

遺伝解析に価値ある画像解析～イネ科の葉形態を例に～

坂本莉沙

日本学術振興会特別研究員 (DC)

東京大学大学院 農学生命科学研究科 生産・環境生物学専攻 博士課程

〒113-8657 東京都文京区弥生 1-1-1 東京大学 農学部 1 号館 328

Lisa SAKAMOTO

Picture-based phenotyping of sorghum leaves for multi-trait QTL analysis

Key words: *phenotyping, QTL analysis, image acquisition, biometrics, image segmentation*

JSPS Research Fellow

Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, The University of Tokyo

DOI: 10.24480/bsj-review.10b4.00157

1. 量的形質の遺伝解析をする。～QTL解析～

1-1. 量的形質の品種改良

新しい有用な品種を作ることは農業上重要である。新しい品種を作るには異なる優良な農業関連形質をもつ品種同士を掛け合わせ、目的の表現型を持つ品種を作ることが一般的である。改良したい形質が質的形質の場合、交配後代を見れば形質関連遺伝子の数や、表現型と遺伝型の対応がわかり効率的に改良できる。問題となるのは目的形質が量的形質の場合である。なぜなら、量的形質では、主導遺伝子だけでなく微動遺伝子による支配、あるいは環境変動の影響を大きく受けるため、多くの場合交配後代で表現型が分離しない。このような場合、質的形質の場合とは異なり、表現型と遺伝子型の対応が容易につかないため、量的形質の改良は一般的に容易ではない。

1-2. QTL解析の原理

量的形質に関わる遺伝様式を理解するために QTL 解析を活用することが出来る。交配集団を用いた QTL 解析では交配後代の表現型とゲノムワイド DNA 多型マーカーとの関連を検定する。つまり、統計学的な検定により有意に表現型と関連しているゲノム領域 (Quantitative trait locus, QTL) を探索することになる。QTL 解析についての詳細な説明は鶴飼 (2000) や Broman *et al.* (2009) を参照してほしい。このように、QTL 解析は交配集団を用意し、集団のゲノムワイド DNA 多型マーカーと対応する表現型値があれば行うことができる。近年、次世代シーケンサーの登場によりゲノムワイド DNA 多型マーカー情報の取得はそれほど大きな課題ではなくなりつつある。一方、各系統の表現型情報をどのように取得するかは、依然として大きな課題である。表現型の計測はフェノタイピング (phenotyping) と呼ばれ、より優れたフェノタイピング技術の開発に向けた研究が近年盛んに行われている (Li *et al.* 2014, Cobb *et al.* 2013, Kim *et al.* 2017)。

L. Sakamoto-1

1-3. QTL解析のための表現型計測

各個体の表現型を計測するのは容易ではない。例えば、止め葉の長さに関わる QTL が知りたいとする。止め葉は穂が出る直前の葉で、最後まで光合成をして穂に栄養を送るとされているため大変重要である。しかし、大抵の作物は年に 1 回しか栽培できず、毎年秋に計測作業が集中してしまう。その上、止め葉の長さを調べるには穂が出てから枯れるまでの限られた期間に長さを測る必要がある。また、統計的な検出力や組替え価などを考慮すると数百系統を計測する必要がある。つまり、短い期間に数百系統の計測を終わらせる必要がある。このため、とにかく作業効率を高めなければ表現型計測を目的数こなすことができない。計測作業効率を高める表現型計測のことを *high-throughput phenotyping* と言う (Araus and Cairns 2013, Andrade-Sanchez *et al.* 2014, Chen *et al.* 2014)。

また、そもそも表現型の評価自体難しいこともある。例えば、葉が赤くなる病害の抵抗性に関わる QTL が知りたいとする。その際、我々は葉の赤さを評価する必要がある。しかし、病害による葉の赤さというのは斑ら、スポット状に現れるため、葉面積、色の濃さ、部位などが関連する複雑な表現型となる。このため、総合的に評価するために達観で病害程度をスコアリングすることが多い。しかし、それでは葉の赤い部分の面積なのか、赤い部分の全体に対する割合なのか、赤色の濃さなのか曖昧なまま評価がなされてしまう。つまり、総合的かつ客観的な評価を行うことは難しい。表現型を明確に定義して網羅的に計測することを *high-resolution phenotyping* と言う (Crowell *et al.* 2014, Crowell *et al.* 2016)。

このような現状を踏まえて表現型の評価に画像解析を活用しようという研究が盛んになっている。*high-throughput phenotyping* も *high-resolution phenotyping* も画像解析を活用した研究が多い。画像解析を活用したフェノタイピングは *image-based phenotyping* と呼ばれる。

