

令和 3 年 6 月 17 日現在

機関番号：32658

研究種目：若手研究

研究期間：2019～2020

課題番号：19K16113

研究課題名（和文）大規模マイクロサテライトデータベースと自動解析処理システムの構築

研究課題名（英文）The construction of large-scale microsatellite database and bioinformatics pipeline system

研究代表者

田中 啓介（TANAKA, Keisuke）

東京農業大学・その他部局等・助教

研究者番号：60747294

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：SNSなどを通じてマイクロサテライト情報の需要がある生物種に対するDNAサンプルの募集を行い、目標の500種を達成することができた。また、マイクロサテライト検出を行うために本研究代表者が考案した実験系では、制限酵素とフォークドアダプターを用いるプロトコルを取り入れることで、従来よりも効率的かつ半分のコストでデータを得ることに成功した。そのほか、公共データベースから配列データを取得し、そこからマイクロサテライト情報の検出を行い、データベースの拡充化を図った。コンピュータ利用による多型解析では、データベース上の配列と相同性検索ができるツールを実装した。

研究成果の学術的意義や社会的意義

これまで品種改良が遅れていた作物・園芸品種に対してDNAマーカー開発が容易となり、マーカー選抜による育種効率の向上や、現場証拠として鑑定が必要な植物個体の識別技術の発展につながることで、さらにDNAマーカー開発への適用は、生態保全や遺伝資源の担保につながられることにも期待される。また、本研究代表者が所属する東京農業大学 生物資源ゲノム解析センターが掲げる農学支援としても非常に有用なゲノムリソースになるといえる。

研究成果の概要（英文）：This study was able to achieve the target of 500 samples via recruitment of genomic DNA for organisms that are in demand for microsatellite information through such as researcher community and SNS. In addition, it succeeded in obtaining data more efficiently and at half the cost than before by adopting a protocol using restriction enzymes and forked adapters in the experimental system of microsatellite detection. It also acquired sequence data from some public databases and detected microsatellite information. On the other hands, a tool that perform homology search with sequences on the database was implemented as computational polymorphism analysis.

研究分野：ゲノム生物学

キーワード：マイクロサテライト 次世代シーケンサー DNAマーカー データベース 解析パイプライン NGS SSR

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

DNA 多型解析は、この 40 年間で様々なアプローチが展開されてきた。中でも最近では、一塩基多型 (SNP) と単純反復配列 (マイクロサテライト、SSR) 多型を利用した研究例が非常に多くみられる。DNA 多型解析を行う研究者にとって、SNP と SSR のどちらを利用するか、目的に応じて選択肢が与えられることは非常に重要である。一般的に SNP は、塩基置換頻度を進化速度として組み込むことができるため、集団遺伝解析や分子系統解析などに効果が発揮される。一方で、作物や畜産動物などの品種識別用の DNA マーカー開発や法科学における個体識別を目的とした場合、SNP 情報ではしばしば精度が低くなることから、マイクロサテライトが利用されることが多い。近年、次世代シーケンサー (NGS) を用いた DNA 多型解析は、様々な分野で利用されるようになった。中でも、1 サンプルの解析コストやデータ量を抑えて大規模多型解析を可能とする RAD-seq や GBS のような SNP ベースの NGS 解析の実施例は年々増加している。ところが、マイクロサテライトベースの NGS 解析は、SNP ベースに比べると、あまり実用化されていない傾向にある。そこで、NGS を用いた DNA 多型解析において、大規模な SNP 検出と共に、大規模なマイクロサテライト検出を行う技術も利用可能になれば、研究者は目的に合わせた解析を選択できるようになり、DNA マーカー開発技術の飛躍的な向上と、さらに高精度な個体識別技術の発展にもつなげられるのではないだろうか。

