プレスリリース

2022 年 10 月 11 日
農　研　機　構
宇　都　宮　大　学
京　都　大　学

野外の生物集団の遺伝子頻度を効率よく推定する統計モデルを開発

－複数個体を一括して抽出したサンプルにおける DNA 量の個体差に対処する－

農研機構と宇都宮大学、京都大学は、複数の生物個体から一括で抽出された DNA サンプルに含まれる各個体由来の DNA 量のばらつきを確率として表現することで、野外の生物集団における対立遺伝子の頻度を推定する統計モデルを開発しました。本成果を定量 PCR や量的 DNA シーケンシング等の遺伝子診断技術に適用することで、個体別診断よりも少ない検査回数で、薬剤抵抗性害虫や外来種の蔓延状況を高精度に把握できます。

野外の生物集団において、在来生物種と交雑の恐れがある外来種や、害虫の殺虫剤抵抗性系統がどの程度の比率で存在するかを把握するために、各個体から DNA を抽出して PCR 検査等で遺伝子型を決定する、個体別遺伝子診断が従来行われてきました。しかし、比率を知りたい対立遺伝子（アリル [1]）をもつ個体が集団中に稀にしか存在しない場合は、推定精度を保つために数十～数百個体以上を診断する必要がありました。そこで実験操作の回数を減らすため、数個体ずつまとめて DNA 抽出された「バルクサンプル（混合 DNA 溶液）」による診断で、対立遺伝子頻度を推定する方法が模索されてきました。

バルクサンプルの DNA 含有量や、対立遺伝子ごとの含有比は、定量 PCR[2]や量的 DNA シーケンシング [3]等の手法によって測定できます。もし各個体が同じ量の DNA を持つならば、バルクサンプル中の DNA の比率は、そのままバルクサンプルに含まれる個体の遺伝子構成を示します。しかし現実には、各個体の体を構成する細胞の数によって DNA 量が異なることに加え、DNA 量は死後の分解によっても減少します。そのためトラップで捕獲した個体からバルクサンプルを作って DNA 量の比を測定すると、野外の個体の存在比から大きくずれる場合があります。

農研機構、宇都宮大学、京都大学の研究グループは、各個体から得られる DNA 量のばらつきを「ガンマ分布」という確率分布で近似することにより、生物集団における対立遺伝子の比率を、その推定値がどの程度確からしいかの指標（信頼区間）とともに推定できる統計モデルを開発しました。複数のバルクサンプルが用意され、その各々が何個体から構成されるかが分かっていれば、適用が可能です。

本モデルを定量 PCR 解析に適用して、対立遺伝子の比率とその信頼区間を簡便に求められるようにするべく、フリーの統計解析環境である R のためのパッケージ “freqpcr” を開発し公式サイトで配布しています（https://cran.r-project.org/package=freqpcr）。すでに本パッケージは、ミカンハダニにおける殺ダニ剤抵抗性遺伝子の地域分布パターンの解析をはじめ、野外で稀な遺伝子の存在比率をより少ない検査回数で高精度に推定する目的で活用されています。また農業害虫のみならず、希少生物種の保全や外来種・系統の侵入警戒を目的としたモニタリングなどにも役立ちます。

＜関連情報＞

予算:農林水産省委託プロジェクト「ゲノム情報等を活用した薬剤抵抗性管理技術の開発」

問い合わせ先など

研究推進責任者：農研機構　植物防疫研究部門　所長　眞岡　哲夫
研 究 担 当 者 ：同 果樹茶病害虫研究領域　主任研究員　須藤 正彬
TEL 0547-45-4101
農研機構　農業環境研究部門 土壌環境管理研究領域　山村 光司
（前　農研機構農業環境変動研究センター　統計モデル解析ユニット長）
農研機構　農業情報研究センター 確率モデルユニット長　山中 武彦
宇都宮大学農学部 生物資源科学科　教授　園田 昌司
京都大学大学院農学研究科 地域環境科学専攻　准教授　刑部 正博
広 報 担 当 者 ：農研機構　植物防疫研究部門 渉外チーム長　松下　陽介
TEL 029-838-6876　プレス用 e-mail IPP-Koho@naro.affrc.go.jp