2. 表現型計測に画像解析を活用する。

本章では具体的な研究例をもとに表現型計測の手順とコツを紹介する。本研究ではソルガム (*Sorghum bicolor* (L.) Moench) の葉形態に関わる QTL の検出を目的とした。ソルガムは世界五大穀物の一つのイネ科作物である。他の穀物が育ちにくい乾燥した土地でも生育することや茎に糖を蓄積することからバイオ燃料材料の可能性があることなどから近年再注目されている作物の一つである。材料はソルガムのモデル品種 BTx623 とタカキビの組換え自殖系統 209 系統、登熟期の各系統 1 個体を使用した。これは岡山大学資源植物科学研究所で 2016 年、2017 年にガラス温室でポット栽培されていたものである。ソルガムの葉はトウモロコシと良く似た 50 cm~1 m のコンブ状の葉で主茎に 5~10 枚展開する。ソルガムの主茎の葉を順番に全て切り取り、表現型計測をした。

2-1. 表現型計測の手順

作業手順の概要を述べる。まず、画像を取得する。カメラの位置による画像の大きさを補正し、そこから葉のみを抽出する。葉のみの画像から表現型計測を行い、QTL 解析を行う。

2-1-1. 画像取得

撮影装置として巨大な三脚 (高さ 1.5m 以上) にカメラを真下に向くように取り付けた (Fig

L. Sakamoto-2

1)。今回、カメラは Canon の一般的な一眼レフカメラ (EOS kiss X5) を使用した。このカメラの後続機は 8 万円程度で購入できる。カメラとパソコンを接続してパソコン上のカメラ操作アプリ (EOS Utility) を使用して三脚上のカメラのシャッターを切った。EOS Utility は Canon の Web ページ

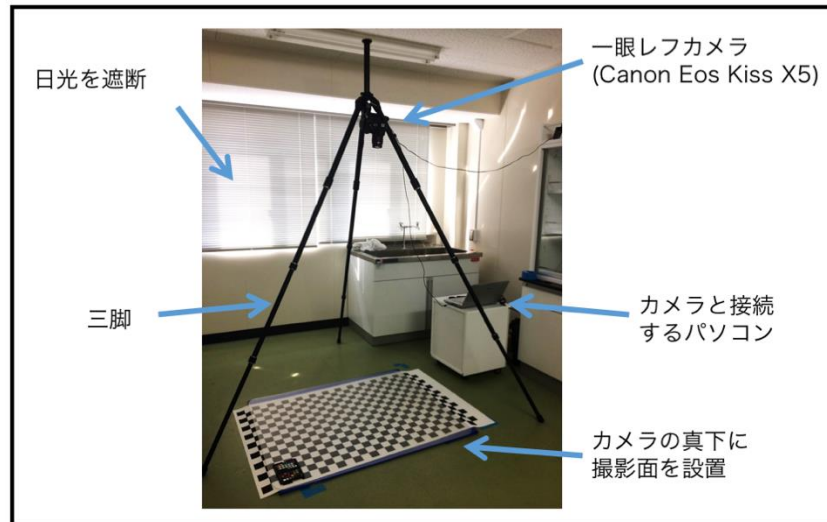


Fig 1 撮影装置

から無償でダウンロードできる。このアプリを使用すると、どのように撮影されるかをパソコン上で見ながらシャッターを押すことができる。また、カメラをコンセント給電しながら撮影できるようにし、バッテリー交換不要にした。この撮影装置により、三脚の下にあらかじめ並べておいた葉を設置すれば撮影を行うことができる。さらに、葉を切り取る人、並べる人、撮影する人など分業を行うことができ、アルバイトを雇い人手を増やせば高速に画像取得をすることができる。今回の計測では 7 人 3 日間で 200 個体以上計測ができた。

今回の画像取得作業と通常の手計測作業を比較しよう。手計測をする際、計測のマニュアルをもとに多くの人手で分担することが多い。しかし、マニュアルに載っていない事態が多発し、現場が混乱し、うまく計測作業が進まないことがある。例えば、葉の長さや幅を計測するだけでも、「葉の先が 2 つに分かれている葉はどうするのか?」「幅の計測が人によって一番太い部分か真ん中か違って計測していたことが休憩時に発覚」などの問題が発生する。これは計測マニュアルの表現型の定義が曖昧であったり、定義の例外が出たりするせいである。しかし、今回のように画像解析を活用する場合は作業の時はとりあえず写真だけとっておけばよく、表現型については後で写真を見ながら考えることができる。実際、今回の撮影作業時に大きな混乱はなかった。ただ、シャッターを切る人は画像解析に慣れている人が行い、画像解析に耐えうるクオリティか確認しながら作業を進めるのが良いだろう。