2. 研究の目的

従来は、コロニーハイブリダイゼーション法や dual suppression PCR 法などによってマイクロサテライト検出されてきたが、いずれも反復領域を含む配列の選抜やキャピラリーシーケンサーを用いることから、時間・コスト・労力がかかることに問題があった。そのような経緯から、本研究代表者は、断片化した DNA に対して SSR プローブで濃縮し、それを NGS でターゲットキャプチャーシーケンスする Microsatellite capture sequencing 法 (MiCAPs) を開発した (Tanaka et al. 2017, 2018, 図 1)。この技術を用いることによって、従来の方法よりも手間をかけず大量かつ安価にマイクロサテライト検出することができるようになった。そこで、本研究課題は、この解析技術の実用化を図り、様々な生物に対して MiCAPs を用いて大量のマイクロサテライト情報を取得し、マイクロサテライトデータベースの構築を目指す。さらに、MiCAPs の Wet 解析だけでなく Dry 解析もまた、誰もが手軽に利用できるようにするために、解析アプリケーションを設計・構築することを目的とする。本研究では、データベースに含める対象生物種の一部を本研究者の所属学会や国内の研究者向けのソーシャルネットワークサービス (SNS) を利用して募集を行い、サンプルを取り入れる。研究者コミュニティを利用することにより、我が国の固有種や品種が得られやすくなり、研究に意義を成す生物種が含まれた国内の研究者に利用しやすいデータベースが完成されると考えられる。また、本研究側にとっても、サンプル収集を迅速化できるので、本研究目的の達成に近づけやすくなるのが期待できる。

