

# IMES DISCUSSION PAPER SERIES

情報爆発時代の切り札へ：  
DNAストレージに関する研究動向とセキュリティ分析

いのうえ しおり  
井上紫織

Discussion Paper No. 2020-J-11

# IMES

INSTITUTE FOR MONETARY AND ECONOMIC STUDIES

BANK OF JAPAN

日本銀行金融研究所

〒103-8660 東京都中央区日本橋本石町 2-1-1

日本銀行金融研究所が刊行している論文等はホームページからダウンロードできます。

<https://www.imes.boj.or.jp>

無断での転載・複製はご遠慮下さい。

備考：日本銀行金融研究所ディスカッション・ペーパー・シリーズは、金融研究所スタッフおよび外部研究者による研究成果をとりまとめたもので、学界、研究機関等、関連する方々から幅広くコメントを頂戴することを意図している。ただし、ディスカッション・ペーパーの内容や意見は、執筆者個人に属し、日本銀行あるいは金融研究所の公式見解を示すものではない。

## 情報爆発時代の切り札へ： DNAストレージに関する研究動向とセキュリティ分析

いのうえ しおり  
井上紫織\*

### 要 旨

近年、世界は情報爆発の時代を迎えており、データ・ストレージへの需要が急増している。現在、データ・ストレージには主にハードディスク・ドライブ（HDD）が用いられているが、その記録密度の伸び率は鈍化傾向にあり、新たなストレージ技術への期待が高まっている。こうした状況を受け、近年、高密度かつ長期間のデータ保存が可能とみられているDNA (deoxyribonucleic acid) を情報記録媒体として用いるDNAストレージの研究が活発化している。最近では、自社のデータ・センターでDNAストレージを実装することを目指して研究を進めているとみられる企業もある。本稿では、DNAストレージの要素技術を解説するとともに、DNAストレージの基本的な概念および実用化に向けた研究動向を紹介する。さらに、DNAストレージを実装するシステムの形態を1つ想定し、セキュリティ面に焦点を当てて分析し課題を考察する。

キーワード: DNAストレージ、情報爆発、情報記録媒体、セキュリティ、  
データ・ストレージ

JEL classification: L65、L86、L96、Z00

\* 日本銀行金融研究所主査 (E-mail: shiori.inoue@boj.or.jp)

本稿の内容は、執筆時点（2020年3月31日）で公表されている情報を基に作成されている。本稿の作成に当たっては、東京大学の萩谷昌己教授および早稲田大学の清水佳奈教授から有益なコメントを頂いた。ここに記して感謝したい。ただし、本稿に示されている意見は、筆者個人に属し、日本銀行の公式見解を示すものではない。また、ありうべき誤りはすべて筆者個人に属する。

## 目 次

1. はじめに.....	1
2. 電子記録媒体と比較した DNA の情報記録媒体としての利点.....	2
(1) 記録密度.....	2
(2) データ保存期間.....	3
3. DNA を用いたデータの書込みと読出し.....	3
(1) 処理の流れ.....	3
(2) 各処理の概要と研究動向.....	4
イ. 変換規則に基づく変換.....	4
ロ. DNA 合成.....	6
ハ. DNA の保存と取出し.....	6
ニ. DNA の増幅.....	7
ホ. シーケンス解析.....	7
(3) DNA ストレージの実用化に向けた取組み.....	8
4. DNA ストレージのセキュリティに関する分析と対策.....	9
(1) セキュリティ分析の方向性と前提.....	9
イ. 分析対象とするシステムとその構成.....	9
ロ. セキュリティ対策の前提と攻撃箇所.....	11
ハ. セキュリティ目標.....	12
ニ. 攻撃者の目的と能力.....	12
(2) 想定される攻撃とその対策.....	12
イ. データの盗取.....	12
ロ. データの改ざん.....	16
ハ. データの読出しの妨害.....	18
(3) セキュリティ分析の総括.....	20
5. 結びに代えて：DNA ストレージへの期待と課題.....	22
補論 1. DNA の役割と基本機能.....	28
補論 2. DNA の増幅.....	29
補論 3. ナノポア・シーケンサーを用いた効率的な読出し方法に関する研究.....	30
補論 4. ランダム・アクセス機能.....	30
補論 5. 改ざんプライマーを用いた DNA の書換え手順.....	32

## 1. はじめに

近年、通信機能を備えた IoT デバイスの増加やビッグデータの利活用の進展等を受け、データ・ストレージへの需要が急速に増加している。全世界の年間データ生成量は、2017 年には 23 ゼタバイト（23 兆ギガバイト）であったが、2025 年には 175 ゼタバイトと年率 30% 程度のペースで増加すると予測されており、こうした状況は「情報爆発時代」とも表現される（喜連川 [2011]、Reinsel, Gantz, and Rydning [2018]）<sup>1</sup>。現在、データ・ストレージには主にハードディスク・ドライブ（HDD）が用いられているが、その記録密度の伸び率は鈍化傾向にあり、新たなストレージ技術への期待が高まっている。

こうした中、高密度かつ長期間にわたりデータを保存しうる技術の 1 つとして、近年、DNA（deoxyribonucleic acid）を情報記録媒体として用いる DNA ストレージに関する研究が活発化している。DNA は、生物の生体形成や生命活動の維持に必要となる、たんぱく質の構造に関する情報を記録した化学物質である。情報は、DNA の構成要素である塩基（アデニン〈A〉、チミン〈T〉、グアニン〈G〉、シトシン〈C〉の 4 種類）の配列により表現される<sup>2</sup>。1950 年代以降、DNA を自由にデザインし操作する技術が活発に研究され、塩基配列を予め決定したうえで、その配列を有する DNA を人工的に合成することもできるようになった。DNA ストレージは、こうした技術の応用事例の 1 つであり、デジタル・データを塩基配列に変換したうえで、当該塩基配列を持つ DNA を人工的に合成することによりデータを記録する<sup>3</sup>。データを読み出す際には、DNA の塩基配列を解析し、その配列をデジタル・データに再変換する。

DNA ストレージに関する最近の研究動向をみると、2012 年に、世界で初めて、ハーバード大学の研究者らにより DNA ストレージにかかる概念実証の事例が報告された（Church, Gao, and Kosuri [2012]）。その後、ワシントン大学とマイクロソフトの研究チームや欧州バイオインフォマティクス研究所、コロンビア大学とニューヨーク・ゲノム・センターの研究チーム等、さまざまな研究チームにより研究が進められ、記録密度の向上、エラー耐性やランダム・アクセス機能の改善、データの読書の自動化等が検討されてきた（例えば、Erlich and Zielinski [2017]）<sup>4</sup>。マイクロソフトでは、自社のデータ・センター内に DNA ストレージを実装することを目指して研究を進めているとみられ、今後実用化に向けた研

---

<sup>1</sup> 本稿におけるデータ等の単位は、メガ（100 万）→ギガ（10 億）→テラ（1 兆）→ペタ（1,000 兆）→エクサ（100 京）→ゼタ（10 垓）の順で 1,000 倍ずつ大きくなる。

<sup>2</sup> DNA の詳細については補論 1 を参照されたい。

<sup>3</sup> データを記録した DNA を大腸菌等の生物の DNA に組み込んで保存する研究事例も報告されている（Wong, Wong, and Foote [2003]）。本手法では、その生物を増殖することで簡単にデータをコピーすることができる一方で、データの書込みや読出しにはより多くの手間を要する。

<sup>4</sup> ここでのランダム・アクセス機能とは、大量の DNA から、目的のデータを記録した DNA のみを抽出することによって、データを効率的に読み出す機能を指す。

究はますます活発になると予想される (Regalado [2017])<sup>5</sup>。

DNA ストレージが実現すれば、高密度かつ長期間のデータ保存が可能になると期待されている一方で、読書きについては、DNA の合成や増幅といった操作が必要となり、処理の速度が HDD 等と比べると遅くなると予想される。こうしたことから、長期間にわたってデータを保存することを主な目的とし、書込みや読出しを高速で実行する必要がなく、いったん書き込んだデータを頻繁に書き替えることがない用途、例えば分散記録台帳の保存やアーカイブにおける活用が想定される。

DNA ストレージを実装したシステムを実用化するうえで、記録したデータが長期間安全に保護されるよう、各種リスクと対策を予め検討しておくことが求められる。地震等の災害への物理的な安全対策を講じておくことはもとより、予期せぬデータ損失への対策、データの機密性や重要性に応じたセキュリティ対策等を検討しておく必要がある。

本稿では、DNA ストレージの基本的な概念と、その実用化に向けた研究動向を紹介する。さらに、DNA ストレージを実装するシステムの形態を1つ想定したうえで、記録したデータを保護するために検討すべき課題のうち、特にセキュリティ面に焦点を当てて分析し、セキュリティ上の課題を考察する。

## 2. 電子記録媒体と比較した DNA の情報記録媒体としての利点

DNA ストレージが実現すれば、主に、記録密度とデータ保存期間に関してメリットがあると考えられる。

### (1) 記録密度

HDD の記録密度は、2000 年から 2010 年にかけて、年率 100%を超えるペースで高まってきたが、従来の垂直磁気記録方式を用いた場合の記録密度は物理的な上限値に達しつつある (佐野・石本 [2013])<sup>6</sup>。こうした中、HDD の高密度記録方式として、近年、熱アシスト磁気記録方式やマイクロ波アシスト磁気記録方式が活発に研究されており、数年以内に 20~40 テラバイト (2 万~4 万ギガバイト) のデータ容量を有する 3.5 インチ HDD が製品化される見込みとなっている (堀内 [2019])<sup>7</sup>。また、長期保存性、消費電力、単位データ量あたりの価

---

<sup>5</sup> Regalado [2017]は、こうした研究が同文献公表時点 (2017 年) から 10 年後を目途に (toward the end of this decade) 進められている旨を報じている。

<sup>6</sup> HDD 上のデータは、1 ビットのデータを格納する単位領域ごとに、それを構成する磁性粒子の磁気の向きによって記録される。垂直磁気記録方式とは、磁性粒子の磁気の向きがディスク面に対して垂直になるように制御する方式である。

<sup>7</sup> HDD の記録密度を向上させるためには、単位領域を微細化すればよい。しかし、領域が細くなるほど、温度変化や隣の領域への書込みの影響を受けて磁気の向きが変わりやすくなり、データが欠損するリスクが高まる。そこで、熱アシスト磁気記録方式では、書込みを行う領域の

格、マルウェア攻撃によるデータ損失のリスク等の観点から、アーカイブ等の用途で活用されている磁気テープについては、収容可能なデータ量は最大で数百テラバイト（数十万ギガバイト）である<sup>8</sup>。

一方、DNA は電子記録媒体と比較して極めて高密度にデータを記録しうる。理論的には、DNA 1 グラムあたり最大で 680 ペタバイト（6 億 8 千万ギガバイト）、1 立方ミリメートルあたり最大で数エクサバイト（数十億ギガバイト）のデータを収容できるとされている（Erlich and Zielinski [2017]、Newman *et al.* [2019]）。これは、HDD の約百万倍の記録密度に相当する。

## （2）データ保存期間

電子記録媒体は、その保存環境を工夫しても経年劣化を避けることは難しく、記録されたデータはいずれ正しく読み出せなくなることが知られている。安全にデータを保存することが可能な期間は、HDD では 5 年程度、フラッシュメモリや磁気テープでは 10 年程度、光ディスクでは 30 年程度、MO では 50 年程度とみられている<sup>9</sup>。

一方、DNA は適切な低温環境下であれば、半永久的にデータを劣化させずに保存することができるとみられている。例えば、スペインの洞窟で発掘された約 40 万年前の人骨から採取された DNA について、その塩基配列の読出しに成功した事例が知られている（Meyer *et al.* [2014]）。

## 3. DNA を用いたデータの書込みと読出し

### （1）処理の流れ

DNA ストレージにおけるデータの書込みプロセスと読出しプロセスの処理の流れのイメージは以下のとおりである。それぞれのプロセスに付した番号は、図表 1 中の番号に対応する。

