

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 13 日現在

機関番号：12102

研究種目：研究活動スタート支援

研究期間：2015～2016

課題番号：15H06071

研究課題名(和文) トマト誘発変異集団の高度利用に向けたエクソーム解析手法の確立

研究課題名(英文) Development of exome mutation database in the micro-tom mutant population

研究代表者

矢野 亮一 (YANO, Ryoichi)

筑波大学・生命環境系・助教

研究者番号：00443044

交付決定額(研究期間全体)：(直接経費) 2,100,000円

研究成果の概要(和文)：エクソーム・シーケンシング(exome sequencing)は遺伝子のタンパク質コード配列をターゲットに次世代シーケンシング解読を実施する手法であり、全ゲノムを標的とする方法に比べてコストパフォーマンスに優れている。本研究では、筑波大学・蔬菜花卉学研究室が保有するトマトのモデル品種マイクロトムの変異体集団においてエクソーム・シーケンシングを実施した。95系統の変異体において延べ240,000以上の変異を同定すると共に、タンパク質機能に影響しうる変異を変異体ごとに同定した。さらに、これらの変異情報をwebブラウザで検索するためのインターネット・アプリケーションを開発した。

研究成果の概要(英文)：Exome sequencing (ES) is a cost-effective next-generation sequencing (NGS) method for rapid identification of genome-wide exon mutations at a population scale. In this study, a 96-plex ES library preparation method was newly developed in the model tomato cultivar 'Micro-Tom'. When exome NGS data was obtained by the Illumina HiSeq paired-end sequencing platform and further subjected to a bioinformatics pipeline using bowtie2, samtools, and GATK, over 240,000 mutations were identified in 95 Micro-Tom mutant individuals. Of these, over 70,000 mutations were expected to affect protein function because they cause amino acid substitution, premature stop codons, or frameshift in the deduced protein amino acid sequences. A web-application system was also developed based on the mutation dataset to enable in silico one-click search of specific mutations or mutants.

研究分野：ゲノム情報科学、園芸育種科学、バイオインフォマティクス

キーワード：トマト 誘発変異集団 エクソームシーケンシング 次世代シーケンシング 変異情報データベース webアプリケーション ゲノムワイド変異検索システム

1. 研究開始当初の背景

本研究の実施機関である筑波大学・蔬菜花卉学研究室では、トマトの実験モデル品種「マイクロトム」において化学変異誘発剤ならびにガンマ線照射による変異体集団を確立しており、これまでにその利活用を通して基礎科学と品種育成(応用)の両面で貢献してきた。前者については、単為結果性や機能性成分蓄積などトマトの重要形質に係る遺伝子の単離と機能解析に貢献してきた。特に、TILLING (Targeting Induced Local Lesions In Genomes) 等の手法によって逆遺伝学的に変異体を単離して遺伝子機能を解析できることから、広く国内外の研究に貢献してきた。また、後者については、民間企業を含め育種素材として利用の実績があり、近年では注目度が増している。このような背景から、マイクロトム変異体集団に存在する変異の実態をゲノムワイドに明らかにすることは、幅広い研究者と育種家にとって、興味のポイントであると考えられる。しかしながら、ゲノムワイドな変異解析については、全ゲノムを対象とした次世代シーケンズ解析(next-generation sequencing, NGS)が限られた数系統のマイクロトム変異体について実施されているのみである。NGSに関わるコストの点から集団レベルでの変異解析は行われておらず、未知な点が多く残っている。

2. 研究の目的

現状、逆遺伝学解析において解析対象遺伝子の変異体を取得する場合、TILLING や High-resolution melting (HRM)、ターゲット・リシーケンシング(対象遺伝子に限り PCR を行って NGS 解析を行う変異体単離手法)を用いることが一般的である。しかし、いずれの手法においても、研究者は遺伝子毎に PCR あるいは NGS 解析を実施して DNA 配列情報を解析する必要がある。この

ため実験毎に人手と時間を要し、必ずしも効率的とはいえない。むしろ、変異体集団に存在するゲノムワイドな変異情報をはじめに情報データベース化し、web ブラウザ等で簡便に検索できるアプリケーションを開発する方が、利便性・コストの両面から効率的であるといえる。何より、インターネットを介して世界中の研究者と育種家による利用を促進できる。一方、ゲノムワイドに変異同定を行う際の一般的手法である全ゲノム NGS では、膨大な費用が必要である。仮に 96 系統の変異体集団を解読する場合、2015 年現在の試算では、二千万円以上の経費を要する。したがって、集団レベルのゲノムワイド変異情報データベースを構築するには、よりコストパフォーマンスに優れた実験手法が必要である。エクソーム・シーケンシング(exome sequencing)は遺伝子領域のタンパク質コード配列(エクソン配列)に限り NGS 解読を実施する手法であり、全ゲノムを標的とする NGS 解読に比べてターゲット領域がエクソンに限定される。このため、全ゲノム NGS に比べて、コストパフォーマンスに優れている。トマトの場合、全ゲノム配列は約 950 Mb 前後と推定されるが、エクソン領域はこのうち約 50 Mb 程度である。したがって、エクソーム・シーケンシング法による 96 系統解読時のコストは、200~300 万円程度に抑えられる(2 度目のライブラリー調整と NGS 解読では、さらに低コストとなる)。しかしながら、市販のエクソームライブラリー調整キットでは、同時解読可能個体数は 24 が上限であるという制約がある。NGS のスループットが上がっている昨今では、さらに highplex な手法が最適である。そこで本研究では市販キットの実験プロトコルを改変し、一度に 96 個体をエクソーム解読するための研究手法開発を目的とした。また、取得したエクソーム変異情報を情報データベー