撮影条件は以下のように整えた (Fig 2)。上位葉 (穂に近い葉) から下位葉 (地際に近い葉) まで左から順に葉先が上に、葉元が下になるようにならべた。葉を並べる際は葉同士が重ならないようにし、葉が丸まらないように葉の裏をテープで止めている。背景を青で統一し、一枚の写真に 1 個体とした。どの写真がどの系統かわかるように系統名をホワイトボードに書き、写真に入れた。スケールがわかるものとして赤い直径 2 cm のシールを背景に貼った。葉が波打つのを抑えるために透明な板を重石として乗せている。

今回、このような撮影条件にしたのは撮影条件の基本である「情報の消失が少なくなるようすること、条件を整えて条件を全写真で統一すること」と撮影状況のバランスを加味した結果である。今回の撮影条件でうまく行くのか、予め数枚写真で予備実験して条件検討しておくのが良いだろう。その際、自らの画像解析スキルに合わせた撮影条件にしておくことが重

要である。

まず、情報の消失が少なくなるようにするには並べること、スケールを入れることが大切である。今回は上位葉から下位葉を並べた。これにより葉の位置についての情報が写真にあることになる。スケールは定規だけでなく、画像処理上使いやすい円の直径や、チェッカーボード（市松模様）を使って欲しい。画像処理では線を扱うのは難しく、定規のメモリは線であるためである。また線に太さがあるため、正確に長さを測るのに向かない。色味の解析をする場合はカラースケールを入れるのを忘れないで欲しい。今回は写真番号と系統の対応関係の情報が無くならないように写真の中に系統名を入れた。画像解析の技術がない場合は無理に系統名を入れなくてもいいが、写真の中に入れてしまったほうがリスクは少ないように思われる。

また、条件を整えて条件を全写真で統一する理由は画像解析をしやすくするためである。それに加えて後から別の表現型を計測したくなった場合にも有効である。今回、著者はカメラのホワイトバランス設定をオートにしてしまい、色味の条件が写真によってやや異なっている。これにより、色の濃さの解析がやや難しくなった。カメラの設定はオートにせず、マニュアルで設定して全写真で統一したほうがよい。しかし、場合によっては条件を統一できない場合がある。野外での撮影は光の強さが変わるため、シャッタースピード等をオートにして写真の白飛び（光を感知しすぎて写真の一部が白く潰れてしまうこと）を防ぐ方がいいだろう。画像解析の煩雑さを増すが白飛びすると表現型を取り出すことすらできなくなるのでそれよりは良い。条件を統一するほど計測は煩雑になり、画像解析は簡単になり、表現型計測は精密になると言う関係性にあることを心に止めて、最適な条件を検討すべきだろう。

画像取得の際に著者が考慮する具体的な項目は、どの機材で画像を取得するか、個体あたりの計測時間、分業のしやすさ、背景色、光源、機材設定（レンズ、ズーム、シャッタースピード、絞り、ISO、ホワイトバランス、フラッシュ、画像サイズ、画像の保存形式）、スケールを何にするか、どう並べるか、自力で画像解析可能か、計測したい表現型がきちんと写真から確認できるか、である。

撮影機材は主に一般的に流通しているカメラかスキャナのどちらかを使用している。カメラは今回のように大型のものも瞬時に画像取得することができ、持ち運びもスムーズである。また、レンズや設定を適切に調節することで様々な大きさ、厚みを持つ撮影対象にも対応が簡単である。しかし、レンズの魚眼効果により画像の端に歪みが生じたり、ピクセル数と実際の大きさの対応を後付けする必要があるったり、照明や撮影設定を調節しないと色の解析が難しいなどの問題がある。スキャナはdpiの情報があるため、大きさの補正を必要とせず、魚眼効果による歪みもなく、光源もスキャナと一体となっている