#### 【書込みプロセスの処理の流れ】

- ① 記録したいデータ（バイナリ・データ）を、一定の規則（変換規則）に従って塩基配列に変換する。

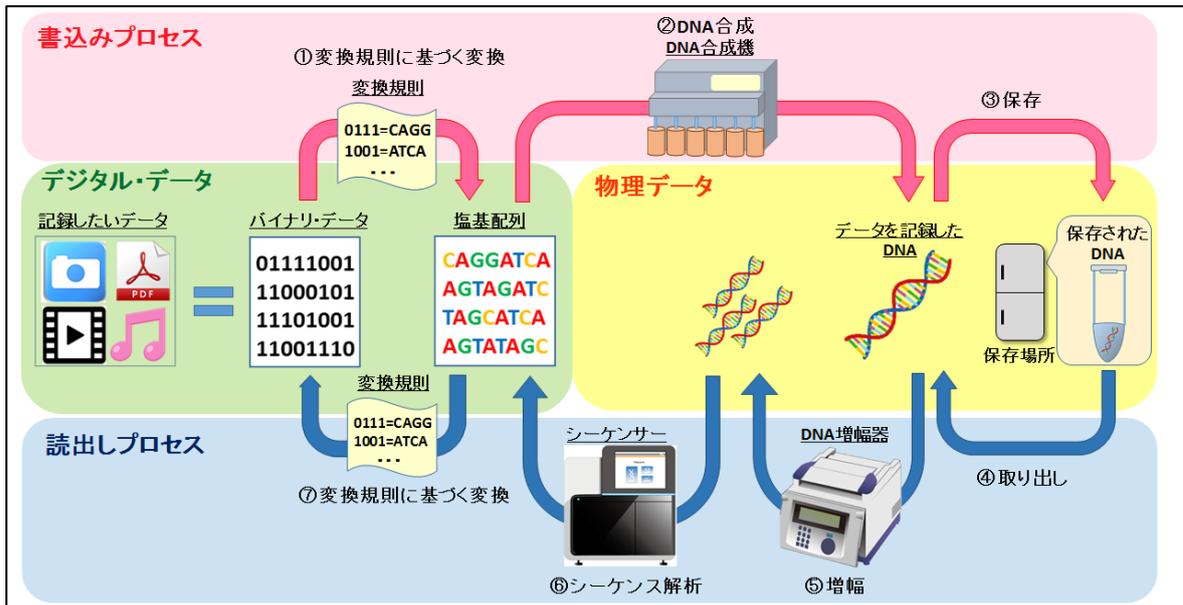
---

みをレーザーで加熱することにより、その領域のみ一時的に磁気の向きを変えてデータを書き込みやすくする。また、マイクロ波アシスト磁気記録方式では、磁性粒子が特定の周波数に共鳴して磁気の向きを変えやすくなる現象を用いて、特定領域のみデータを書き込みやすくする。

<sup>8</sup> 2017 年には、カートリッジ 1 巻あたり 330 テラバイトのデータ（従来の 20 倍に相当）を収容可能な磁気テープが開発されている（日本アイ・ビー・エム [2017]、ソニー・ソニーストレージメディアソリューションズ [2017]）。

<sup>9</sup> 安全にデータを保存することが可能な期間は製品により異なるが、それぞれの電子記録媒体の標準的な期間が示されている（例えば、<https://www.guardian-r.com/blog/osusume/20190315/1739/>）。

図表 1. DNA を用いたデータ記録の概要（イメージ）



備考：シーケンサーと DNA 増幅器のイメージは DBCLS TogoTV によるものを使用（© 2016 DBCLS TogoTV）。

- ② DNA 合成機を用いて、①で得られた塩基配列からなる DNA を合成する。
- ③ ②で合成した DNA を適切な環境下で保存する。

【読出しプロセスの処理の流れ】

- ④ ③で保存した DNA を取り出す。
- ⑤ DNA 増幅器を用いて、④で取り出した DNA を塩基配列の読出し（シーケンス解析）に必要となる十分な量まで増幅する。
- ⑥ ⑤で増幅した DNA を、シーケンス解析を行う装置（シーケンサー）に投入して塩基配列を読み出す。
- ⑦ ⑥で読み出した塩基配列を、変換規則に従ってバイナリ・データに変換する。

（2）各処理の概要と研究動向

本節（1）の①および⑦で用いる変換規則の適用や③における保存の一部は、DNA ストレージ固有の処理である。その他の処理は、バイオテクノロジー分野において一般的に用いられており、多くの知見の蓄積がある。以下では、それぞれの処理について、研究開発の現状と課題を整理する。

イ. 変換規則に基づく変換

変換規則を検討する際に考慮すべき主な観点として、①記録密度（DNA の単位量あたり記録可能なデータ量）、②DNA の構造安定性、③エラー耐性が挙げら

れる。これまでに、各観点に着目した変換規則が提案され、その概念実証が行われてきたが、筆者が知る限り、本稿執筆時点（2020年3月末）においては、すべての観点が考慮された汎用的な変換規則は報告されていない<sup>10</sup>。

#### （イ）記録密度の向上を企図した変換規則

データ表現に必要な塩基数が少ないほど、DNAストレージの記録密度は高くなる。例えば、出現頻度が高い文字に対して短い塩基配列を対応させるという発想を応用することにより、すべての文字に同じ塩基数の塩基配列を対応させる場合と比較して、全体のデータ量を削減する変換規則が提案されている（Ailenberg and Rotstein [2009]）。

#### （ロ）DNAの構造安定性を考慮した変換規則

同じ塩基が連続した配列を有するDNAや、「G」と「C」の合計の含有率が高すぎるDNAは、その構造が不安定となり、設計通りに合成できない場合がある。そこで、バイナリ・データを「A」、「T」、「C」の3塩基のみを用いて塩基配列に変換したうえで、1種類の塩基が4つ以上連続した場合には、3番目の塩基を「G」に変換することにより、DNAの構造安定性を高めた変換規則が提案されている（Goldman *et al.* [2013]）。

#### （ハ）エラー耐性を考慮した変換規則

DNAの増幅では、塩基が欠損したり他の塩基に置き換わったり（変異）するエラーが、一定の確率で発生しうる<sup>11</sup>。そこで、これらのエラーへの耐性を備えた変換規則として、誤り訂正符号を用いる方法や、同一のデータを異なる複数のDNAに保存して、それぞれの読出しの結果を総合することでバイナリ・データを復元する方法等が提案されている（Goldman *et al.* [2013]、Grass *et al.* [2015]、Blawat *et al.* [2016]）<sup>12</sup>。いずれの手法もデータに冗長性を持たせることにより記録密度が低下する点が課題となっていたが、近年では、エラー耐性と高い記録密度を両立する新たな手法が提案されている（Erlich and Zielinski [2017]）<sup>13</sup>。本手

---

<sup>10</sup> すべての観面で十分な条件を満たす変換規則の開発は、非常にハードルが高く実現困難であるとの見方もある。

<sup>11</sup> DNAは4種類の塩基を持つため、論理的には2ビットのデータを保存することができるが、一定確率で発生するエラーにより、実際には1.83ビットが上限であるといわれている（シャノン限界）。

<sup>12</sup> 誤り訂正符号とは、ノイズが加わりうる環境においてデータを取り扱う際に、そのデータに加えられたノイズを検出して訂正する技術である。

<sup>13</sup> 提案手法は、誤り訂正符号の一種である噴水符号を応用した手法であり、「DNAファウンテン」と名付けられている。本手法では、記録対象のデータを一定量のデータからなる複数のセグメントに分割したうえで、それらの中からランダムに選択した複数のセグメントを用いて一定

法を用いて、テキストおよび画像ファイル等からなるデータを、DNA1 グラムあたり 215 ペタバイト（2 億 1,500 万ギガバイト）の密度で書き込み、それを正しく読み出すことに成功したことが報告されている。

## ロ. DNA 合成

DNA は、糖、リン酸、塩基からなる構成単位（ヌクレオチド）が連なっている（補論 1 参照）。糖およびリン酸はすべてのヌクレオチドで共通であり、塩基の種類に対応してヌクレオチドも 4 種類存在する。DNA 合成とは、4 種類のヌクレオチドを材料として、これらを化学反応により順次連結することで目的の DNA を得ることである<sup>14</sup>。

DNA 合成の技術は、1950 年代後半に世界で初めて成功事例が報告され、その後、合成効率の向上、エラー率の低減、合成の自動化等に向けた研究が重ねられてきた<sup>15, 16</sup>。近年では、ウェブ上で希望の塩基配列を入力するだけで、その塩基配列からなる DNA の合成を依頼できる合成受託サービスが普及している。本稿執筆時点では、塩基数が 100 程度までの DNA は 1 塩基あたり数十円～数百円程度の価格、数日程度の納期で入手することができる<sup>17</sup>。それ以上の塩基数を有する DNA についても、1 万塩基程度までであれば 1 塩基あたり数十円程度の価格、数週間～1 か月程度の納期で入手可能である<sup>18</sup>。もっとも、DNA ストレージの実用化を展望すると、こうしたコストをさらに低減させることが求められると指摘されている（Bishop, Mccorkle, and Zhirnov [2017]）。

## ハ. DNA の保存と取出し

DNA は、マイナス 80℃程度の適切な環境下で凍結し、その状態を維持することにより、半永久的に安定して保存することができる。データの読出しの際には DNA を融解する必要があるが、凍結と融解を繰り返すと一部の DNA が分解す

---

の演算を繰り返し実行する。その結果から、合成する DNA の塩基配列を決定する。

<sup>14</sup> より詳細には、塩基数が 100 程度までの DNA（オリゴ DNA）は、ヌクレオチドを順次連結することにより合成し、100 から 1 万程度の塩基数の DNA（長鎖 DNA）は、まず、パーツとなる複数のオリゴ DNA を合成した後、それらを連結することにより合成する。

<sup>15</sup> 化学反応では、一部の原料が未反応のまま残存するなどして、理論上想定される生成物の量に対して、実際に得られる生成物の量が少なくなる場合がある。前者に対する後者の割合を合成効率と呼ぶ。

<sup>16</sup> 合成の自動化に関して、オリゴ DNA については、1 つの装置で 384 種類の DNA を並列して自動合成することが可能となっている一方で、長鎖 DNA については複数種類を並列して自動合成する技術は本稿執筆時点では確立されていない（野地ほか [2019]）。

<sup>17</sup> 数多くの企業がオリゴ DNA の合成受託サービスを提供している（例えば、<https://www.biologica.co.jp/products-service/custom-synthesis/nucleic-acid/dna-oligo/>）。DNA の精製度や濃度、合成場所（国内あるいは海外）等により納期や価格が異なる。

<sup>18</sup> 長鎖 DNA の合成受託サービスについても、多数の企業が提供している（例えば、[https://fasmac.co.jp/gene\\_craft/artificial\\_service](https://fasmac.co.jp/gene_craft/artificial_service)）。

るため、増幅過程で新しく複製された DNA を保存する。

DNA は、なるべく省スペースで保存することが望ましい。また、保存された複数の DNA から特定の DNA の取出しを可能とする設計は、データの読出し効率の向上につながる。これらの観点を考慮し、DNA を高密度かつ効率的に保存する手法として、約 1 テラバイト (1,000 ギガバイト) のデータを記録した DNA をガラス板上に格子点状 (各格子点は直径 0.275 ミリメートル) に配置する手法が提案されている (Newman *et al.* [2019])。本手法では、異なるデータを記録した DNA を各格子点に保存し、データを読み出す際には、読み出したいデータを記録した DNA が存在する格子点の DNA だけを取り出すことができる。

## 二. DNA の増幅

DNA の増幅手法としては、ポリメラーゼ連鎖反応 (polymerase chain reaction : PCR) と呼ばれる化学反応を用いて、ヌクレオチドを順次連結していく手法が普及している。PCR を用いた DNA の増幅には、増幅対象となる DNA の両端の塩基配列に依存する、短い DNA (プライマー) が必要となる (DNA の増幅手順の詳細については、補論 2 を参照)。DNA 増幅器を用いることにより、ごく微量の DNA を簡便かつ短時間で大量に増幅することが可能である<sup>19</sup>。

## ホ. シーケンス解析

### (イ) 次世代シーケンサー

1990 年に開始されたヒトゲノム解読プロジェクトでは、ヒト 1 人の DNA のシーケンス解析に約 13 年の歳月と 30 億ドルのコストを要した (Hayden [2014])<sup>20</sup>。しかし、2000 年代になると、複数の DNA を並列してシーケンス解析できる「次世代シーケンサー」の開発が進み、シーケンス解析に要する時間とコストはムーアの法則を上回るスピードで大幅に低下した。2017 年には、ヒト 1 人の DNA は約 1 日の時間と 1,000 ドルのコストで解析できるようになった (Wetterstrand [2019])<sup>21</sup>。

次世代シーケンサーの本体価格は数百万円～数千万円と高額であるほか、データ解析には専門的な知識を要する。近年では、シーケンス解析の受託サービ

---

<sup>19</sup> DNA の増幅は、まず、1970 年代に大腸菌等の生物を用いる手法が開発された。その後、1980 年代半になると、生物を用いず化学的に増幅する手法として、PCR を用いる手法が開発された。本稿執筆時点では、10 マイクロリットル程度の反応溶液を用意すれば、自動的に複数種類の DNA を並列して増幅できる装置が開発されており、10 分以内に DNA を 1 兆倍に増幅できる超高速装置も発売されている (例えば、<https://www.xpresspcr.com/>)。