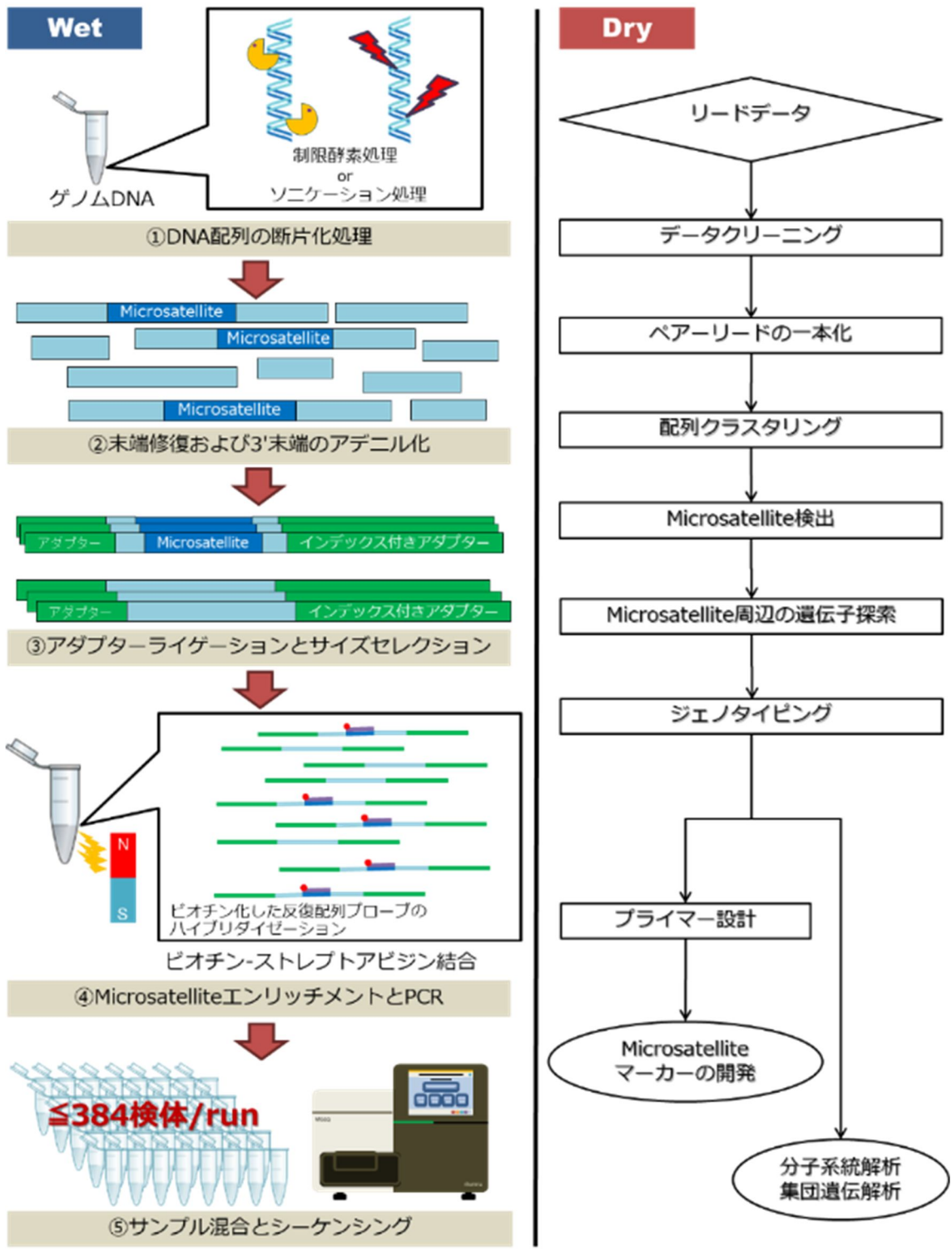


図1 Microsatellite capture sequencing法 (MiCAPs) の概要

3. 研究の方法

本研究の目的を達成するために、500種の生物から大規模なマイクロサテライト情報を検出してマイクロサテライトデータベースを構築すること、MiCAPsを誰もが手軽に利用できるようにするための解析アプリケーションを設計・構築すること、これら2点を実施した。具体的には以下の通りである。

まず、マイクロサテライトデータベースの構築では、系統識別や個体識別が必要とされる分野の利用者にとって有益なデータベースになることを目指した。特に農学分野向けには、非モデル系生物が多く、DNAマーカーの開発があまり進められていない農業・林業・畜産品種を用いた。さらに、本研究者の所属学会や国内の研究者向けのSNSを介して取得した生物種を用いた。本研究課題では、500種を目標としてサンプル収集を行い、それらをMiCAPsによって大規模なマイクロサテライト情報を検出し、マイクロサテライトデータベースを構築した。そして、データベースには配列情報だけではなく、モチーフやマイクロサテライト近傍の遺伝子アノテーション情報なども付加することによって充実化を図った。

続いて、解析アプリケーションの設計・構築では、MiCAPsによって得られたマイクロサテライト情報から、データ上で比較する系統・個体間で多型検出する*in silico*多型解析法の開発を試みた。この方法では、まずマイクロサテライトデータベースもしくは手持ちのマイクロサテライトデータをリファレンスデータとする。そして、NGSによって得られたシーケンスデータをインプットとしてリファレンスにマッピングし、そこからコンセンサス配列を作成する。そして、本研究代表者が作成したスクリプトをもとに各系統・個体のマイクロサテライトを検出することで、比較する系統・個体間の多型情報が与えられるというプログラムである。また、MiCAPsのDry解析部分にあたるマイクロサテライト検出のための解析パイプラインを組み立て、利用者にとって手軽に扱えるツールの構築を目指した。