本資料は農政クラブ、農林記者会、農業技術クラブ、文部科学記者会、科学記者会、栃木県政記者クラブ、筑波研究学園都市記者会、京都大学記者クラブに配付しています。

開発の社会的背景

遺伝子診断は、生物を扱う現代の学問および産業の基盤をなす技術です。農業分野では殺虫剤への抵抗性アリル（対立遺伝子）を持つ害虫が、地域の害虫集団に存在する頻度が剤の使用可否の判断基準となります。生物多様性の保全においても、在来種の生息地に外来種の（あるいは同種の別地域から侵入した系統に由来する）アリルが含まれる割合を知ることは、優先すべき保全対象地域の決定や交雑リスクの算定に重要です。生物の進化や地理的分布を研究対象とする分子系統学や生物系統地理学といった分野もまた、DNA を構成する塩基に現れる多型の頻度を調べることで、種が過去に分岐した過程を明らかにしてきました。

あるアリルが生物集団に存在する頻度を調べるために、各個体から DNA を抽出し、PCR などの実験手法を用いて遺伝子型を決定する、個体別遺伝子診断が主に用いられてきました。しかし問題のアリルが数%を下回るほどの低頻度でしか存在しない場合、たとえば抵抗性発達の初期段階で抵抗性アリルを検出するには、1 か所で数百を超える検査個体数を要することもあり、個体別遺伝子診断には大きなコストが掛かることが課題でした。

研究の経緯

個体別遺伝子診断のコストは野外での採集労力と、DNA を抽出・検査する人手や試薬、消耗品の費用に大別されます。採集労力を下げるために、農業害虫ではフェロモンや光で虫を集め、粘着板トラップなどで捕集する手段が用いられてきました。検査回数の削減手段としては、複数個体から一括で DNA を抽出して得た混合 DNA 溶液（以下、バルクサンプル）の活用が試みられています。

定量 PCR や量的 DNA シーケンシングといった手法で、バルクサンプルに含まれるアリルごとの DNA 量を計測すれば、その割合がアリル頻度の推定値（点推定値）に一致します。しかし各個体から抽出された DNA 量が大きな振れ幅を持つため、推定値の確からしさ（信頼区間）を求めることが困難でした。そこで本研究ではバルクサンプル中のアリル存在量から、当該アリルが野外集団に存在する頻度の信頼区間を求める手法として、1 個体から得られる DNA 量のばらつきを確率分布で表現する統計モデルを構築しました。

研究の内容・意義

1. 生物集団に特定のアリルが占める頻度を知る基本的な手段は個体別遺伝子診断です。対象アリルの塩基配列に相当する DNA の有無を、PCR 増幅して検知する手法がしばしば用いられます。遺伝子診断を個体別に行うにせよバルクで行うにせよ、まず生息場所から多数の個体をランダムに採集します。
2. 個体別遺伝子診断では、当該アリルが全採集数に占める割合を直接調べた値が、集団のアリル頻度の推定値となります。推定値が収まる信頼区間も、二項分布[4)]という確率分布で容易に計算できます（図 1）。なお単数体生物では、各個体は一種類のアリルしか持たず、PCR 増幅の有無で遺伝子型が分かります。多くの動物は 2 倍体なので、検出したいアリルと他のアリルを併せ持つ個体（ヘテロ接合体）も現れますが、各々のアリルを対象に PCR 検査を行えば、個体の遺伝子型を把握できます。
3. 測定したいアリルが低頻度と予想される場合に、検査コストを抑えるために個体別ではなく複数個体まとめて DNA を抽出し、バルクサンプルとして分析する手法が用いられます。バルクサンプルに通常の PCR を行うと、最低 1 個体でも当該アリルが含まれるか、全く含まれないかの判別しかできないため、代わりに定量 PCR や量的 DNA シー