<sup>20</sup> ヒトの DNA は、約 60 億の塩基の DNA がペアとなっていることが知られている。これは、1 塩基を 2 ビットのデータで表現できるとした場合、1.5 ギガバイトのデータに相当する。

<sup>21</sup> 次世代シーケンサーは、検出原理上、DNA 増幅器を用いた DNA の増幅とは別に、シーケンサー内部で再度 DNA を増幅し、その増幅過程で生じる蛍光等から間接的に塩基配列を解析する。

スを利用することにより、1 千億塩基の DNA について、数十万円の価格、数週間の納期で解析結果を得ることができる<sup>22</sup>。

#### (ロ) ナノポア・シーケンサー

シーケンス解析において、その場で解析結果を得ようとする場合には、高額なシーケンサーを入手する必要がある、受託サービスを利用する場合には、解析結果の入手に時間を要する点が課題となる。近年では、安価かつその場での解析を可能とする「ナノポア (nanopore)・シーケンサー」が注目されている。ナノポア・シーケンサーとは、微小な穴 (ナノポア) に DNA をくぐらせるときに発生する微小な電流をもとに DNA 構造を解析するものである。本稿執筆時点では、解析精度が 80~90%と低い点や、複数の DNA を並列して解析できない点が課題として残されているものの、①解析時間が相対的に短い、②本体価格が数十万円程度と安価である、③コンパクトで携帯できる、④解析可能な DNA の長さに理論的には上限がないなど、多数のメリットが存在する<sup>23</sup>。DNA ストレージにおいてナノポア・シーケンサーを活用するための研究も行われている (補論 3 を参照)。

#### (3) DNA ストレージの実用化に向けた取組み

DNA ストレージにおける一連の処理を実現する取組みも進められている。2012 年、ハーバード大学のジョージ・チャーチ教授らが、DNA ストレージの概念を初めて実証した結果を報告した (Church, Gao, and Kosuri [2012])。当研究では、バイナリ・データを「0」「1」「2」からなる三進数に変換し、さらにそれぞれのデータを 5 つの塩基配列に対応させた。この変換規則に基づき、約 5 万 3 千語のテキストおよび画像からなる 1 冊の本 (約 658 キロバイト) を、DNA1 グラムあたり 1.28 ペタバイト (128 万ギガバイト) の密度で記録し、さらにそのデータを読み出すことにも成功した。その後の研究により、2016 年には 200 メガバイト、2019 年には 1 ギガバイトのデータを、それぞれ DNA ストレージで記録できることが実証されている (Brunker [2016]、Langston [2019])。

また、最近の研究では、データの書込みおよび読出しにかかるすべての処理を自動化した装置の概念実証の事例が報告されている (Takahashi *et al.* [2019]、

---

<sup>22</sup> 次世代シーケンサーによる受託サービスも、さまざまな企業が提供している (例えば、[http://catalog.takara-bio.co.jp/jutaku/basic\\_info.php?catcd=B1000482&subcatcd=B1000707&unitid=U100005162](http://catalog.takara-bio.co.jp/jutaku/basic_info.php?catcd=B1000482&subcatcd=B1000707&unitid=U100005162))。

<sup>23</sup> ナノポア・シーケンサーでは、解析対象の DNA の分子構造を直接解析するため、解析に要する時間が相対的に短い。また、解析ごとに生じるコスト (ランニングコスト) は、専用の溶液と容器の費用であるため、塩基数によらず一定となる。なお、次世代シーケンサーでは、解析対象の塩基数が多いほどランニングコストは増加する。

Langston [2019])。本研究では、「hello」という 5 バイトのデータの書込みと読出しに 21 時間を要したほか、装置の製造費用として約 1 万ドルを要しているものの、DNA ストレージの全自動化の可能性を示す研究成果として評価されている<sup>24</sup>。もっとも、実用化に際しては、記録可能な容量の増加、書込みと読出しに要する時間の短縮、装置の費用の削減、メンテナンス方法の検討、操作性の向上等、検討すべき課題が残されているのが現状である<sup>25</sup>。

#### 4. DNA ストレージのセキュリティに関する分析と対策

DNA ストレージを実用化するうえでは、記録対象のデータを適切に保護するための各種対策を講じておくことが必要である。それらの 1 つとして、以下では、記録対象のデータのセキュリティに焦点を当てて、そのリスクや対策について検討する<sup>26</sup>。

##### (1) セキュリティ分析の方向性と前提

本節では、DNA ストレージを実装するシステムとして想定される形態を 1 つ取り上げ、一般的な情報システムと同様、取り扱われるデータやシステムの機能の機密性 (confidentiality)・完全性 (integrity)・可用性 (availability) の観点からセキュリティ分析を試みる。

##### イ. 分析対象とするシステムとその構成

3 節 (3) でみたとおり、DNA ストレージの自動化装置は概念実証が報告されているが、顧客の環境下に設置して顧客自らがデータの読書きを行う汎用的な製品にするためには、多くの課題がある。そこで、本稿執筆時点での技術を基に、今後研究開発が進展したときに実現可能性の高いケースとして、さまざまな顧客から預かったデータを DNA の形で保存するデータ・センターのシステム (DNA ストレージ・システム) の形態の 1 つを想定する。DNA ストレージ・システムにおける各プロセスの処理の流れ、および前提事項は以下のとおりである。それぞれのプロセスに付した番号は、図表 2 中の番号に対応する。

---

<sup>24</sup> 現状では、DNA 合成機や DNA 増幅器、シーケンサーといった各工程で用いる装置は自動化が進んでいるものの、前後の処理には人手を要するのが一般的である。

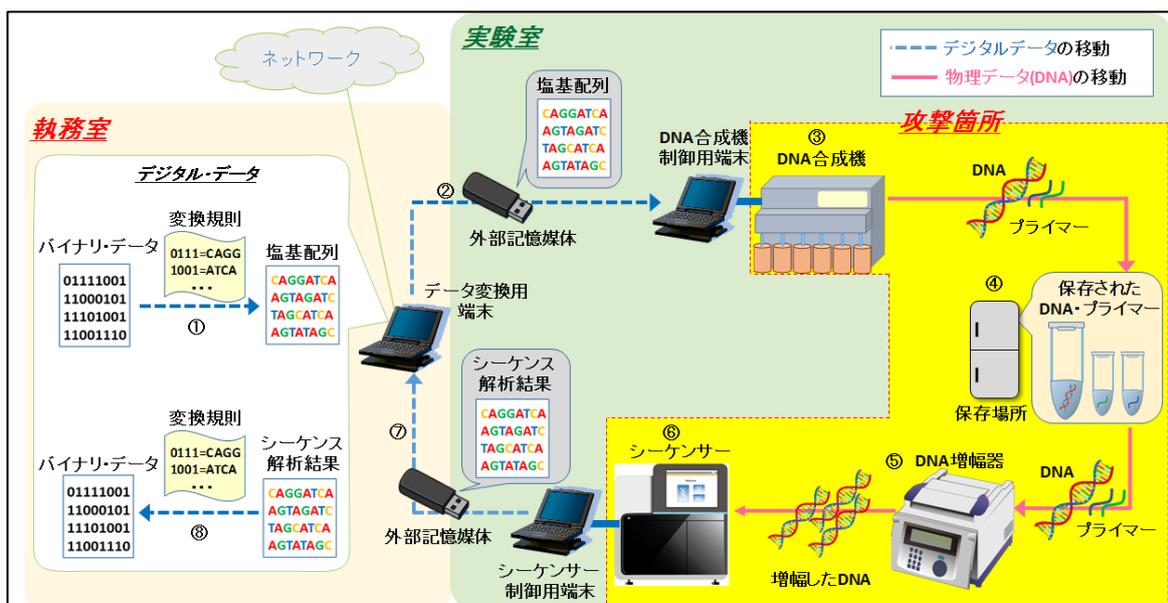
<sup>25</sup> DNA ストレージの実用化にあたっては、ランダム・アクセス機能を実装することも重要である。バイナリ・データを変換して得られた塩基配列を加工することにより、こうした機能を実現する手法も提案されている (補論 4 を参照)。

<sup>26</sup> DNA ストレージによって長期にわたりデータを記録・保存するサービスを提供するためには、DNA を保存する装置や設備の長期的な安定性や高い信頼性の実現が前提となる。例えば、各種装置やそれらに電力を供給する電源設備を安定稼働させる必要があるほか、地震等の災害発生時の対応も充実させておくことが求められる。こうした物理的な安全対策に関する検討も、今後必要になってくると考えられる。

## 【書込みプロセス】

- ① サービス提供業者の執務室にあるデータ変換用端末上で、顧客によって持ち込まれたバイナリ・データに変換規則を適用して塩基配列に変換する。
  - ・変換規則は公知である。
  - ・データ変換用端末は、インターネットやメール・システムを利用できるよう、外部ネットワークに接続している。
- ② 得られた塩基配列を USB メモリ等の外部記憶媒体に記録して実験室に持ち運び、DNA 合成機に直接接続された端末（DNA 合成機制御用端末）に接続して塩基配列をコピーする。
  - ・DNA 合成機制御用端末は、外部ネットワークには接続していない。
- ③ 顧客のデータを記録した DNA とそれを増幅する際に用いるプライマーを、DNA 合成機を用いて合成する。
  - ・合成された DNA とプライマーは、シリアル番号を付して管理する。各作業担当者は、シリアル番号からその DNA とプライマーの属性（依頼者やデータの概要等）を照合し入手可能である。
  - ・DNA とそれに対応するプライマーは常にセットで取り扱われ、DNA 合成機から保存場所、保存場所から DNA 増幅器、DNA 増幅器からシーケンサーへの運搬は、すべて人手を介して行われる。
- ④ 合成された DNA とプライマーのセットを適切な条件下で保存する。

図表 2. DNA ストレージ・システムの構成（概念図）



備考：シーケンサーと DNA 増幅器のイメージは DBCLS TogoTV によるものを使用（© 2016 DBCLS TogoTV）。

### 【読出しプロセス】

- ⑤ 保存場所から DNA とプライマーのセットを取り出し、DNA 増幅器で DNA を増幅する。
- ⑥ シーケンサーを用いてシーケンス解析を行う。シーケンス解析結果は、シーケンサーに直接接続された端末（シーケンサー制御用端末）に出力される<sup>27</sup>。
  - ・シーケンス解析には、次世代シーケンサーまたはナノポア・シーケンサーを用いる。
  - ・シーケンサー制御用端末は、外部ネットワークには接続していない。
- ⑦ 得られたシーケンス解析結果を、USB メモリ等の外部記憶媒体に記録して執務室に持ち運び、データ変換用端末に接続してシーケンス解析結果をコピーする。
- ⑧ データ変換用端末上で、シーケンス解析結果に変換規則を適用し、バイナリ・データに変換する。

### ロ. セキュリティ対策の前提と攻撃箇所

執務室と実験室は常時施錠されており、入退室には ID カードを用いたドアの解錠が必要であるものとする<sup>28</sup>。

デジタル・データ（バイナリ・データ、塩基配列、シーケンス解析結果）を扱う各端末（データ変換用端末、DNA 合成機制御用端末、シーケンサー制御用端末）および外部記憶媒体は、十分なセキュリティ対策を講じているとする。例えば、外部ネットワークを介したサイバー攻撃に対しては、一般的な情報システムや制御システムと同様に、ファイアウォールや不正侵入検知システムの導入、デジタル・データの暗号化や各端末へのマルウェア対策ソフトの導入等が挙げられる<sup>29</sup>。また、サービス提供者自体は組織的な不正を行わないものの、各端末や外部記憶媒体を取り扱う作業担当者が攻撃者と結託して不正行為を行う可能性があるとして想定する。そこで、各端末で操作記録（ログ）を取るとともに、作業

<sup>27</sup> DNA の塩基配列を読み出した後、シーケンサー内部で一定のデータ形式に整えたり圧縮したりするなどの後処理を実行したものが、シーケンス解析結果として出力されるケースがある。

<sup>28</sup> 入退室が行われた際の ID と時間が記録・管理されているものとする。外部業者等、ID カードを所有していない者が入室する際は、予め ID カードの貸与を受ける必要がある。

<sup>29</sup> 近年、制御システムのセキュリティ・リスクの分析とその対策方針にかかる検討が進展しており、ガイドライン等が公表されている（例えば、情報処理推進機構セキュリティセンター [2020]）。制御システムとは、製造業における生産ラインのシステムや鉄道や航空機のシステム等、情報系でないシステムを制御するための情報システムの総称である。DNA ストレージ・システムは、DNA 合成機、DNA 増幅器、シーケンサーといった、DNA の取扱いに特化した機器と、それらを制御するための端末や外部記憶媒体から構成されているとみられる。そうした観点に立てば、端末や外部記憶媒体を制御システムとみなし、リスク分析やセキュリティ対策を検討する際に上記のガイドライン等を活用することもできると考えられる。