以上の内容により、マイクロサテライトベースのNGS解析の実用化が果たされ、MiCAPsが多くの研究者にとって有用なツールになると期待される。

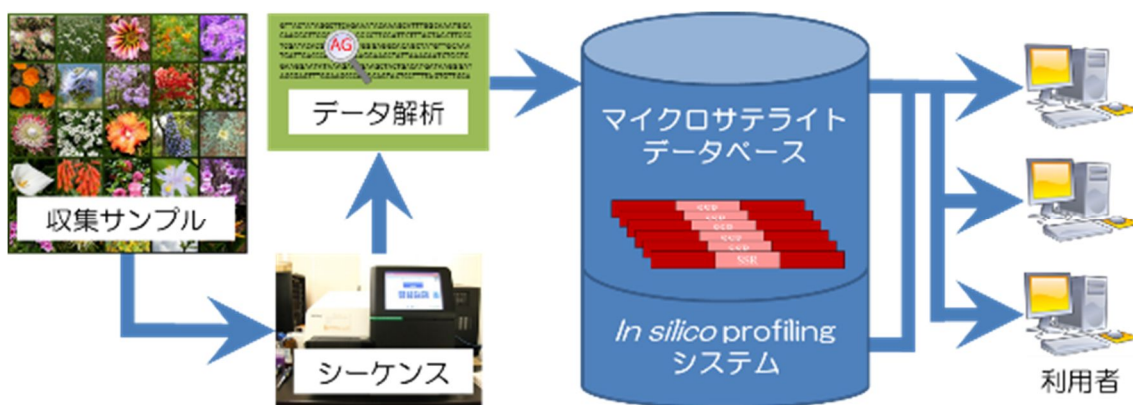


図2 本研究の構想

4. 研究成果

(1) 初年度

我が国の研究者にとって有益なマイクロサテライトデータベースの構築を目指すために、本研究者の所属学会や研究者間コミュニティ、知人を介した周知により、マイクロサテライト情報の必要がある生物種に対するDNAサンプルの募集を行った。サンプル提供者には、本研究の趣旨に対する理解を得た上で、提供サンプルから得られたデータはデータベース公開前であっても提供者が利用できるような互いの利益になるよう進めた。現状では、目標のおおよそ半数の個体を集めることができた。これらは順次、本研究代表者が考案した実験系を適用し、NGSを用いて配列データを出力した。さらに、これらデータを公開するためのウェブページ制作を行い、カタログ形式によるデータベースの基盤をつくった。

*in silico*多型解析の構想は、シーケンスによって得られたデータから対象生物種の系統間や個体間の多型性を比較することである。そのために、まずは個々のマイクロサテライト領域に対するプロファイル情報が必要となる。NGSによって得られた配列データからマイクロサテライト情報を得るために、データクリーニング、ペア配列の一本化、重複配列の除去、マイクロサテライト領域の識別が行われる。今年度は、これら一連の工程をパイプライン化した。

(2) 最終年度

昨年度に続き、研究者間コミュニティ、SNS を利用した周知により、マイクロサテライト情報の需要がある生物種に対する DNA サンプルの募集を行い、目標の 500 種を達成することができた。また、MiCAPs の実験系では、制限酵素とフォークドアダプターを用いる改良 RAD-seq 法のプロトコル (Ando *et al.* 2018) を取り入れることで、ソニケーション法と制限酵素法として異なるプロトコルを与えることができるようになった。さらに、バーコーディングしたサンプルをプリーングする際のクオリティーチェックを省略する迅速法を付属させることもできた。そのため、従来よりも効率的かつ半分のコストでデータを得ることに成功した。そのほか、公共データベースから配列データを取得し、そこからマイクロサテライト情報の検出を行い、データベースの拡充化を図った。*in silico* 多型解析では、データベース上の配列と相同性検索ができる BLAST 解析ツールを実装した。一方、MiCAPs のデータ解析に当たる一連のパイプラインは、昨年度に作成することができたが、データベースを公開するために使用するサーバーの能力が十分ではなかったため、データベース上の 1 つのツールとして実装することは難しかった。この点については、パイプラインをユーザー側の端末で解析できるプログラムを配布できるか今後の研究展開として検討していきたい。本研究において構築したデータベースは、「MiCAPs DB」として以下の URL 先に公開する。