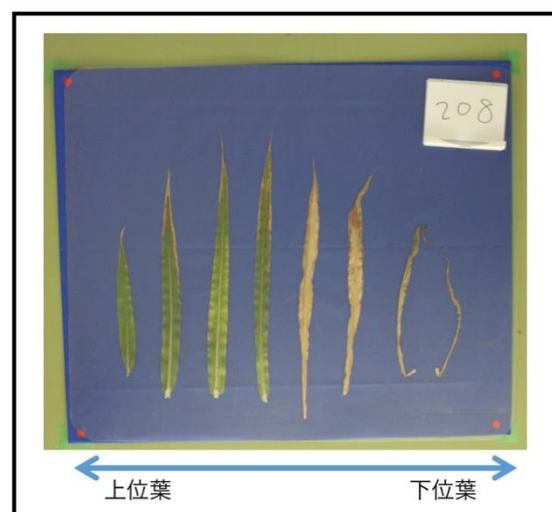


Fig 2 撮影条件
ホワイトボードに書いてある数字は系統名を示す。
背景の角にある赤丸は直径2cmのスケールである。
上位葉から下位葉まで左から順に葉を並べた。

ため、画像解析がしやすい。しかし、読み取り面から離れるとボケるため厚みのある場合は適さず、大きすぎると読み取り時間がかかる。カメラに比べて持ち運びに適さない機材が多いのも難点である。

光源は種類と位置に注意する。太陽光は時間とともに変化するのでそれに応じた画像の補正が必要になってくる。人工光は写真内に光の反射や影が発生しない位置におく。光の反射は画像解析を困難にするので反射しやすい素材を画像に入れないことも大事である。

画像の保存方式はできるだけ情報を保持できるものを選ぶ。画像の保存形式には画像を圧縮する形式 (JPEG 形式など) と圧縮しない形式 (TIFF 形式など) がある。圧縮すると輪郭部分が曖昧になるので圧縮しない形式を勧める。また、センサーの情報をそのまま保存する RAW データも保存できるならした方がよい。今後さらに現像方法が良くなってより良い画像を得られる可能性があるからである。

シャッタースピード、絞り、ISO はそれぞれが関連する重要なカメラ設定である。この 3 つをうまく調節してブレがなく、解析対象にピントが合い、黒つぶれも白飛びもない画像が取れるように調節する。

画像取得に関して補足的な内容にはなるが、撮影画像は数百枚にのぼり、管理方法が問題となる。どの画像が何を写したものが分からなくなることをのまないようにしたい。著者は画像のファイル名を「撮影年度_系統番号.tiff」となるように名前の変換表を作り、プログラムによる一括変換を行った。ピンボケなど撮影に失敗した写真は「reject 撮影年度_系統番号.tiff」としてファイル名で区別がつくようにした。その後の画像解析に使用する写真を 1 つのフォルダにいった。これにより、画像解析で一括処理がしやすくなり、ファイル名と表現型値を表データとして出力すると系統と表現型値の対応がつくようになった。数百枚を手動で名前変更をすると表記揺れ、タイプミスが発生するので避けたい。

2-1-2. 画像処理による表現型値の取り出し

三脚の高さが変わると写真によって 1 cm あたり何ピクセルかバラバラになる。よって、画像処理による大きさの補正を行った。MATLAB の Camera Calibrator Toolbox を利用した。三脚を動かすたびに交点間距離が 50 mm のチェッカーボードを撮影した。そして、チェッカーボードの交点間が 150 ピクセルになるようにホモグラフィ変換 (Hartley and Zisserman 2003) を作った。どの画像がどの位置の三脚だったかに対応したホモグラフィ変換を行い、全ての画像で 150 ピクセルを 50 mm として扱うことができるようにした。もしこのような画像解析を行う技術がないのであれば三脚を絶対に動かさず、完全に真下を向くようにしてカメラで撮影するか、スキャナーで dpi の設定を固定にしてスキャンするのが良い。スケールとして入れている赤い直径 2 cm のシールが画像上で 600 ピクセルになっていることを確認した。600 ピクセルからどれくらいずれているかを見れば今回の実験の精度を確認できる。このように目的の画像解析をする前に画像の補正を行うことを画像の前処理と呼ぶ。今回は行っていないが、レンズの魚眼効果の補正、照度の補正、色味の補正、部分切り出しなどを必要に応じて前処理する。撮影条件と求める測定精度によってこれらの前処理の必要性が決まる。

量的形質を取り出すための画像解析は目的の部分以外を黒くしたマスク画像から行うのが一般的である。画像上のどのピクセル部分がどのオブジェクトに対応しているかを処理する

ことを画像のセグメンテーションという。画像のセグメンテーションをして目的以外を黒くすればマスク画像ができる。画像のセグメンテーションは画像解析の一大分野であり次々に新しいセグメンテーション手法が提案されている。画像の特徴によってうまくいくセグメンテーション手法は異なるので「画像 セグメンテーション」で Web 検索して色々な手法を理解して実践すると良いだろう。今回は MATLAB の Image Processing Toolbox を利用し色の閾値処理によってセグメンテーションした。背景が青色であるため青いピクセルに囲まれているオブジェクトが葉であるとするのでマスク画像を作成した。