担当者がログを改変できないかたちで管理することにより、事後的に不正を検知するなどのセキュリティ対策を講じているものとする。

一方、DNA は、①DNA 合成機、②保存場所、③DNA 増幅器、④シーケンサーでの作業過程において、攻撃者と結託した作業担当者による物理的な攻撃を受けるリスクがある。DNA は、目に見えないほどの少量であっても、漏洩すると塩基配列が解析される可能性があるほか、異物が混入されるとその物質の影響を受けて各作業が意図したとおりに行われなくなる可能性がある。そのため、こうした物理的な攻撃が実行されうると想定する<sup>30</sup>。

## ハ. セキュリティ目標

サービス提供者は、顧客から預かったデータについて、予め定められた期間内において、機密性の観点からは漏洩しないようにすること、完全性の観点からは改ざんされないようにすることが目標となる。可用性の観点からは、顧客が、DNA からデータを読み出したい時に、円滑に読出しを行うことができるようにすることが目標となる。

## 二. 攻撃者の目的と能力

攻撃者は、DNA ストレージ・システムにおいて扱われるデータの盗取（機密性への攻撃）、改ざん（完全性への攻撃）、読出しの妨害（可用性への攻撃）を試みる。攻撃者は、DNA ストレージ・システムの運用に関与しない第三者とするが、外部業者を装って実験室に入室したり、上記の作業担当者と結託したりすることができる想定する。

### （２）想定される攻撃とその対策

#### イ. データの盗取

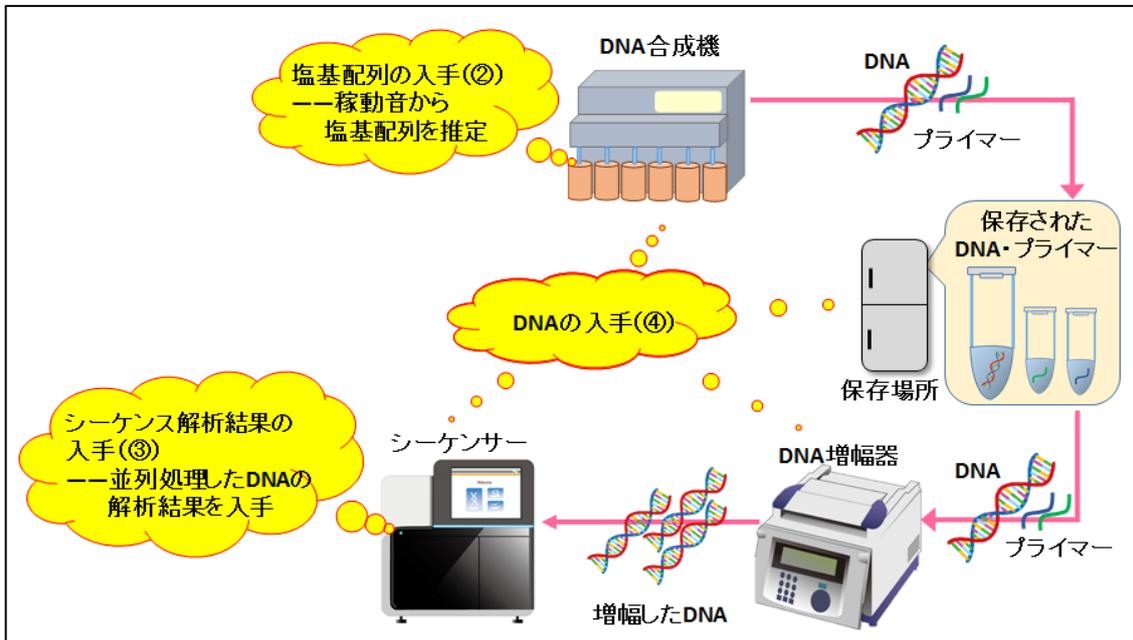
##### （イ）攻撃の方針

DNA に記録されるデータを攻撃者が入手する方法として、次の４つが考えられる。①バイナリ・データを直接入手する。②塩基配列を入手し、それに変換規則を適用してバイナリ・データに変換する。③シーケンス解析結果を入手し、それに変換規則を適用してバイナリ・データに変換する。④DNA を入手した後、そのシーケンス解析を行い、得られたシーケンス解析結果に変換規則を適用してバイナリ・データに変換する。これらのうち、①については、本節（１）ロ. のとおり、バイナリ・データのセキュリティが確保されていることから、攻撃は

---

<sup>30</sup> なお、DNA ストレージ・システムを構成する各種の装置や設備に関しては、攻撃に伴うものではない故障や自然災害への対策を十分に実施しており、サービス提供に求められるレベルの安定性や信頼性を維持していることとして議論を進める。

図表 3. データの盗取にかかる攻撃の概要



備考：シーケンサーと DNA 増幅器のイメージは DBCLS TogoTV によるものを使用（© 2016 DBCLS TogoTV）。また、図表中の②～④は 4 節（2）イ.（イ）の本文中の番号に対応する。

成功しない。したがって、攻撃者は、②～④の方法による攻撃を試みると想定し、以下でそれぞれについて検討する（図表 3 を参照）。

#### （ロ）塩基配列の入手（DNA 合成機の脆弱性の悪用）

攻撃者または攻撃者と結託した作業担当者は、端末や外部記憶媒体を不正に操作して入手することができないため、DNA 合成機から何らかの手段を用いて塩基配列を入手する必要がある。そうした方法として、DNA 合成機の稼働音をもとに、合成された DNA の塩基配列を非侵襲的に推定する手法が提案されている（Faezi *et al.* [2019]）。本手法では、ヌクレオチドが DNA 合成機の内部を通過する時に生じる音から、合成に用いられたヌクレオチドの種類を順次特定していく。Faezi *et al.* [2019]では、1 台の DNA 合成機を対象に提案手法を実験している。その合成機から 70 センチメートル以内に録音装置を設置して DNA 合成時の稼働音を録音して分析したところ、実験室内にある程度の雑音が存在する状況下であっても、合成された DNA の塩基配列のうち約 88%の配列を推定することに成功したと報告している。

この手法を用いるケースの 1 つとして、攻撃者と結託した作業担当者が DNA 合成機の稼働音を録音しそのデータを攻撃者に渡すケースが考えられる。DNA 合成機に近づくことが可能であって合成対象の DNA の属性を把握可能な作業

担当者は、目的のデータを DNA に書き込むタイミングを予め特定したうえで攻撃を実行することが可能である。

一方、攻撃者が外部業者を装って実験室に盗聴器を設置したり、実験室内に持ち込まれるスマートフォンに盗聴に使用できるアプリをインストールしたりして、DNA 合成機の稼働音を継続的に入手するケースも考えられる<sup>31</sup>。こうしたケースでは、攻撃者は入手した稼働音から塩基配列を推定し、さらにその塩基配列をバイナリ・データに変換するまで、その稼働音がどのデータに関するものかについて識別することはできない。したがって、攻撃者の観点からは、入手した情報を得られない可能性があり、攻撃の効果は相対的に低い<sup>32</sup>。

もっとも、サービス提供業者の観点からは、いずれのケースにおいても、顧客データが漏洩しないようにするという機密性にかかるセキュリティ目標を達成できないことになるため、サービス提供業者として対策を採ることが必要になると考えられる。

対策として、データを書き込む際にバイナリ・データを暗号化してから塩基配列に変換するという方法が考えられる。これにより、塩基配列が盗取された場合にも記録したデータの漏洩を防ぐことができる<sup>33</sup>。ただし、DNA ストレージ・システムではデータを長期的に保存することが想定されるため、その間、常に安全な状態で暗号を適用できるようにすることが求められる。例えば、暗号化に用いた鍵を安全に保管する体制の整備や、量子コンピュータの脅威が顕在化する場合に備えて、それに耐性のある暗号（耐量子計算機暗号）を採用すること、採用した暗号が危殆化した場合に、安全な暗号への移行を検討すること等が挙げられる。

このほか、運用上の対策としては、①各 DNA のシリアル番号に紐づく属性情報を作業担当者に知られない（攻撃対象とする DNA<標的 DNA>を特定できない）ようにする、②実験室への入退室管理を金属探知機等によって厳しく行い、

---

<sup>31</sup> 盗聴に使用できるアプリの一部では、インストールされても画面に表示されないなど、スマートフォンの持ち主でも認識困難なものが存在している（例えば、<https://sat-sagasu.com/keitaidenwa-koteidenwa-tochoki>）。攻撃者が、こうしたアプリを善意の作業担当者が自分のスマートフォンにインストールするように誘導し、そのアプリを遠隔操作して稼働音を録音するといった手法もありうる。

<sup>32</sup> 攻撃の目標が、「任意の DNA に格納されている情報をどれでもよいので入手し、それを暴露することによって、当該サービス提供業者のレピュテーションを低下させる」というものであるならば、攻撃は成功することになる。

<sup>33</sup> DNA に書き込むデータ全体を一度に暗号化する場合、そのデータの一部だけを読み出そうとすると、復号するために必要なデータに対応する塩基配列の箇所を特定できず、復号困難なケースが想定される。そうしたケースでは、目的のデータのみを効率的に読み出すランダム・アクセス機能を適用することが難しくなる。データをセグメントごとに分割したうえで、それぞれのセグメントを個別に暗号化することも考えられるが、データの規模が大きくなると、セグメント数も増えることから、暗号化にかかる手間が大きくなる。

録音装置を持ち込ませないようにする、③DNA 合成機を遮音された個室に設置し、当個室へ入室できる人員をより厳格に制限する等が考えられる。

#### (ハ) シーケンス解析結果の入手（次世代シーケンサーの脆弱性の悪用）

次世代シーケンサーを用いて複数の DNA を並列して解析する場合、それぞれの DNA にかかるシーケンス解析結果の一部が、少量ではあるものの、並列処理した他の DNA のシーケンス解析結果の一部に反映されてシーケンサーから出力される可能性があると報告されている (Ney *et al.* [2017])<sup>34</sup>。このシーケンサーの脆弱性を悪用する攻撃者は、DNA ストレージ・サービスの顧客として、自らが保存を依頼したデータの読出しを依頼し、シーケンス解析結果を入手する。得られたシーケンス解析結果から、同時にシーケンス解析された別の顧客の DNA のシーケンス解析結果の一部を入手する攻撃が想定される。

攻撃者がシーケンサーの作業担当者と結託する場合、標的 DNA と同じタイミングで攻撃者の DNA のシーケンス解析を行うようにコントロールすることで、標的 DNA に記録されたデータを選択的に入手しうると考えられる。ただし、得られるデータが DNA のどの箇所のものかはランダムであり、コントロールすることはできない。攻撃者がシーケンサーの作業担当者と結託しないケースでは、標的 DNA と同じタイミングでシーケンス解析を行うようにすることも困難であると考えられる。

対策として、データを書き込む際にバイナリ・データを暗号化してから塩基配列に変換することが考えられる<sup>35</sup>。このほか、DNA のシリアル番号に紐づく属性情報を作業担当者に知られないようにする、異なる顧客のシーケンス解析を同時に行わないようにするといった運用が挙げられる。また、こうした脆弱性が存在しないナノポア・シーケンサーを用いることが考えられる。

## (二) DNA の入手

DNA の入手のシナリオとして、以下が想定される。まず、攻撃者は、DNA 合成機からシーケンサーに至るまでのプロセスの作業担当者と結託する。その作業担当者は、標的 DNA の一部を取り出して持ち出すか、その DNA が付着している実験器具の一部（パーツ）等を持ち出して攻撃者に渡す。攻撃者は、入手し

---

<sup>34</sup> 他の DNA のシーケンス解析結果に反映されるデータ量は、使用するシーケンサーの種類や解析対象の DNA の特性等、さまざまな要因の影響を受けうるが、Ney *et al.* [2017]による検証では、標的 DNA の塩基配列の 0.007%が、並列処理した他の DNA のシーケンス解析結果に反映された。

<sup>35</sup> 攻撃者の DNA と並列でシーケンス解析した他の顧客の DNA が同じ鍵で暗号化されている場合には、攻撃者の DNA のシーケンス解析結果に反映された他の顧客の DNA のシーケンス解析結果が正しく復号され、情報が漏洩する可能性が残る。

た標的 DNA を必要に応じて増幅したうえで、シーケンス解析を行う。ただし、標的 DNA の増幅にはプライマーが必要となるため、それも持ち出して入手しておく必要がある。