ケンシングで、アリルごとの DNA 含有量を直接測定します（図 1 右側、図 2)。各個体の DNA 収量が一定であれば、この測定値はバルクサンプルにおいて、各アリルを持つ個体数（単数体生物の場合）に比例するはずです。

4. しかし実際のバルクサンプルでは、各アリルの DNA 量は、個体数に比例した期待値の周りにランダムにばらつきます。1 個体の生物が持つ DNA 量が一定でなく、生体を構成する細胞の数（個体の大きさや齢期）に応じて異なるためです。またトラップ上に捕獲された生物個体は、死亡した時点から体内で DNA の分解が始まります。長期間の設置後にトラップを回収すると、死亡後の日数に応じて分解度が異なる個体が混在し、DNA 収量の振れ幅はさらに大きくなります。これらの関係を数学的に厳密に表現すると、アリル頻度の信頼区間を求める際の計算量が増大してしまう難点がありました。
5. そこで本研究では、個体あたりの DNA 収量をガンマ分布（図 3）という確率分布で近似する統計モデルを提案しました。集団のアリル頻度と、バルクサンプルにおけるアリル含有量の測定値の関係（図 1）は、バルクサンプルを構成する個体数（二項分布モデル）と、DNA 収量（ガンマ分布モデル）の両方を考慮して立式されます。
6. なお実験操作も、DNA 量の偏りや測定誤差を生じる原因となります。偏りや誤差の形は遺伝子診断の手法によって異なるため、量的シーケンス（発表論文 1）と定量 PCR（発表論文 2）の各手法について、フリーの統計解析環境である R を用いて集団のアリル頻度の推定値とその信頼区間を計算するパッケージを個別に開発しました。とりわけ定量 PCR に対応するよう作成した “freqpcr” は、R 言語の公式なパッケージ配布サイトである CRAN でも公開されています。

今後の予定・期待

“freqpcr” が農業分野で活用できる例として、害虫であるミカンハダニにおけるキチン合成酵素阻害剤抵抗性遺伝子の地理的分布の解析に適用し、抵抗性遺伝子を持つ系統が九州の一部に局在することを明らかにしました（Tadatsu et al. 2022 *Pest Management Science*, DOI: 10.1002/ps.7021)。地域ごとに複数個体のハダニを採集し、一括抽出した DNA に対し、定量 PCR を用いて測定した抵抗性遺伝子頻度を当初は点推定値で示していましたが、 “freqpcr” が開発され、遺伝子頻度の信頼区間を統計上の根拠と共に示せるようになりました。

生物 1 個体からの DNA 収量をガンマ分布で近似するアイデアは、幾つかの先行研究に存在しますが、本研究で提示したモデルでは、生体において各個体の DNA 量がほぼ平均値に等しい場合から、トラップ上の死骸、さらに体外に放出された DNA（環境 DNA）のように DNA 量に大きな個体差がある場合まで、単一の確率分布で柔軟に表現できます。

動物、植物、昆虫を対象とした野外での生態・疫学調査のための、個体を対象とした採集に限らず、生物集団の遺伝子頻度をバルクサンプルから定量する、様々な用途に適用可能です。たとえば実験室での微生物培養においてサイズの異なる複数コロニーがカウントされ、一部が変異株であるとき、これを懸濁した溶液からでも、ガンマ分布を仮定すれば本モデルにより遺伝子頻度を区間推定できると考えられます。パッケージはオープンソースで公開されているため、アリル頻度推定を必要とする、あらゆる生物関連分野の検査キットや分析機器、解析プログラムに本モデルを組み込めます。

用語の解説

1）アリル（対立遺伝子）

生物の形や性質を決める遺伝子情報は、DNA を構成するポリヌクレオチド鎖上の特定の位置（遺伝子座）にある、塩基の配列として保持されています。塩基が別の種類に置き換わると、同種の生物であっても形や性質の異なる個体が現れるようになります。同一遺伝子座上に異なる塩基配列を持つ個体が、種内に混在するとき、異なる型の遺伝子それぞれを対立遺伝子（アリル）と呼びます。