マスク画像作成には 3 つのポイントがある。1 つ目は処理の仕方である。画像処理は全ての画像に同じ処理を for 文で行うと一番労力がかからない。そのためには撮影時に全写真で条件を統一した部分を軸にして解析するとうまく行くことが多い。条件を統一しているので同じ処理によって同じ効果を得られる可能性が高いからである。今回の場合、緑色やベージュ色が葉であるとしてしまうと葉の赤い着色がある部分が背景とされてしまう。遺伝解析を行う際は表現型にバリエーションがあるため全てのバリエーションに対応した処理を考えるのはやや難しい。できるだけ例外が発生しないような処理を考えるとうまく行くことが多い。統一した処理をするのが難しい場合はグラフカット手法 (Rother *et al.* 2004) による半自動セグメンテーションを活用するのがいいだろう。マスク画像作成がうまくいかない場合は画像の前処理を行うと良い場合もある。それでもうまくいかない場合は Adobe Illustrator などのペイントソフトで手動で領域指定するか、画像を撮影し直すのが良い。何枚かうまくいかない程度であればそれらは欠測値として扱う。2 つ目はマスク画像の精度である。マスク画像の精度は目的によって許容できるものが異なる。面積を求めたいのであれば先端等がやや欠けていてもおおよそ面積は誤差の範囲で変わらないので問題ない。輪郭を求めたいのであれば、先端が欠けると輪郭形状としては大きな違いとなるので許容できない。3 つ目は目視による確認である。マスク画像を作成できたら、まずは人間の目でどのような表現型があるかを確認すると良い。マスク画像には背景などの余計な情報がなく、全ての個体を同じ条件下で高速に何度でもみることが出来るからである。普段は気がつかなかったバリエーションに気がついたり、明らかに異常な個体がないかチェックしたりできる。

本研究では high-resolution phenotyping を目指し、明確な定義で網羅的に計測をした。まず、葉一枚ずつの解析を可能にするため個葉ごとに別々の画像として保存した。具体的にはオブジェクト解析 (Haralick *et al.* 1992) によって自動的に個葉を検出し、左から順に、つまり止葉から順に 1 から番号を振り保存した。2016 年は合計で 1845 枚の葉があった。保存した葉一枚の画像から欲しい表現型値を取り出した。まず、葉一枚の画像を上下に分割し、葉の上半分の葉先の表現型を取り出す場合と、下半分の葉元の表表現型を取り出す場合と、分割せずに一枚の葉全体から表現型を取り出す場合の 3 通りを適用した。これを葉のどの部分から表現型を取り出したかと呼ぶことにする。次に、色の閾値を設定し、閾値を超えた部分のピクセル数を面積とした。全ての色のピクセル数を葉面積、緑色のピクセル数を緑面積、緑色以外のピクセル数を枯面積、赤色のピクセル数を赤面積とした。また、葉面積から緑面積を葉面積で割ったものを緑割合、赤面積を葉面積で割ったものを赤割合とした。また、緑割合が 10%に満たない葉は枯れていると判定した。これにより、葉一枚ごとの表現型値を得ることができたが、QTL 解析を行うには系統ごとの表現型値を得る必要がある。今回は 1 系統

1 個体であるので個体内の合計なのか平均なのかといった統計量を求めることで葉一枚ごとの表現型値から系統ごとの表現型値を求めた。さらに、その際に枯れていると判定した葉を取り除くか、のぞかずに全ての葉を使うかの2通りを適用した。また、止め葉の表現型値なのか、止め葉以外の葉の表現型値なのか、止め葉も含めて全ての葉を使うのかの3通りも適用した。以上のように表現型を計測したため、形質数が掛け算方式に増えた。何年の栽培か、枯葉を含めるか、止め葉を含めるか、葉のどの部分か、どの色のピクセル数または割合か、個体内の合計なのか平均なのかといった統計量によって表現型が評価された。つまり、2016年の枯葉を除く主茎の葉全ての一枚全部の葉面積の合計、つまり2016年の枯葉を除く総葉面積というふうに表現型が取り出された。全部で490形質であった。