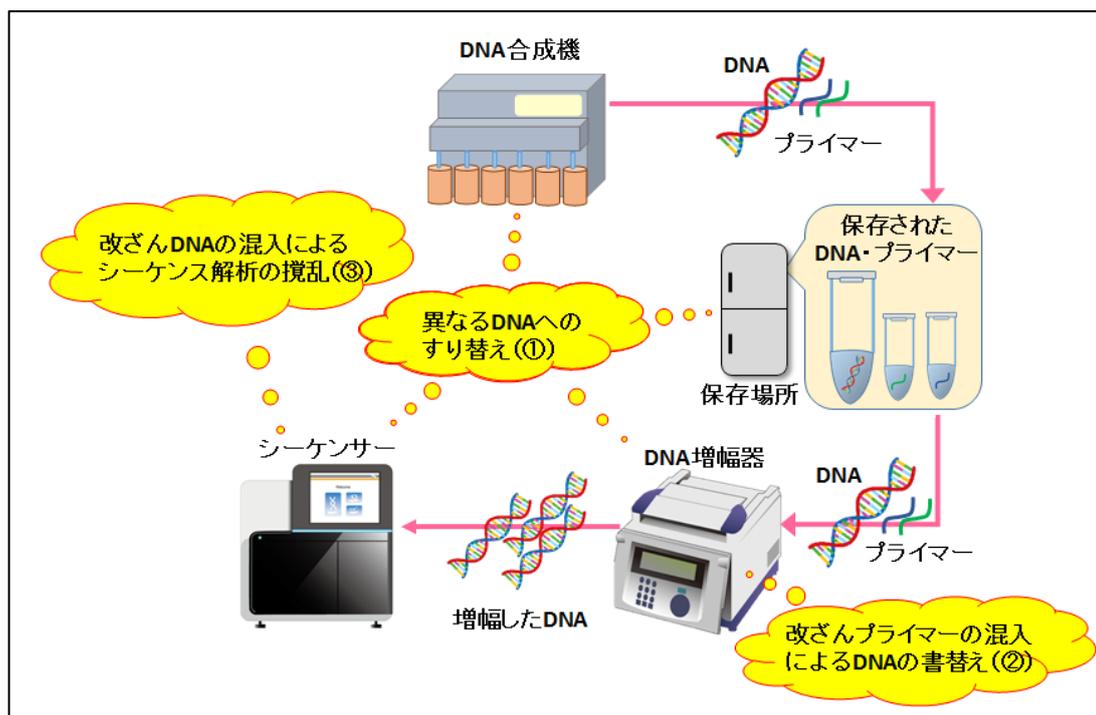
対策として、各 DNA のシリアル番号に紐づく属性情報を作業担当者に知られないようにすることや、データを書き込む際に、バイナリ・データを暗号化してから塩基配列に変換することが考えられる。また、実験室への入退室管理を厳しく行い、盗取した DNA を収容した容器や DNA が付着したパーツの持出しをより難しくすることも一定の効果があると考えられる<sup>36</sup>。

## ロ. データの改ざん

### (イ) 攻撃の方針

攻撃者は、①DNA のすり替え、②増幅時における DNA の書換え、③シーケンス解析の攪乱により、DNA ストレージ・システムに記録されたデータの改ざんを試みる (図表 4 を参照)。

図表 4. データの改ざんにかかる攻撃の概要



備考：シーケンサーと DNA 増幅器のイメージは DBCLS TogoTV によるものを使用 (© 2016 DBCLS TogoTV)。図表中の①～③は4節 (2) ロ. (イ) の本文中の番号に対応する。

<sup>36</sup> DNA は少量であってもその情報を解析することができるため、攻撃者と結託した作業担当者は、例えば、身の回りの所有物に付着させて持ち出すことが考えられる。このため、DNA の持出しを完全に防止することは困難であるとみられる。

#### (ロ) DNA のすり替え

攻撃者は事前に DNA 合成受託サービス等を利用して、標的 DNA に反映させたいデータを記録した DNA (改ざん DNA) とそれに対応するプライマーを準備する。攻撃者と結託した作業担当者は、改ざん DNA およびプライマーを実験室内に持ち込み、標的 DNA およびそれに対応するプライマーとすり替える。攻撃のタイミングは、DNA 合成機からシーケンサーに至るまでのプロセスが想定される。DNA を取り扱うのは作業担当者に限られることから、この攻撃は作業担当者との結託が必要となる。

運用上の対策として、各 DNA のシリアル番号に紐づく属性情報を作業担当者には知られないようにすることや、実験室への入退室管理を厳しく行うことにより、改ざん DNA の持込みをより困難にすることが考えられる<sup>37</sup>。攻撃検知の観点からは、バイナリ・データに予めデジタル署名を付して記録することにより、データの読出し後にデータの改ざんを検知できるようにしておくことが考えられる (Kar *et al.* [2018]、Kar and Ray [2019])。暗号化によりデータを保護するケースと同様、デジタル署名についても長期的に安全な方式を採用する必要があるほか、署名方式が危殆化した際には、署名方式をより安全なものに移行するとともに、そうした方式によって署名を生成し直すことも求められる。また、署名の検証に用いる鍵等を長期間安全に管理し続ける必要もある。

#### (ハ) 増幅時における DNA の書換え

攻撃者は、標的 DNA の増幅時に、正規のプライマーの代わりに標的 DNA に反映させたいデータを含む短い DNA (改ざんプライマー) を用いることにより、標的 DNA の一部を書き換えることが理論的には可能なケースが考えられる (補論 5 を参照)<sup>38</sup>。

攻撃者と結託した DNA 増幅器の作業担当者は、改ざんプライマーを実験室に持ち込み、それを標的 DNA の増幅時に使用する。なお、改ざんプライマーの設計にあたって、攻撃者は、本節 (2) イ. (二) の手法等により標的 DNA を盗取し、予めそのシーケンス解析を実行して塩基配列を入手しておく必要がある。また、改ざんプライマーは標的 DNA と結合可能な塩基配列である必要があることから、書換え可能な塩基の数は、DNA のすり替えの場合と比較して少なくなる。

本攻撃への対策としては、本節 (2) ロ. (ロ) のケースと同様、①各 DNA のシリアル番号に紐づく属性情報を作業担当者には知られないようにすること、②

<sup>37</sup> 改ざん DNA は非常にコンパクトな容器に収容することが可能であるため、持ち物のチェック等により改ざん DNA の持込みを完全に防止することは困難であると考えられる。

<sup>38</sup> 本攻撃は、DNA の塩基を置換する際にバイオテクノロジー分野において一般的に用いられる手法を応用することにより、理論上可能であると考えられる手法であり、実験によってその有効性が確かめられたものではない。

実験室への入退室管理を厳しく行い、標的 DNA の持出しや改ざんプライマーの持込みをより困難にすること、③バイナリ・データに予めデジタル署名を付して記録することにより、読出し後にデータの改ざんを検知できるようにしておくことが考えられる。

## (二) シーケンス解析の攪乱

シーケンス解析時に改ざん DNA を混入することにより、改ざん DNA の塩基配列が標的 DNA のシーケンス解析結果に反映され、標的 DNA のシーケンス解析が攪乱されるケースがある (Ney *et al.* [2017])<sup>39</sup>。こうしたケースを悪用し、攻撃者と結託したシーケンサーの作業担当者が改ざん DNA の混入を実行するケースが想定される。また、攻撃者は、シーケンス解析の操作過程で使用する実験器具のうち、標的 DNA に触れる使い捨てパーツ等に予め改ざん DNA を塗布しておき、外部業者を装って実験室に侵入した際にそれを置いておくことが考えられる (Islam *et al.* [2019])。この場合、当該パーツが使用されるタイミングを攻撃者は制御できず、標的 DNA とは別の DNA のシーケンス解析が攪乱される可能性がある。

上記のいずれのケースにおいても、改ざん DNA の塩基配列に制約はないものの、標的 DNA における攪乱対象の部位をコントロールすることは困難であると考えられる。したがって、攻撃者が標的 DNA の塩基配列のどの部分でも構わないので書き換えたいと考える場合に、この攻撃を試みると想定される。

本攻撃の対策については、①各 DNA のシリアル番号に紐づく属性情報を作業担当者に知られないようにすること、②実験室への入退室管理を厳しく行い改ざん DNA の持込みをより困難にすること、③バイナリ・データに予めデジタル署名を付して記録することにより、読出し後にデータの改ざんを検知できるようにしておくことが考えられる。

## ハ. データの読出しの妨害

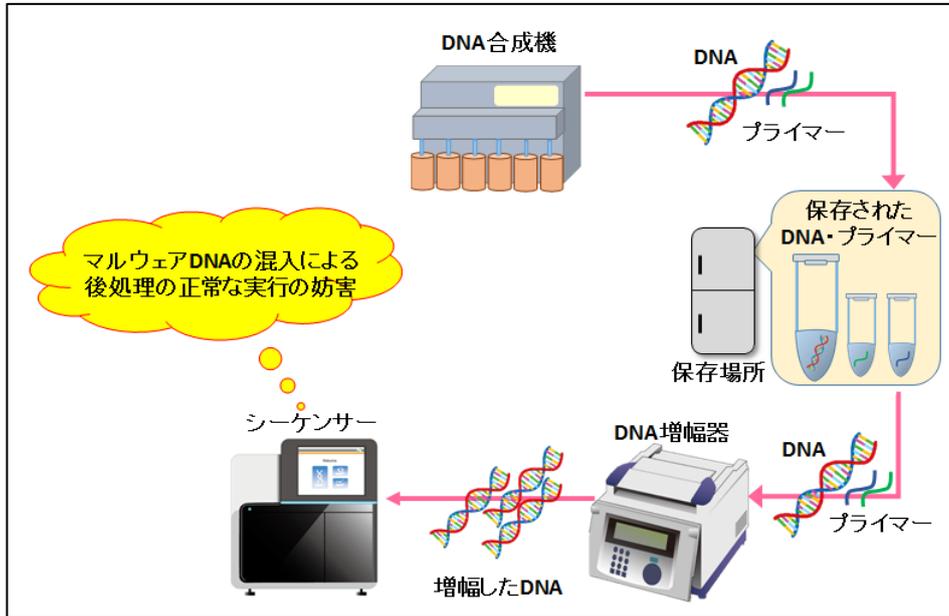
### (イ) 攻撃の方針

シーケンサーは、DNA の塩基配列を読み出した後、それを一定のデータ形式に整えたり圧縮したりするなどの後処理を内部で実行したうえで、その結果をシーケンサー制御用端末に出力するというケースがありうる。こうした後処理

---

<sup>39</sup> こうしたケースに加えて、次世代シーケンサーの場合、本節(2)イ.(ハ)で取り上げたシーケンサーの脆弱性を悪用した攻撃として、攻撃者が、標的 DNA と攻撃者の DNA を並列処理するようにコントロールし、標的 DNA のシーケンス解析結果に攻撃者の DNA のシーケンス解析結果を反映させて攪乱する攻撃が考えられる。

図表 5. データの読出しの妨害にかかる攻撃の概要



備考：シーケンサーと DNA 増幅器のイメージは DBCLS TogoTV によるものを使用 (© 2016 DBCLS TogoTV)。

用のプログラムに脆弱性が存在した場合、攻撃者がそれを悪用することによって、後処理の正常な実行を妨害することを企図した攻撃を試みる可能性がある (図表 5 を参照)<sup>40</sup>。このような攻撃が成功すると、シーケンサーがそのタイミングにおいて DNA のシーケンス解析結果を出力できない場合が想定される。

一方、DNA 増幅器については、小型かつ安価な装置であり、一般的に容易に代替機器を入手・使用できるため、攻撃者がそれを攻撃するインセンティブは働かないものとする。

#### (ロ) 攻撃手法

攻撃者は、攻撃対象のシーケンサーに内蔵されている後処理用のプログラムの脆弱性に関する情報を入手する。次に、プログラムの一部の書換えが発生するようなコードを設計する。そのうえで、そのコードに対応する塩基配列を特定し、当該塩基配列を有する DNA (マルウェア DNA) を合成する<sup>41</sup>。

<sup>40</sup> Ney *et al.* [2017]は、次世代シーケンサーを用いて得られたシーケンス解析結果のデータの後処理や分析用のプログラムを対象に、バッファ・オーバーフロー攻撃に対する脆弱性の有無を調査したところ、脆弱性につながりうる関数呼出しが多数含まれていたとの結果を報告している。なお、バッファ・オーバーフロー攻撃は、攻撃対象のソフトウェアが入力許容容量を超えたデータの処理を行う際に、処理しきれないデータが、本来想定していない領域のプログラムを上書きしてしまう脆弱性を悪用した攻撃手法である。

<sup>41</sup> Ney *et al.* [2017]は、まず、バッファ・オーバーフローの脆弱性を有する、塩基配列のデータを圧縮するプログラムを準備するとともに、バッファ・オーバーフロー攻撃用のコードを作成してそれを塩基配列として埋め込んだマルウェア DNA を合成している。そのうえで、マルウェア