2）定量 PCR（Polymerase Chain Reaction、ポリメラーゼ連鎖反応）

特定の配列を持つ DNA の断片を複製する化学反応が PCR です。増やしたい DNA 配列の端に、相補的な塩基配列を持つ短い一本鎖 DNA（プライマー）を結合させ、これを起点に DNA 合成酵素を作用させると、1 回の反応で約 2 倍ずつ DNA 量が増幅されます。一定の濃度に達するまでに要した増幅回数から、反応液中に当初存在していた DNA 量を逆算する手法が定量 PCR です。また、複数のアリルのうち一種類だけ結合・増幅するようにプライマーの配列を工夫することで、アリル頻度を測定できます（図 2 を参照）。

3）DNA シーケンシング技術を用いた DNA 量の測定（量的 DNA シーケンシング）

以前より用いられている DNA シーケンシング（サンガー法）は DNA の塩基配列を、通常は定性的に調べるものです。ただしシーケンサーの機器内では、4 種類の塩基（アデニン、グアニン、シトシン、チミン）がそれぞれ発する蛍光強度を測定しています。異なる遺伝子型の複数個体が混合された DNA 溶液、すなわちバルクサンプルをシーケンサーに掛けると、蛍光のピーク高からアリルの含有比を簡易的に測定でき、この手法を量的 DNA シーケンシングといいます。

4）二項分布

アリル頻度が p（$0 \leqq p \leqq 1$）である生物集団（単数体生物とします）から n 個体を採集したとき、うち当該アリルを持つ個体の数である m は二項分布 $B(n, p)$ という確率分布に従います。採集数 n が極めて多ければ $p = m / n$ とみなせますが、実際の採集数は有限であるため p の推定値には不確かさが伴います。このような推定値が収まる妥当な幅を、確率分布（ここでは二項分布）を用いて計算することを区間推定といいます。

発表論文

1. Sudo M, Yamamura K, Sonoda S, Yamanaka T (2021). Estimating the proportion of resistance alleles from bulk Sanger sequencing, circumventing the variability of individual DNA. *Journal of Pesticide Science* 46(2): 1-8.（オープンアクセス）
https://doi.org/10.1584/jpestics.D20-064

2. Sudo M, Osakabe M (2022) freqpcr: estimation of population allele frequency using qPCR ΔΔCq measures from bulk samples. *Molecular Ecology Resources* 22 (4) 1380-1393.（オープンアクセス）
https://doi.org/10.1111/1755-0998.13554