<http://www.nodai-genome.org/MiCAPs/>

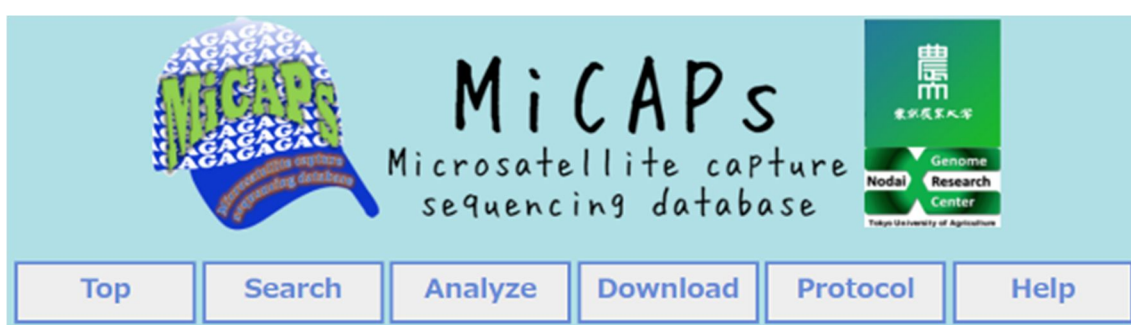


図3 大規模マイクロサテライトデータベース「MiCAPs DB」のトップページ

これまで、マイクロサテライト情報に関するデータベースは、海外のいくつかの研究グループによって公開されてきた。しかしながら、登録されている生物種の多くは、われわれ日本人研究者にとって身近なものではなく、利用しにくい状態にあった。そこで本研究は、マイクロサテライト検出の効率化に成功した MiCAPs を用いて、日本国内の研究者向けにデザインしたデータベースがつかれないか検討した。また、本研究代表者が開発した MiCAPs と同じような方法は、いくつか報告例がある。しかし、特定の制限酵素を用いて DNA 断片化したものからターゲットキャプチャーシーケンスを行う方法、*in silico* 多型解析法は、本研究代表者の提案によるものであり、この方法を使えばホモ接合とヘテロ接合を明瞭に識別することができるので、多型検出力が高いことが示されている。本研究が構築したマイクロサテライトデータベースが活かされれば、例えばこれまで品種改良が遅れていた作物・園芸品種に対して DNA マーカー開発が容易となり、マーカー選抜による育種効率の向上や、現場証拠として鑑定が必要な植物個体の識別技術の発展につながることで、さらに DNA マーカー開発への適用は、生態保全や遺伝資源の担保につながることも期待される。また、本研究代表者が所属する東京農業大学 生物資源ゲノム解析センターが掲げる農学支援としても非常に有用なゲノムリソースになるといえる。

【引用文献】

Tanaka K, Ohtake R, Yoshida S, Shinohara T. Effective DNA fragmentation technique for simple sequence repeat detection with a microsatellite-enriched library and high-throughput sequencing. *Biotechniques* 62: pp180-182. (2017)

Tanaka K, Ohtake R, Yoshida S, Shinohara T. [Capter 2] Microsatellite Capture Sequencing. In: *Genotyping*. IntechOpen: pp13-30. (2018)

Ando T, Matsuda T, Goto K, Hara K, Ito A, Hirata J, Yatomi J, Kajitani R, Okuno M, Yamaguchi K, Kobayashi M, Takano T, Minakuchi Y, Seki M, Suzuki Y, Yano K, Itoh T, Shigenobu S, Toyoda A, Niimi T. Repeated inversions within a pannier intron drive diversification of intraspecific colour patterns of ladybird beetles. *Nat Commun.* 9(1): 3843. (2018)

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 0件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 田中啓介	4. 巻 4
2. 論文標題 大規模マイクロサテライトデータベース” MiCAPs DB ” の構築に向けて	5. 発行年 2021年
3. 雑誌名 アグリバイオ	6. 最初と最後の頁 70-75
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計1件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 田中啓介
2. 発表標題 Microsatellite capture sequencing 法を用いた in silico 多型解析パイプラインの構築
3. 学会等名 日本農芸化学会2020年度大会(福岡)
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------