攻撃者は、結託したシーケンサーの作業担当者にマルウェア DNA を渡し、シーケンス解析対象となる DNA に混入させる。また、攻撃者は、シーケンス解析の操作過程で使用する実験器具の使い捨てパーツ等に予め改ざん DNA を塗布しておき、外部業者を装って実験室に侵入した際にそれを置いておくことが考えられる (Islam *et al.* [2019])。

もともと、攻撃者は、シーケンサー内蔵のソフトウェアの脆弱性に関する情報を予め入手しておく必要がある。また、マルウェア DNA は、攻撃の実行に必要なコードに対応する塩基配列を有しつつ、その塩基配列が物理的に合成可能な配列となるように設計する必要もあり、実行のハードルは相応に高いものと考えられる。

対策として、まず、シーケンサーに内蔵されているプログラムの脆弱性を解消することが求められる。プログラムがセキュリティに配慮して生成されているか (セキュア・コーディングが実践されているか) を確認するとともに、セキュリティ・パッチを適切に適用することが有用である。また、実験室への入退室管理を厳しく行い、悪意ある DNA や実験器具のパーツの持込みをより困難にすることが考えられる。

### (3) セキュリティ分析の総括

本節 (2) で取り上げた攻撃の概要の総括を図表 6 に示す。

それぞれの攻撃手段について影響の多寡を比較する。データの盗取については、漏洩しうる情報量が多い順に「DNA の入手」、「塩基配列の入手 (DNA 合成機の脆弱性を悪用)」、「シーケンス解析結果の入手 (シーケンサーの脆弱性の悪用)」となる。データの改ざんについては、改ざんの影響が大きい順に「DNA のすり替え」、「増幅時における DNA の書換え」、「シーケンス解析の攪乱」となる。

セキュリティ対策を講じる際には、①各攻撃手段の影響の多寡を考慮してどの攻撃手段への対策を講じるべきかを判断するとともに、②どのようなタイプの攻撃への対策を講じるべきか (攻撃者と作業担当者の結託は想定するか、対策の対象は特定の DNA を標的とする攻撃か不特定の DNA を標的とする攻撃かなど) を判断する。そのうえで、③候補となる各対策手法の効果や実行可能性、コスト等を総合的に評価・比較しながら、対策手法を決定する。1つの対策手法では充足できないと考えられる場合には、複数の対策手法を組み合わせることも検討する。

---

ア DNA のシーケンス解析結果を次世代シーケンサーによって得た後、それを当該プログラムに入力したところ、一定の確率で、意図したとおりにプログラムが書き換えられたことを確認している。

図表 6. DNA ストレージ・システムにおいて想定されうる攻撃と影響・対策

目的	攻撃				攻撃に必要な行為	攻撃の影響	対策手法の例		
	手段	攻撃の場所	結託者	攻撃対象			技術的な対策	運用上の対策	
データの盗取	塩基配列の入手 (DNA 合成機の脆弱性を悪用)	DNA 合成機	実験室内の作業担当者	特定の DNA	録音装置の設置	合成された DNA の一部の塩基配列が推定されうる	バイナリ・データの暗号化	<ul style="list-style-type: none"> <li>属性情報の管理の厳格化</li> <li>DNA 合成機へのアクセスの厳格化</li> <li>実験室の入室管理を厳格化</li> </ul>	
			—	不特定の DNA					
	シーケンス解析結果の入手 (次世代シーケンサーの脆弱性の悪用)	シーケンサー (次世代シーケンサー)	左記の作業担当者	特定の DNA	データ読出しの依頼	並列処理された DNA の塩基配列の一部が推定されうる (位置は制御困難)			<ul style="list-style-type: none"> <li>DNA 合成機へのアクセスの厳格化</li> <li>実験室の入室管理の厳格化</li> </ul>
			—	不特定の DNA					
DNA の入手	DNA 合成機、保存場所、DNA 増幅器、シーケンサー	左記の作業担当者	特定の DNA	—	盗取された DNA のすべてのデータが漏洩しうる	<ul style="list-style-type: none"> <li>属性情報の管理の厳格化</li> <li>実験室の退室管理の厳格化</li> </ul>			
データの改ざん	DNA のすり替え	DNA 合成機、保存場所、DNA 増幅器、シーケンサー	左記の作業担当者	特定の DNA	改ざん DNA とそのプライマーの合成	すり替えられた DNA のすべてのデータが改ざんされうる	バイナリ・データへのデジタル署名付加	<ul style="list-style-type: none"> <li>属性情報の管理の厳格化</li> <li>実験室の入室管理の厳格化</li> </ul>	
	増幅時における DNA の書換え	DNA 増幅器	左記の作業担当者	特定の DNA	<ul style="list-style-type: none"> <li>標的 DNA の入手とシーケンス解析</li> <li>改ざんプライマーの合成</li> </ul>	標的 DNA の一部 (予め意図した箇所) が書き換えられうる			
	シーケンス解析の攪乱	シーケンサー	左記の作業担当者	特定の DNA	改ざん DNA の合成	シーケンス解析結果の一部が攪乱されうる (その箇所は制御困難)			実験室の入室管理の厳格化
—			不特定の DNA	<ul style="list-style-type: none"> <li>改ざん DNA の合成</li> <li>実験室への侵入</li> </ul>					
読出しデータの妨害	シーケンサーの脆弱性の悪用	シーケンサー	左記の作業担当	後処理用プログラム	マルウェア DNA の合成	後処理用プログラムの誤動作が誘発されうる	プログラムの脆弱性解消	実験室の入室管理の厳格化	
			—		<ul style="list-style-type: none"> <li>マルウェア DNA の合成</li> <li>実験室への侵入</li> </ul>				

例えば、①次世代シーケンサーの脆弱性を悪用してシーケンス解析結果の入手を試みる攻撃のうち、②不特定の DNA を対象とした攻撃への対策を講じる必要があると判断した場合を考える。候補となる対策手法として、バイナリ・データの暗号化、シーケンス解析の並列処理の禁止、ナノポア・シーケンサーの使用が考えられる。

バイナリ・データの暗号化については、バイナリ・データの漏洩を防ぐ効果が高い反面、暗号が危殆化した場合の対応や鍵管理を適切に行うことが必要となり、実装には相応のコストが必要となる。シーケンス解析の並列処理の禁止については、並列処理しないことにより、解析に要する時間が増大することになる。ナノポア・シーケンサーの使用については、次世代シーケンサーを使用する場合と比較して、本体価格を低く抑えることができる反面、並列処理できないことにより解析に要する時間が増大する可能性がある。これらの観点からコスト対効果を評価し、どの対策手法が最適かを決定する。

このほか、実験室への入退室管理の厳格化が対策手法の候補となる場合、少量であっても攻撃に悪用しうるといふ DNA の特徴を考慮して、具体的な管理方法を検討し、その効果を評価することが求められる。DNA 合成機へのアクセスの厳格化については、それに伴い追加的に発生するコスト（例えば、DNA 合成機を設置するための個室の設置費用）を評価する。また、属性情報の管理の厳重化は、特定の DNA を標的とする攻撃に対しては有効であるが、不特定の DNA を標的とする攻撃に対しては有効ではない点を踏まえる必要がある。

## 5. 結びに代えて：DNA ストレージへの期待と課題

近年、金融分野では、フィンテックをはじめ、最先端の情報技術を活用した金融サービスを開発し提供していこうとする流れが活発となっている。ブロックチェーンにおける分散台帳のデータや AI・機械学習における訓練データ等、今後、新しい金融サービスにおいて取り扱われるデータの量も一段と増加していく可能性が高いとみられる。DNA ストレージは、こうした金融分野における情報爆発時代にとって、中長期的に有望なデータ・ストレージ技術となりうる。

DNA ストレージは、バイオテクノロジーと情報技術という異なる技術分野が融合した技術であるといえる。2 節と 3 節において説明したように、現時点では発展途上の段階にあるが、DNA ストレージを実装したシステムが実際に開発・運用されるようになれば、取り扱われるデータのセキュリティを確保することが必要になることは明らかである。近年、情報システムのセキュリティを検討する際には、設計段階からセキュリティをどう確保し維持していくかを予め検討し必要な機構をシステムに組み込んでおく「セキュリティ・バイ・デザイン」の考え方が注目されている。DNA ストレージにおいても、こうしたアイデアに基

づき、両分野の観点からリスク分析を行いセキュリティ対策の方針等を検討しておくことが有用である。また、長期間にわたりデータを安全に保存していくうえでは、暗号技術の危殆化や、量子コンピュータによる脅威など、セキュリティにかかる最新の動向を注視したうえで、継続的にセキュリティ対策を更新していくことが求められる。

バイオテクノロジーと情報技術の融合は比較的新しい分野であり、研究課題も多い。これらの分野の融合によって新たにリスクが生じる可能性についても留意する必要がある。それぞれの分野の専門家が、互いの分野に対する知見を深めつつ、共に議論し研究を進めていくことが重要になる。DNA ストレージの今後の研究開発が進展することを期待したい。

以 上

## 【参考文献】

- 喜連川 優、「情報爆発のこれまでとこれから」、『電子情報通信学会誌』、94(8)、電子情報通信学会、2011年、662～666頁
- 佐野正和・石本健志、「ストレージの技術トレンド—ストレージはどこに向かうのか—」、『PROVISION』、No.77、日本アイ・ビー・エム、2013年、26～31頁
- 情報処理推進機構セキュリティセンター、「制御システムのセキュリティリスク分析ガイド 第2版 ～セキュリティ対策におけるリスクアセスメントの実施と活用～」、情報処理推進機構、2020年
- ソニー・ソニーストレージメディアソリューションズ、「業界最高の面記録密度 201Gbit/inch<sup>2</sup> を達成した磁気テープストレージ技術を開発」、ソニー、2017年 (<https://www.sony.co.jp/SonyInfo/News/Press/201708/17-070/>、2020年6月15日)
- 日本アイ・ビー・エム、「磁気テープ・ストレージで新記録を樹立、クラウド・ストレージにおける磁気テープの競争力を強化」、日本アイ・ビー・エム、2017年 (<https://jp.newsroom.ibm.com/announcements?item=122512>、2020年6月15日)
- 野地博行・田端和仁・景山茂樹・奥野大地・高久春雄、「ImPACT 野地プログラム調査報告書 長鎖 DNA 合成技術の進展と課題」、科学技術振興機構、2019年 (<https://www.jst.go.jp/impact/noji/progress/pdf/20190301.pdf>、2020年6月15日)
- 堀内義章、「2019年世界経済とストレージ・HDDの業界展望」、日本HDD協会、2019年 (<http://www.idema.gr.jp/common/pdf/news/tenbo2019.pdf>、2020年6月15日)
- Ailenberg, Menachem, and Ori D. Rotstein, “An Improved Huffman Coding Method for Archiving Text, Images, and Music Characters in DNA,” *BioTechniques*, 47(3), Future Science, 2009, pp. 747-754.
- Appuswamy, Raja, Kevin Lebrigand, Pascal Barbry, Marc Antonini, Oliver Madderson, Paul Freemont, James MacDonald, and Thomas Heinis, "OligoArchive: Using DNA in the DBMS Storage Hierarchy," paper presented at the Conference on Innovative Data Systems Research (CIDR) 2019, CIDR Conference, 2019 (available at <http://cidrdb.org/cidr2019/papers/p98-appuswamy-cidr19.pdf>、2020年6月15日)。
- Bishop, Bryan, Nathan Mccorkle, and Victor Zhirnov, “Summary Report: Technology Working Group Meeting on Future DNA Synthesis Technologies,” Semiconductor Research Corporation, 2017 (available at <https://www.src.org/program/grc/semisynbio/semisynbio-consortium-roadmap/6043->