今回のように画像解析を活用することで網羅的に表現型を計測することができる。しかし、それでも今回の画像解析では計測できなかった表現型がある。例えば光合成効率や受光体制について興味がある場合は今回の計測では主茎と葉の角度の情報がない。今回は破壊調査で葉を切り取ったからである。主茎と葉の角度の情報を画像取得段階で失っているため画像解析をしても計測することができない。画像の撮影の際にどのような情報消失が起きてしまうのかがあらかじめ分かっているならば画像解析以外の手法で（大抵の場合は手作業で）計測することを勧める。対照的にうまく表現型計測ができたものもある。一枚の葉の画像を上半分、下半分に分割するだけで葉先と葉元に分割することができたのは全ての葉で向きを揃えておいたからである。葉先と葉元に分割したのはどの肥料が不足するかによっては先から枯れるか、元から枯れるかが変わると知っているからだ。つまり、葉先と葉元では異なる情報を持っていることを知っている。生育した植物を見て植物がどんな情報を持っているか考えてそれらが失われないように撮影し、画像解析で表現型を取り出すことが大切である。情報が失われないようにするにはランダムにしないで揃えて並べるのが基本である。著者はこれができるのは日頃から植物に接している人、植物に造詣が深い人だと考えている。植物の表現型にどんな意味があるのか。これをよく理解しているからこそより良い表現型計測をできる。

2-2. 画像解析の長所

ここで、表現型計測に画像解析を活用する利点について述べる。今回の画像解析では特殊なことはしていない。画像処理の入門書に書いてある処理しかしていない。その上、プログラミングも Web で手に入るサンプルコードをつなぎ合わせて入力を今回の写真に変えるぐらいしかしていない。「画像解析は敷居が高い」と思う読者もいるかもしれないが、それはすでに手元にある画像で画像解析を行おうとするからである。自らが画像解析を行うことを前提として自ら写真をとれば画像解析は簡単にできる。というよりも自らの画像解析スキルでできる範囲の画像を取得するように心がけるようにすべきである。自分のスキルに合わせて柔軟に取り入れられるのが画像解析のいいところである。

達観による評価と今回の表現型計測を比較する。達観による評価では難しいほど多数の表現型を評価することができた。また、達観では区別の難しい緑面積と緑色割合を別々のものとして評価することができた。葉一枚一枚の緑色割合と緑面積の散布図 (Fig 3) をみるとこの2つは全く異なる表現型であることがわかる。また、面積や割合をスコアリングではなく

連続値で絶対評価することができた。画像解析を活用することでレゾリューションの高い計測が可能となる。

今回の計測でかかったコストについて見る。計測の際にかかるコストとして葉を切って並べるのに7人の人手を必要としたが、計測期間は3日で済む。カメラや三脚も特殊なものを使用していないので他の実験への汎用性が高く、価格も安い。また、画像解析も今回はそこまで特殊な解析をしていないので通常の Mac mini で解析できる。画像解析ソフトは有料の MATLAB を使用しているが、無料の open CV を使用しても問題ない。MATLAB を使用しているのは公式のサンプルコードやアプリが充実しており、プログラミング労力コストが低く済むからである。画像解析を活用した表現型計測のコスト的なハードルは低いといえる。

3. 計測した表現型を解析する。

画像解析を使うことでたくさん計測ができたが、これは生データを手に入れたということであってこれがすなわち結果というわけでない。データ解析をして結果を見て考察しなくてはならない。これは基本的な事だが、実験計画作成段階で使用予定の解析手法についての基本的な知識は得ておいたほうが良い。適用できる条件や手法の限界、必要とする生データの形式を理解する事で解析可能な生データを得ることができる。今回の研究では 209 行 490 列の表現型データと 209 行 1799 列のゲノムワイド DNA 多型マーカーの関連を解析する。しかし、どのように解析したらいいのだろうか？ここからは大量の表現型データを手に入れた後の解析方法について紹介する。今回はたくさん計測をしたいということで画像解析をした。しかし、たくさん計測するという事は一体どういうことなのだろうか？

まず、たくさん「個体」を計測する場合である。これは high-throughput phenotyping の発想である。表データでは行数を増やすことに当たる。すると統計学的な検出力の向上と計測の労力の削減につながる。もしも high-throughput phenotyping によって測定誤差が増えるのであればそれによる統計学的な検出力の低下と天秤にかける必要がある。

次に、たくさん「表現型」を計測する場合である。これは high-resolution phenotyping の発想である。表データでは列数を増やすことに当たる。すると検定回数が増え、多重比較の問題が発生する。また、解析結果の数が増え、考察が大変になる。さらに、相関の強い形質をたくさん計るだけとも見ることができデータを水増ししているだけとも見て取れる。いい結果となった表現型だけを論文にすることは統計学的に推奨されない。つまり、たくさん表現型を測定する場合は何らかの工夫をする必要がある (Granier and Vile 2014)。このようなたくさん表現型を解析する解析は多変量解析、オーム解析、オミックス解析と呼ばれる。一般的に多変量解析は勉強