参考図

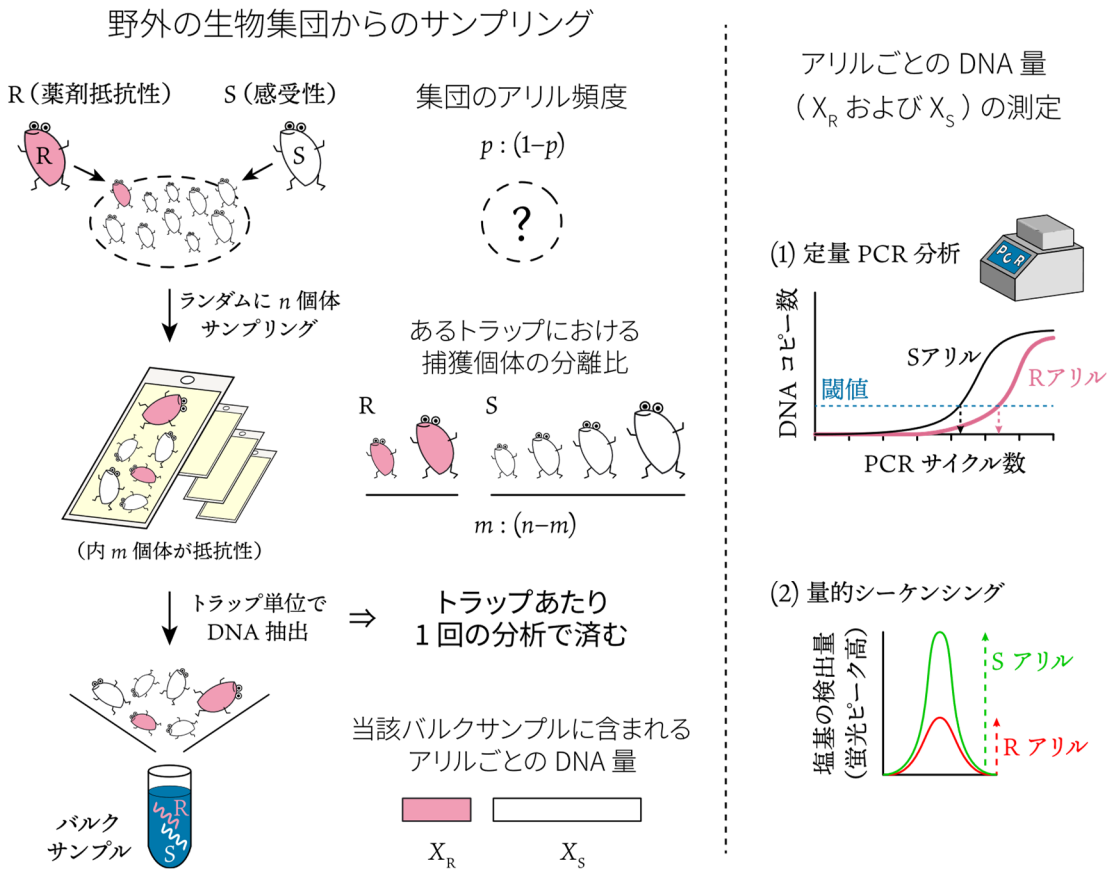

図 1　野外の生物集団からのサンプリング

（左）ある単数体生物の集団が、薬剤抵抗性のアリル（R）を頻度 p で持っており、残りの $1-p$ が感受性アリル（S）だったとします。粘着板トラップで n 個体を採集したとき、抵抗性個体が m 個体含まれる確率は二項分布で表せます。このトラップから一括抽出された DNA 溶液（バルクサンプル）に含まれる、アリルごとの DNA 量（X_R, X_S）は m 個体ないし $n-m$ 個体分の合計量です。（右）アリルごとの DNA 量を測定して比率を求めるために、定量 PCR 分析の増幅サイクル数（Cq 値：図 2 に詳細）や、量的 DNA シーケンシングの塩基検出量といった指標が用いられます。トラップあたり 1 回の分析でよいため労力を削減できますが、DNA 量の個体差があるため含有比の測定値が m ： $(n-m)$ とは厳密には一致せず、統計理論（図 3）に基づいた補正が必要となります。