- full-report-dna-based-storage-final-twg1-4-18.pdf、2020年6月15日)。
- Blawat, Meinolf, Klaus Gaedke, Ingo Hütter, Xiao-Ming Chen, Brian Turczyk, Samuel Inverso, Benjamin W. Pruitt, and George M. Church, “Forward Error Correction for DNA Data Storage,” *Procedia Computer Science*, 80, Elsevier, 2016, pp. 1011-1022.
- Brunker, Mike, “Microsoft and University of Washington Researchers Set Record for DNA Storage,” Microsoft, 2016 (available at <https://blogs.microsoft.com/ai/synthetic-dna-storage-milestone/>、2020年6月15日)。
- Church, George M., Yuan Gao, and Sriram Kosuri, “Next-Generation Digital Information Storage in DNA,” *Science*, 337(6102), 2012, p. 1628.
- Erlich, Yaniv, and Dina Zielinski, “DNA Fountain Enables a Robust and Efficient Storage Architecture,” *Science*, 355(6328), 2017, pp. 950-954.
- Faezi, Sina, Sujit Rokka Chhetri, Arnav Vaibhav Malawade, John Charles Chaput, William Grover, Philip Brisk, and Mohammad Abdullah Al Faruque, “Oligo-Snoop: A Non-Invasive Side Channel Attack against DNA Synthesis Machines,” paper presented at the Network and Distributed System Security (NDSS) Symposium 2019, Internet Society, 2019 (available at [https://www.ndss-symposium.org/wp-content/uploads/2019/02/ndss2019\\_05B-1\\_Faezi\\_paper.pdf](https://www.ndss-symposium.org/wp-content/uploads/2019/02/ndss2019_05B-1_Faezi_paper.pdf)、2020年6月15日)。
- Goldman, Nick, Paul Bertone, Siyuan Chen, Christophe Dessimoz, Emily M. LeProust, Botond Sipos, and Ewan Birney, “Towards Practical, High-Capacity, Low-Maintenance Information Storage in Synthesized DNA,” *Nature*, 494, 2013, pp. 77-80.
- Grass, Robert N., Reinhard Heckel, Michela Puddu, Daniela Paunescu, and Wendelin J. Stark, “Robust Chemical Preservation of Digital Information on DNA in Silica with Error-Correcting Codes,” *Angewandte Chemie*, 54(8), Wiley-VCH, 2015, pp. 2552-2555.
- Hayden, Erika Check, “Technology: The \$1,000 Genome,” *Nature*, 507, 2014, pp. 294-295.
- Islam, Mohd Siblee, Stepan Ivanov, Eric Robson, Triona Dooley-Cullinane, Lee Coffey, Kevin Doolin, and Sasitharan Balasubramaniam, “Genetic Similarity of Biological Samples to Counter Bio-Hacking of DNA-Sequencing Functionality,” *Scientific Reports*, 9:8684, Springer Nature, 2019 (available at <https://www.nature.com/articles/s41598-019-44995-6>、2020年6月15日)。

- Kar, Diptendu Mohan, and Indrajit Ray, “That’s My DNA: Detecting Malicious Tampering of Synthesized DNA,” *Proceedings of IFIP Conference on Data and Applications Security and Privacy (DBSec) 2019, Lecture Notes in Computer Science*, 11559, Springer-Verlag, 2019, pp. 61-80.
- , ———, Jenna Gallegos, and Jean Peccoud, “Digital Signatures to Ensure the Authenticity and Integrity of Synthetic DNA Molecules,” *Proceedings of the New Security Paradigms Workshop (NSPW) 2018*, Association for Computing Machinery, 2018, pp. 110-122.
- Langston, Jennifer, “With a “Hello,” Microsoft and UW Demonstrate First Fully Automated DNA Data Storage,” Microsoft, 2019 (available at <https://news.microsoft.com/innovation-stories/hello-data-dna-storage/>、2020年6月15日).
- Lopez, Randolph, Yuan-Jyue Chen, Siena Dumas Ang, Sergey Yekhanin, Konstantin Makarychev, Miklos Z. Racz, Georg Seelig, Karin Strauss, and Luis Ceze, “DNA Assembly for Nanopore Data Storage Readout,” *Nature Communications*, 10:2933, Springer Nature, 2019 (available at <https://www.nature.com/articles/s41467-019-10978-4>、2020年6月15日).
- Meyer, Matthias, Qiaomei Fu, Ayinuer Aximu-Petri, Isabelle Glocke, Birgit Nickel, Juan-Luis Arsuaga, Ignacio Martínez, Ana Gracia, José María Bermúdez de Castro, Eudald Carbonell, and Svante Pääbo, “A Mitochondrial Genome Sequence of a Hominin from Sima de los Huesos,” *Nature*, 505, 2014, pp. 403-406.
- Newman, Sharon, Ashley P. Stephenson, Max Willsey, Bichlien H. Nguyen, Christopher N. Takahashi, Karin Strauss, and Luis Ceze, “High Density DNA Data Storage Library via Dehydration with Digital Microfluidic Retrieval,” *Nature Communications*, 10:1706, Springer Nature, 2019 (available at <https://www.nature.com/articles/s41467-019-09517-y>、2020年6月15日).
- Ney, Peter, Karl Koscher, Lee Organick, Luis Ceze, and Tadayoshi Kohno, “Computer Security, Privacy, and DNA Sequencing: Compromising Computers with Synthesized DNA, Privacy Leaks, and More,” *Proceedings of the USENIX Conference on Security Symposium (SEC) 2017*, USENIX Association, 2017, pp. 765-779.
- Organick, Lee, Siena Dumas Ang, Yuan-Jyue Chen, Randolph Lopez, Sergey Yekhanin, Konstantin Makarychev, Miklos Z. Racz, Govinda Kamath, Parikshit Gopalan, Bichlien H. Nguyen, Christopher N. Takahashi, Sharon Newman, Hsing-Yeh Parker,

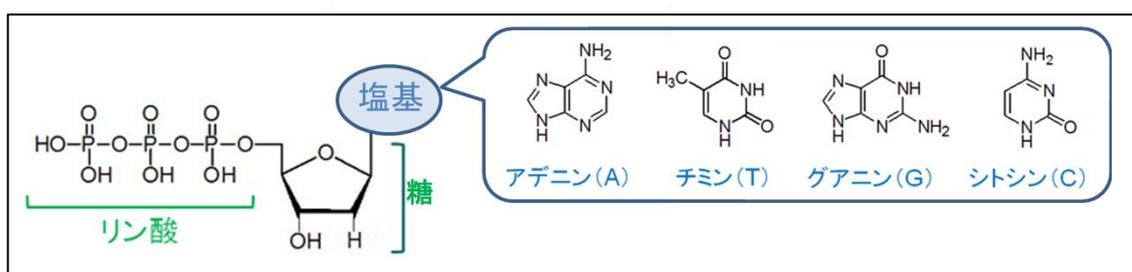
- Cyrus Rashtchian, Kendall Stewart, Gagan Gupta, Robert Carlson, John Mulligan, Douglas Carmean, Georg Seelig, Luis Ceze, and Karin Strauss, “Random Access in Large-Scale DNA Data Storage,” *Nature Biotechnology*, 36, Springer Nature, 2018, pp. 242-248.
- Pray, Leslie A., “Discovery of DNA Structure and Function: Watson and Crick,” *Nature Education*, 1(1):100, Nature Education, 2008 (available at <http://www.nature.com/scitable/topicpage/discovery-of-dna-structure-and-function-watson-397/>, 2020年6月15日).
- Regalado, Antonio, “Microsoft Has a Plan to Add DNA Data Storage to its Cloud,” MIT Technology Review, Massachusetts Institute of Technology, 2017 (available at <https://www.technologyreview.com/2017/05/22/68387/microsoft-has-a-plan-to-add-dna-data-storage-to-its-cloud/>, 2020年6月15日).
- Reinsel, David, John Gantz, and John Rydning, “Data Age 2025: The Digitization of the World From Edge to Core,” International Data Corporation, 2018 (available at <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>, 2020年6月15日).
- Takahashi, Christopher N., Bichlien H. Nguyen, Karin Strauss, and Luis Ceze, “Demonstration of End-to-End Automation of DNA Data Storage,” *Scientific Reports*, 9:4998, Springer Nature, 2019 (available at <https://www.nature.com/articles/s41598-019-41228-8>, 2020年6月15日).
- Wetterstrand, Kris A., “DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP),” National Human Genome Research Institute, 2019 (available at <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>, 2020年6月15日).
- Wong, Pak Chung, Kwong-Kwok Wong, and Harlan Foote, “Organic Data Memory Using the DNA Approach,” *Communications of the ACM*, 46(1), Association for Computing Machinery, 2003, pp. 95-98.
- Yazdi, Hossein Tabatabaei, Yongbo Yuan, Jian Ma, Huimin Zhao, and Olgica Milenkovic, “A Rewritable, Random-Access DNA-Based Storage System,” *Scientific Reports*, 5:14138, Springer Nature, 2015 (available at <https://www.nature.com/articles/srep14138>, 2020年6月15日).

## 補論 1. DNA の役割と基本機能

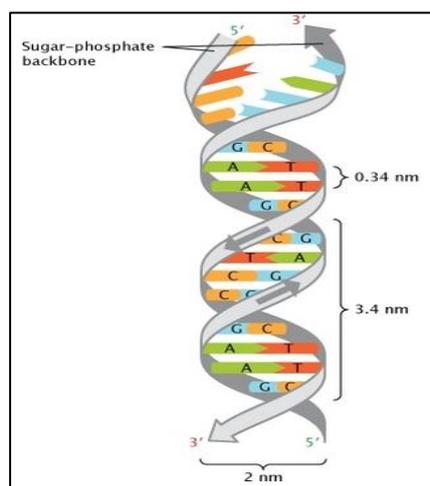
DNA は、生物の生体形成や生命活動の維持に必要となるたんぱく質の構造に関する情報を記録した化学物質である。1 つの個体を形成するすべての細胞は、その個体に固有かつ共通の DNA を有しており、DNA に記録された情報に基づき、随時必要なたんぱく質を発現している。そのため、DNA は「生命の設計図」ともいわれる<sup>42</sup>。

化学的な構造をみると、DNA はヌクレオチドと呼ばれる構成単位が連なった高分子である。ヌクレオチドは、糖、リン酸、塩基から構成され、塩基は、アデニン (A)、チミン (T)、グアニン (G)、シトシン (C) の 4 種類存在する (図表 A-1)。このうち「A」と「T」、「G」と「C」はペアとなって結合する。そのため、細胞内では、互いに対をなす塩基配列からなる 2 本の DNA が二重らせん構造をとっている (図表 A-2)<sup>43</sup>。

図表 A-1. DNA の構成単位 (ヌクレオチド) の構造



図表 A-2. DNA の二重らせん構造



資料 : Pray [2008] の Figure.3

<sup>42</sup> 人工的に合成された DNA は、それ単体で「生命の設計図」として機能することはない。

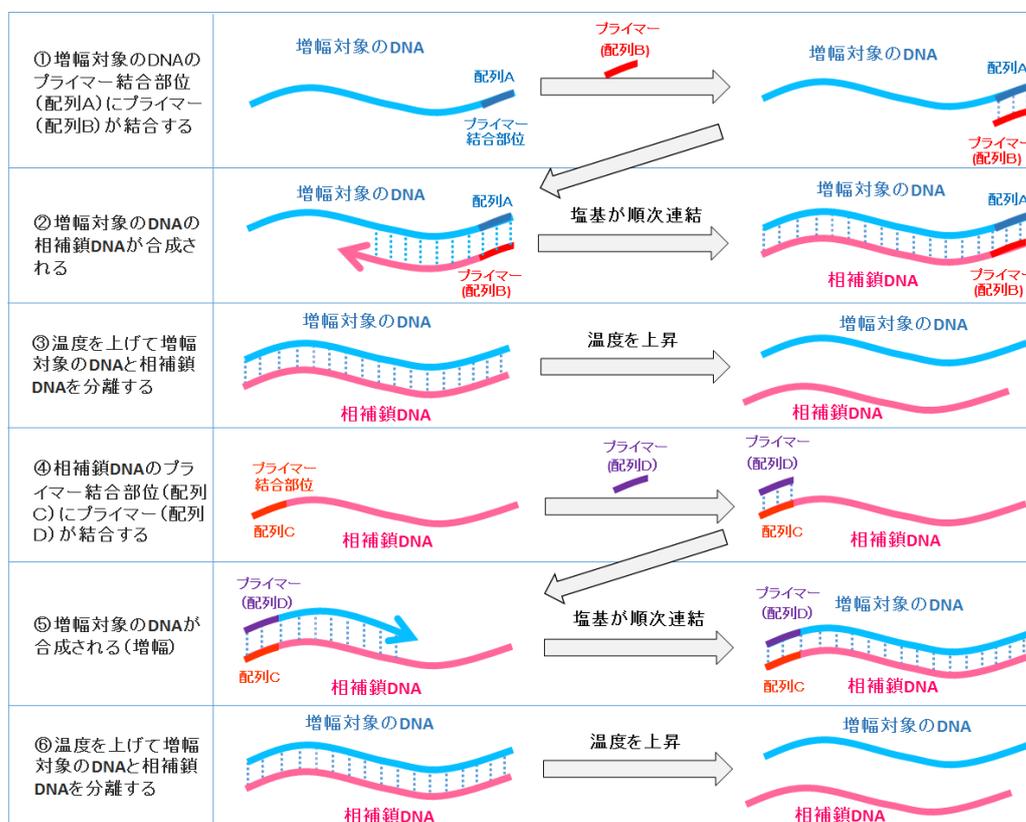
<sup>43</sup> DNA の二重らせん構造は、1953 年、分子模型を構築する手法を用いてジェームズ・ワトソンとフランシス・クリックによって提唱された。