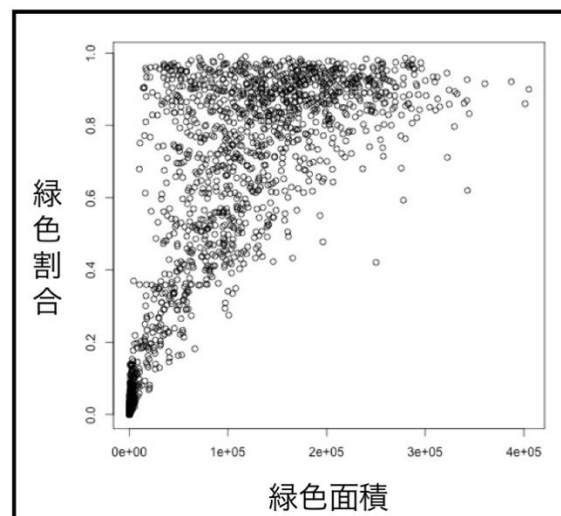


Fig 3 葉ごとの緑色割合と緑色面積
各点は葉1枚を示す。画像解析によって得られた葉全てをプロットした。縦軸は各葉の緑色面積を、横軸は各葉の緑面積を表す。

して使えるようになることも実際のデータに適用して解析することも難易度が高めである。その上、個体数が限られている場合はさらに難易度が上がる。もしも **high-resolution phenotyping** をしようとしている場合は統計学がわかる人が研究チームにいないと生データを解析することすら困難、論文化しようとしても解析の問題点を指摘される事態になる。

形質が複数ある時の QTL 解析手法を調べたところ、そのような手法はあったが (Topp *et al.* 2013, Alimi *et al.* 2013, Li *et al.* 2006) 条件検討の結果、今回の事例では使うことができなかった。実はこれは著者が予想していなかった事態である。大抵の QTL 解析では複数の形質について議論されることが常であるので今回の事例にあった既存手法はすでにあると思込んでいた。そこで、著者は手法を開発することにした。もし作れない場合はどうすればいいのだろうか？それには 3 つの行動が考えられる。一つ目は気合を入れて数多くの結果を見て解釈することである。つまり、時間と労力で解決を図る行動である。二つ目は表現型相関の結果を見て相関が高い形質があれば片方を捨てたり、主成分分析等を使用したりして次元削減することである。つまり、問題の原因であるたくさんの表現型を減らすことで解決を図る行動である。三つ目は専門家に助けを求めることである。つまり、多変量解析を専門とする人に相談することである。また、先行研究の著者らに相談するも良いだろう。あるいは、視野を広げて異分野である統計学や情報学系の研究者や会社を頼ることも考えていだろう。また、専門家と組むことで誤った解析、解釈をするリスクを大幅に回避できる。有名な統計学者 R. A. フィッシャーの言葉に「実験が終わってしまった後で統計学者に相談をするのは、検死解剖をどのように行なえばよいかを尋ねるようなものだ」というものがある。生データの取り方によって使える多変量解析手法が変わるので、できれば実験計画作成段階から専門家に相談してほしい。本当は日頃からコツコツと多変量解析の勉強をして手法適用、開発ができるようにすることが望ましいだろう。生データを取得してから多変量解析の学習をすると論文化するまでに莫大な時間がかかるので勧めない。生データを取得してからはその時点で持っている知識で既存手法を簡単に改良する程度にとどめたい。

今回の研究では著者が提案する QTL 解析手法で 15 領域の QTL を検出することができ、それぞれがソルガムの葉形態に対してユニークな効果を持つことが確認できた。詳細な QTL 解析手法については現在論文執筆中であるためここでは述べない。

4. まとめ

画像解析を活用し主観を挟むことなく大量の表現型の計測を行うことができた。データの質と量を画像解析で高めることができる。そのためには入念な撮影条件の設定、データ解析の素養が不可欠であると言えよう。

謝辞

本稿で紹介した研究を遂行するにあたり東京大学大学院農学生命科学研究科の岩田洋佳准教授には指導教官として数々のご指導をいただいた。また同研究科の鐘ヶ江弘美特任助教、石森元幸博士には遺伝解析の具体的な方法やソルガムの遺伝子やデータベース等をご指導いただいた。また同研究科の藤本優准教授、高梨秀樹助教、堤伸浩教授にはソルガムの表現型についてのアドバイス、ソルガムの遺伝子についての情報提供、計測計画の作成、計測方法