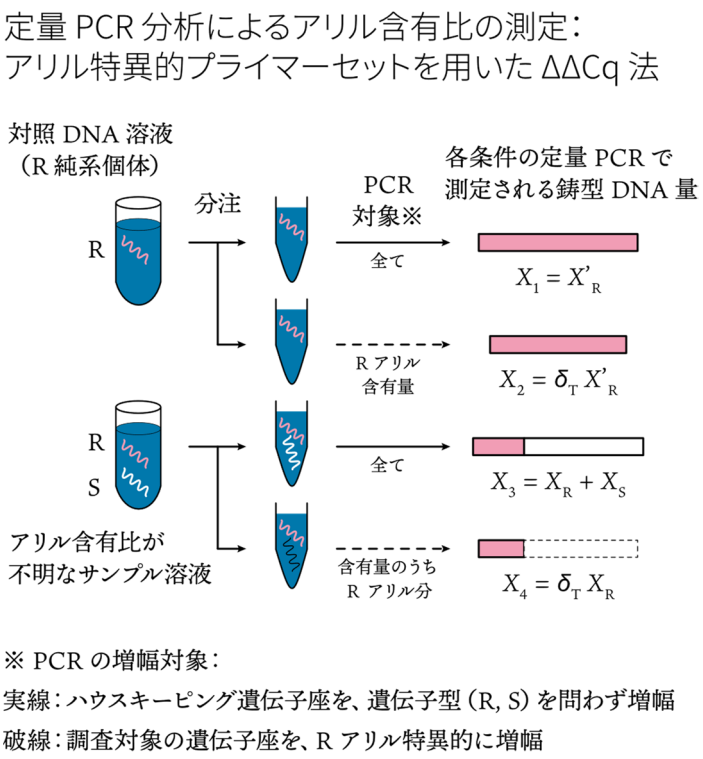

図 2　定量 PCR 分析によるバルクサンプルのアリル含有比の測定

本研究の発表論文 2 で開発した　“freqpcr”　パッケージでは、バルクサンプルに含まれる 2 種類のアリル（R と S、あるいは R とそれ以外）の比率を求めるために、幾つかの定量 PCR 手法を利用できます。代表的な手法が、アリル特異的プライマーセットを用いた ΔΔCq 法（Maeoka et al. 2020 *Applied Entomology and Zoology*, DOI:10.1007/s13355-020-00686-7）です。アリルの含有比が不明な「検査サンプル」の他に、R アリルだけを持つことが分かっている純系個体から抽出した DNA 溶液を、「対照サンプル」として用意します。検査サンプルと対照サンプルの各々をさらに二分割し、1 つは対象とする遺伝子座上の、R アリルだけを増幅するプライマーセットで PCR 増幅します。もう 1 つはサンプルの DNA 濃度の基準とするために、遺伝子型に係わらず体内の存在量が安定している遺伝子（ハウスキーピング遺伝子）を増幅します。これら 4 つの測定値 X_1〜X_4 から計算される（X_4 / X_3）/（X_2 / X_1）に相当する指標（ΔΔCq 値）が、およそ R アリルの含有比の近似である X_R /（X_R + X_S）となります。さらに freqpcr を用いることで、バルクサンプルが取られた個体群における、R アリル頻度の信頼区間を推定できます。

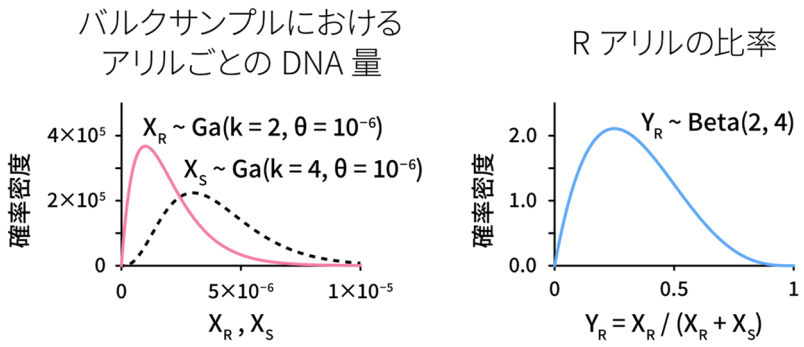

図 3　バルクサンプルのアリル含有量やその比率の確率分布による表現

（左）生物の 1 個体から得られる DNA の量は、理論上あらゆる正の値を取ります。この量を簡便に表現できるのが、ガンマ分布という確率分布です。たとえば 1 個体分の DNA 量が形状母数 $k = 1$ のガンマ分布に従うと仮定します。バルクサンプルが 2 個体の R と 4 個体の S で構成される場合、2 個体分の R アリルの合計含有量である X_R は形状母数 $k = 2$ のガンマ分布、4 個体分の S アリルである X_S は $k = 4$ のガンマ分布で表現できます。（右）さらに、このバルクサンプルに含まれる R アリルの比率である $Y_R = X_R / (X_R + X_S)$ は、ベータ分布という確率分布に従います。2 つのガンマ分布の代わりに 1 つのベータ分布で式を表すことで、コンピュータによるアリル頻度推定が高速になります。