## 補論 2. DNA の増幅

DNA の増幅におけるプロセスは以下のとおりである。それぞれのプロセスに付した番号は、図表 A-3 中の番号に対応する。

- ① 増幅対象の DNA の末端（プライマー結合部位〈配列 A〉）に、ペアとなる塩基配列を有するプライマー（配列 B）が結合する。
- ② PCR により、増幅対象の DNA のペアとなる塩基がプライマーに順次連結され、増幅対象の DNA とペアになる塩基配列（片側が A なら反対側は T、片側が G なら反対側は C）を有する DNA（相補鎖〈そうほさ〉DNA）が合成される。
- ③ 温度を上げて増幅対象の DNA と相補鎖 DNA を分離する。
- ④ 相補鎖 DNA の末端（プライマー結合部位〈配列 C〉）に、ペアとなる塩基配列を有するプライマー（配列 D）が結合する。
- ⑤ PCR により、相補鎖 DNA のペアとなる塩基がプライマーに順次連結され、相補鎖 DNA とペアになる塩基配列を有する DNA（増幅対象の DNA）が合成される。
- ⑥ 温度を上げて増幅対象の DNA と相補鎖 DNA を分離する。
- ⑦ 増幅対象の DNA が必要な量に達するまで、①から⑥のプロセスを繰り返す。

図表 A-3. DNA の増幅プロセス



### 補論 3. ナノポア・シーケンサーを用いた効率的な読出し方法に関する研究

ナノポア・シーケンサーは、複数の DNA を並列して解析することができないという欠点がある。例えば、ある長さのデータを 1 つの DNA に記録した場合と、複数の DNA に分割して記録した場合（補論 4 のランダム・アクセス機能を用いるケース等）を比較すると、原理的に、読出しに要する時間は後者の方が長くなる。これは、シーケンス解析では、解析の都度、前後の処理に相応の手間と時間を要し、複数の DNA に分割した場合、読出しの回数だけ、前後の処理が追加的に発生するためである。

そこで、複数の DNA にまたがる連続したデータを効率的に読み出す手法として、複数の DNA が混ざり合っているときに、読み出したいデータを記録した DNA だけを連結して 1 つの DNA としたうえでシーケンス解析を行う手法が提案されている（Lopez *et al.* [2019]）。本手法を用いることで、選択した DNA を連結する処理が追加的に発生するものの、すべての DNA のデータを読み出す処理を実行するよりは、全体として効率的な読出しが可能となる<sup>44</sup>。

### 補論 4. ランダム・アクセス機能

DNA ストレージを実用化するうえでは、記録されたデータから、真に必要なデータのみを読み出す機能を実装することが求められる。代表的な機能として、読み出したいデータが記録されている（情報記録媒体上の）箇所を特定したうえで、それを読み出すランダム・アクセス機能がある<sup>45</sup>。これを実現する手法として、次のような手順が提案されている（Yazdi *et al.* [2015]、Organick *et al.* [2018]、図表 A-4）。

#### 【書込みプロセス】

- ① 記録したいバイナリ・データを、複数のセグメント（例えば、書籍であれば章ごと）に分割する。
- ② 各セグメントのバイナリ・データをそれぞれ塩基配列に変換する。
- ③ それぞれのセグメントの両端に、データ検索用の固有の配列をデータ検索用タグとして付加する。各データ検索用タグがどのセグメントに付加されているかを表す情報を記録しておく。

<sup>44</sup> ここで複数の DNA から 1 つの DNA を生成する際に用いられる連結手法はギブソン・アセンブリ（Gibson Assembly）と呼ばれ、オリゴ DNA から長鎖 DNA を生成する際に用いられる最も一般的な手法の 1 つである。ギブソン・アセンブリでは、複数の DNA が PCR を用いて連結され、1 つの DNA となる。

<sup>45</sup> ランダム・アクセス機能が利用できない場合、一部のデータのみを読み出したい場合にも、記録したすべてのデータを読み出したうえで、読み出したいデータを検索することになる。

図表 A-4. ランダム・アクセス機能を実現するための書き込みプロセス



- ④ データ検索用タグが付加された各セグメントの塩基配列に対応する DNA をそれぞれ合成する。
- ⑤ 合成した DNA は 1 つの容器に混合して保存する。

#### 【読み出しプロセス】

- ⑥ 読み出したいデータのセグメントに付加されているデータ検索用タグを特定する。
- ⑦ データ検索用タグの配列からなるプライマーを合成する。
- ⑧ ⑤で混合した DNA に、⑦で合成したプライマーを混合して増幅にかかるプロセスを開始する。これにより、読み出したいデータのセグメントが含まれる DNA のみが増幅される。
- ⑨ ⑧で増幅した DNA をシーケンス解析し、塩基配列を得る。
- ⑩ 塩基配列をバイナリ・データに変換する。

なお、データの内容によっては、表形式のデータベース(リレーショナル・データベース)で管理することにより、データへのアクセスの利便性が向上する場合がある<sup>46</sup>。DNA ストレージでは、表形式のデータをセル群ごとに分割して、それ

<sup>46</sup> リレーショナル・データベースは、データを複数の表として管理し、表と表間の関係を定義することで、複雑なデータを関連付けて記録できるデータベースである。例えば、ある商店にお

ぞれに固有のタグを付加することによって各セル群を特定可能とし、標準化されたデータベース言語である SQL を用いて目的のデータを読み出すことに成功したとの研究結果が示されている (Appuswamy *et al.* [2019])。

## 補論 5. 改ざんプライマーを用いた DNA の書換え手順

DNA 増幅時に改ざんプライマーを用いることにより、標的 DNA の塩基配列の一部を置換したり、新たな配列を挿入したりすること (標的 DNA の書換え) が可能なケースがある。ここでは、標的 DNA の 3 塩基を異なる配列 (「ATG」を「CAT」) に書き換える手順を説明する。それぞれのプロセスに付した番号は、図表 A-5 中の番号に対応する。

### ① 2つの改ざんプライマー (A)と(B)を設計して合成する。

＜改ざんプライマー(A)の設計手順＞

- i. 書き換えたい内容を反映した DNA (改ざん DNA) の配列を決定する。
- ii. 標的 DNA のプライマー結合部位と書換え箇所を含む領域を選択する。
- iii. 上記 ii において選択した領域に対応する、改ざん DNA の領域を選択する。
- iv. 上記 iii において選択した領域の塩基配列に対してペアになる塩基配列を、改ざんプライマー(A)の塩基配列とする<sup>47</sup>。

＜改ざんプライマー(B)の設計手順＞

- v. 改ざん DNA の塩基配列のうち、標的 DNA のプライマー結合部位と反対側の末端部位の塩基配列を、改ざんプライマー(B)の塩基配列とする。

② 標的 DNA、改ざんプライマー(A)、改ざんプライマー(B)を混合し、DNA の増幅にかかるプロセスを開始する。まず、標的 DNA に改ざんプライマー(A)が結合する<sup>48</sup>。それを契機に、PCR により標的 DNA のペアとなる塩基が順次連結され、改ざん DNA とペアになる塩基配列からなる DNA (相補鎖改ざん DNA) が合成される。

③ 温度を上げて標的 DNA と相補鎖改ざん DNA を分離する。

④ 相補鎖改ざん DNA に改ざんプライマー(B)が結合する。それを契機に、PCR

---

いて、販売した商品とその価格を整理した売上管理表と、商品ごとの製造元メーカーを整理した商品情報管理表をリレーショナル・データベース上で管理すると、両者のデータを統合することにより、製造元メーカー別の売上高を検索することができる。

<sup>47</sup> 書換え箇所と標的 DNA のプライマー結合部位が離れている場合には、改ざんプライマー(A)は相応に長くなる。

<sup>48</sup> 標的 DNA と改ざんプライマー(A)の書換え箇所の塩基配列は相補的でない。もっとも、書換え箇所の塩基数が、改ざんプライマー(A)全体の塩基数と比較して十分に小さい場合、書換え箇所以外の塩基が標的 DNA の塩基と対をなすことにより、改ざんプライマー(A)は標的 DNA に結合することができる。

により相補鎖改ざん DNA のペアとなる塩基が順次連結され、改ざん DNA が合成される。

- ⑤ 温度を上げて相補鎖改ざん DNA と改ざん DNA を分離する。
- ⑥ 改ざん DNA に改ざんプライマー(A)が結合する。それを契機に、PCR により改ざんプライマーのペアとなる塩基が順次連結され、相補鎖改ざん DNA が

図表 A-5. 改ざんプライマーを用いた DNA の書換えプロセス

<p>i. 改ざんDNAの配列を決定</p>	<p style="text-align: center;">書換え箇所</p> <p>標的DNAの配列 <span style="border: 1px solid black; padding: 2px;">C···T·····GC···ATG·····AT</span></p> <p style="text-align: center;">↓</p> <p>改ざんDNAの配列 <span style="border: 1px solid black; padding: 2px;">C···T·····GC···CAT·····AT</span></p>
<p>ii. 標的DNAのプライマー結合部位と書換え箇所を含む領域を選択する</p>	<p style="text-align: center;">書換え箇所      標的DNAのプライマー結合部位</p> <p>標的DNAの配列 <span style="border: 1px solid black; padding: 2px;">C···T·····GC···ATG·····AT</span></p>
<p>① iii. 上記 ii において選択した領域に対応する改ざんDNAの領域を選択</p>	<p>標的DNAの配列 <span style="border: 1px solid black; padding: 2px;">C···T·····GC···ATG·····AT</span></p> <p style="text-align: center;">↓</p> <p>改ざんDNAの配列 <span style="border: 1px solid black; padding: 2px;">C···T·····GC···CAT·····AT</span></p>
<p>iv. 改ざんプライマー(A)の配列を決定</p>	<p>改ざんDNAの配列 <span style="border: 1px solid black; padding: 2px;">C···T·····GC···CAT·····AT</span></p> <p style="text-align: center;">ペアとなる(相補的な)配列</p> <p style="border: 1px solid black; padding: 2px; text-align: center;">CG···GTA·····TA</p> <p style="text-align: center;">改ざんプライマー(A)の配列</p>
<p>v. 改ざんプライマー(B)の配列を決定</p>	<p>改ざんDNAの配列 <span style="border: 1px solid black; padding: 2px;">C···T·····GC···CAT·····AT</span></p> <p style="text-align: center;">改ざんプライマー(B)の配列</p>
<p>② 標的DNAに改ざんプライマー(A)が結合し、相補鎖改ざんDNAが合成される。</p>	<p>標的DNA</p> <p style="text-align: center;">← 改ざんプライマー(A)</p> <p style="text-align: center;">塩基が順次連結</p> <p>相補鎖改ざんDNA      改ざんプライマー(A)</p>
<p>③ 温度を上げて標的DNAと相補鎖改ざんDNAを分離する。</p>	<p>標的DNA</p> <p>相補鎖改ざんDNA</p> <p style="text-align: center;">温度を上昇</p> <p>標的DNA</p> <p>相補鎖改ざんDNA</p>
<p>④ 相補鎖改ざんDNAに改ざんプライマー(B)が結合し、改ざんDNAが合成される。</p>	<p>改ざんプライマー(B)</p> <p style="text-align: center;">→ 相補鎖改ざんDNA</p> <p style="text-align: center;">塩基が順次連結</p> <p>改ざんDNA      相補鎖改ざんDNA</p>
<p>⑤ 温度を上げて改ざんDNAと相補鎖改ざんDNAを分離する。</p>	<p>改ざんDNA</p> <p>相補鎖改ざんDNA</p> <p style="text-align: center;">温度を上昇</p> <p>改ざんDNA</p> <p>相補鎖改ざんDNA</p>
<p>⑥ 改ざんDNAに改ざんプライマー(A)が結合し、相補鎖改ざんDNAが増幅される。</p>	<p>改ざんDNA</p> <p style="text-align: center;">← 改ざんプライマー(A)</p> <p style="text-align: center;">塩基が順次連結</p> <p>相補鎖改ざんDNA      改ざんプライマー(A)</p>
<p>⑦ 温度を上げて改ざんDNAと相補鎖改ざんDNAを分離する。</p>	<p>改ざんDNA</p> <p>相補鎖改ざんDNA</p> <p style="text-align: center;">温度を上昇</p> <p>改ざんDNA</p> <p>相補鎖改ざんDNA</p>

増幅される。

- ⑦ 温度を上げて改ざん DNA と相補鎖改ざん DNA を分離する。
- ⑧ ④と⑤および⑥と⑦をそれぞれ繰り返すことにより、改ざん DNA（および相補鎖改ざん DNA）を増幅する。