学的手法で直接読み取ることから，最終世代と呼ばれることもある．図１に示したようなナノスケールの化学構造を直接読み取るという極めてチャレンジングな課題であり，多くの研究機関，企業が開発に取り組んだ．現在，実用化レベルに到達しているのは英国 Oxford Nanopore Technologies（ONT）社である．計測原理を簡単に説明する．電解質がメンブレンで分離されており，このメンブレンにナノサイズの孔（ナノポア）を持つタンパク質（バイオポア）が形成されている．このナノポアを一本鎖化された DNA が通過する際に，このナノポアを通過するイオン電流が変調を受け，その値は塩基種に依存することが知られている．つまり，ナノポアに DNA を導入し，通過するイオン電流を計測すれば，通過している塩基種が分かり，DNA 配列解析が可能になる．ONT 社のポータブル型シーケンサである MinION では 512 個のバイオポアが形成可能であり，これが並列度に相当する．ONT 社では MinION のフローセルを最大 48 個並列化した PromethION もリリースしている．ナノポア方式には原理的には読取長の上限は存在しないが，ONT 社の方式ではサンプル調製の関係で平均 10k 塩基長となっている．第三世代の Sequel II と異なり繰り返し読み取りが困難なため，読取精度に課題が残っているものの，使い勝手の良さから第二世代との組み合わせで新規配列の決定などに使われるケースも増えてきている．

ONT 社以外にも第四世代の開発は継続中である．最先端研究開発支援プログラム（FIRST）の成果を活用した DNA

シーケンサとして Quantum Biosystems（QB）社が開発を継続している[9]．ナノギャップ電極間に DNA を通過させ，トンネル電流が塩基種に依存する特徴を用いて DNA 配列解析する手法である．ONT 社ではタンパク質のナノポアを用いているが，無機材料を用いたソリッドナノポア型の開発も精力的に続けられている[10]．

５． むすび

ヒトゲノム解読完了後，ポストゲノム時代としてゲノム情報に基づくさまざまな研究，応用が花開いた．DNA シーケンサを含む理化学機器は縁の下の力もちとして，この発展を支えてきた．しかし，いまだ生命・生物には未知の領域が多数残っている．新しい分析手法，計測手法は未知の世界につながる新たな扉である．応用物理に携わる研究者，学生の方々の力が今，必要とされている．

謝 辞

原稿執筆の機会を頂いた応用物理学会 Web コラム編集委員会の皆様に感謝申し上げます．また，原稿執筆に関して有意義な意見を頂いた（株）日立製作所研究開発グループの植松千宗氏，横井崇秀氏に感謝いたします．

文 献

1) J. D. Watson and F. H. C. Crick, “A Structure for. Deoxyribose Nucleic Acid.”, Nature Vol. 171, No. 4356（1953）, pp. 737-738.
2) (国研)理化学研究所　ニュースリリース(2001 年 2 月 12 日)：https://www.riken.jp/medialibrary/riken/pr/press/2001/20010212_1/20010212_1.pdf
3) National Human Genome Research Institute ホームページ，https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data
4) (国研)理化学研究所　ニュースリリース(2002 年 1 月 4 日)：https://www.riken.jp/medialibrary/riken/pr/press/2002/20020104_1/20020104_1.pdf
5) H. Kambara and S. Takahashi, “Multiple-sheathflow capillary array DNA analyzer”, Nature Vol. 361 (1993), pp. 565–566.
6) イルミナ社ホームページ，https://jp.illumina.com/content/dam/illumina-marketing/apac/japan/documents/pdf/2013_illumina_techsupport_session19.pdf
7) イルミナ社ホームページ，https://jp.illumina.com/content/dam/illumina-marketing/apac/japan/documents/pdf/specsheet_hiseq_x_series_system_j.pdf
8) トミーデジタルバイオロジー社ホームページ，https://www.digital-biology.co.jp/allianced/products/pacbio/
9) Quantum Biosystems ホームページ，https://quantumbiosystems.com/
10) 例えば，Y. Goto et al., “Solid-state nanopores towards single-molecule DNA sequencing”, J. human genetics Vol.65 (2019), pp.69-77.

脱稿日　2020 年 6 月 16 日

武田　健一（たけだ　けんいち）

1992 年東北大学工学院研究科博士前期課程修了．博士（工学）．株式会社日立製作所中央研究所にてシリコン半導体の研究開発に従事，その後，バイオ・ヘルスケア分野の研究開発に従事．応用物理学会会員，IEEE 会員．