についてご指導いただいた。岡山大学資源植物科学研究所の坂本亘教授，大西紀和助教，小童谷利恵氏には組替え自殖系統の作出と栽培，ソルガムの表現型についてのアドバイス，ソルガムの遺伝子についての情報提供していただいた。九州大学大学院，JST・PRESTOの野下浩司博士には画像取得のコツ，画像解析の手法についてご指導いただいた。明治大学農学部的小林正明博士，矢野健太郎教授には組替え自殖系統のDNA多型マーカーを提供していただいた。本稿で紹介した研究はJST,CREST「葉緑体機能改変によるステイグリーン植物の創出」(JPMJCR11B2)および「高速ジェノタイピングを利用したエネルギー作物のテーラーメイド育種技術の開発」(JPMJCR12B5)およびJSPS，科研費，特別研究員奨励費「作物形態の定量化とゲノムとの関連の解析」(17J05431)の支援を得て遂行した。

引用文献

- Andrade-Sanchez, Pedro, Michael A. Gore, John T. Heun, Kelly R. Thorp, A. Elizabete Carmo-Silva, Andrew N. French, Michael E. Salvucci, and Jeffrey W. White. 2014. Development and Evaluation of a Field-Based High-Throughput Phenotyping Platform. *Functional Plant Biology* 41 (1): 68–79.
- Araus, José Luis, and Jill E. Cairns. 2014. Field High-Throughput Phenotyping: The New Crop Breeding Frontier. *Trends in Plant Science* 19 (1): 52–61.
- Cobb, Joshua N, Genevieve Declerck, Anthony Greenberg, Randy Clark, and Susan McCouch. 2013. Next-Generation Phenotyping: Requirements and Strategies for Enhancing Our Understanding of Genotype – Phenotype Relationships and Its Relevance to Crop Improvement. *Theoretical and Applied Genetics* 126:867–887.
- Crowell, S., A. X. Falcao, A. Shah, Z. Wilson, A. J. Greenberg, and S. R. McCouch. 2014. High-Resolution Inflorescence Phenotyping Using a Novel Image-Analysis Pipeline, PANorama. *Plant Physiology* 165 (2): 479–95.
- Crowell, S., Pavel Korniliev, Alexandre Falca, Abdelbagi Ismail, Glenn Gregorio, Jason Mezey and Susan McCouch. Genome-wide association and high-resolution phenotyping link *Oryza sativa* panicle traits to numerous trait-specific QTL clusters. *Nature Communications*. 7:10527
- Granier, Christine, and Denis Vile. 2014. ScienceDirect Phenotyping and beyond: Modelling the Relationships between Traits. *Current Opinion in Plant Biology* 18: 96–102.
- Kim, Song Lim, Nita Solehati, In Chan Choi, Kyung Hwan Kim, and Taek Ryouon Kwon. 2017. Data Management for Plant Phenomics. *Journal of Plant Biology* 60 (4): 285–97.
- Li, Lei, Qin Zhang, and Danfeng Huang. 2014. A Review of Imaging Techniques for Plant Phenotyping. *Sensors* 14:20078–111.
- Li, Renhua, Shirng-wern Tsaih, Keith Shockley, Ioannis M Stylianou, Jon Wergedal, Beverly Paigen, and Gary A Churchill. 2006. Structural Model Analysis of Multiple Quantitative Traits. *Plos Genetics* 2 (7): e114.
- Rother, C., Kolmogorov, V., Blake, A. 2004. “GrabCut” — Interactive Foreground Extraction using Iterated Graph Cuts. *ACM Transaction on Graphics* 70(2): 109-131.
- Topp, Christopher N, Anjali S Iyer-pascuzzi, Jill T Anderson, Cheng-ruei Lee, and Paul R Zurek. 2013. 3D Phenotyping and Quantitative Trait Locus Mapping Identify Core Regions of the Rice Genome

Controlling Root Architecture. *Proceedings of the National Academy of Sciences* 110 (18): E1695-E1704

Wubs, M, N A Alimi, M C A M Bink, J A Dieleman, E Heuvelink, J Magan, R E Voorrips, et al. 2013. Genetic and QTL Analyses of Yield and a Set of Physiological Traits in Pepper. *Euphytica* 190:181–201.

Younes, Mamoun, Lia V. Lechago, Jacqueline R. Somoano, Moni Mosharaf, and Juan Lechago. 1996. Wide Expression of the Human Erythrocyte Glucose Transporter Glut1 in Human Cancers. *Cancer Research* 56 (5): 1164–67.