ス化し、web ブラウザ等で簡便に情報検索するためのインターネット・システム開発を目的とした。

3. 研究の方法

(1) 96plex エクソームライブラリーの調整

本研究では、95 系統のマイクロトム変異体と 1 系統の野生型系統 (コントロール) の計 96 系統を解析した。これらの系統は果実におけるカロテノイド含量や植物体形態、開花性にバリエーションを示す集団である。

本研究に先立ち、トマトのエクソームキャプチャー用プローブを民間企業の受託サービスを利用して合成済であった。実験では、はじめに Illumina NGS アダプターを両末端に付与したインサートサイズ 200 bp の DNA ライブラリーを変異体系統毎に作成して混合し、次いで、エクソームキャプチャー用プローブとターゲット配列をハイブリダイゼーションさせて濃縮した。この際、プローブとエクソン DNA の複合体は磁性体により回収されるが、イントロン DNA 等のノンターゲット配列は洗浄過程で分離され除去される。前述のように、市販キットでは一度のエクソームキャプチャーで扱えるマルチプレックス数上限が 24 である。非特異ハイブリダイゼーションを防ぐための HE オリゴ DNA (hybridization enhancing oligo DNA) もこれに合わせて市販されるが、24plex 用 HE オリゴは本研究の 96plex エクソームキャプチャーには利用できない。このため、独自に HE オリゴ DNA を設計した。96 系統の混合ライブラリーを一度に濃縮する方法と 48 系統ずつ 2 回に分けて濃縮する方法の 2 通りを検討し、定量 PCR による濃縮効率を調査した。最終的に遺伝子間の偏りが少ない後者の方法を採用した。5 遺伝子平均でターゲット濃縮効率は 14.7 倍であり、ゲノム 950 Mb に対してターゲット 50 Mb であることを考慮すると、妥当な濃

縮率であると判断された。

(2) Illumina Hiseq による 96plex エクソームライブラリーのシーケンシング

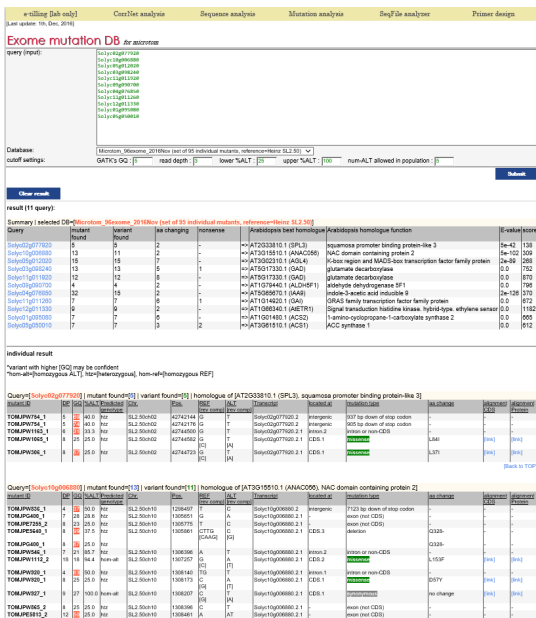
(1) で作成したエクソームライブラリーを Illumina Hiseq-2000 に供し、96 系統のショートリード情報を取得した (100 bp ペアエンド解析)。Hiseq による解析は Dual index モードで実施し、計 3 レーン分のデータを取得した。

4. 研究成果

Hiseq によるエクソーム解読後の総 DNA 配列長は、96 系統全体で 119 Gb、個体平均では 1.2 Gb (標準偏差は 0.2 Gb) であり、順当な結果であった。FASTQ 情報 (シーケンス・データ) はクオリティコントロール後、bowtie2 にてトマトのリファレンス・ゲノム配列情報にアライメントした。次いで、Genome analysis toolkit により、全エクソムの遺伝子変異情報を 96 系統個別に解読した。ターゲット領域 (エクソン配列) のカバー率は平均 98.4% であったことから、(1) の 96plex エクソームキャプチャーは成功と判断した。また、平均 read depth は 12.8 であり、DNA 変異検出に必要な最小限の情報が取得できた。一方、マイクロトムは野生型系統自体が完全な純系ではないため、上記で解読されるすべての DNA 配列の相違が変異であるとは限らない。むしろ、個体間に存在する intra-cultivar variation (品種内多型) である可能性も残存する。このため、全ゲノム NGS およびエクソーム・シーケンシングで解読した 9 系統の野生型マイクロトム個体をコントロールとして、変異体集団の変異情報をサブトラクションした。結果として、95 系統に存在する約 240,000 の変異を同定した (depth \geq 3、genotyping quality \geq 5)。これらのうち、98% はそれぞれの変異体に特有

であり、自然変異集団(natural variation)と異なって個体に固有な変異が集団内に多いことが明らかであった。さらに上記のうち、70,000以上はタンパク質アミノ酸配列の変化を引き起こす変異であり、約9,000は終止コドンの生成を伴うものであった。フレームシフトを生じる変異も約3,000存在した。さらに本研究では、これらのエクソン変異情報をSQLデータベース化し、webブラウザを介して検索可能なインターネット・アプリケーションを独自のLAMPサーバ上に開発した(論文公表と共に公開予定)。当該アプリケーションを利用すると、ユーザーは遺伝子IDから変異体と変異情報を取得できる(図1)。実際に、変異体における表現型が分かっている遺伝子について、当該アプリケーションにより新奇変異アレルを同定して解析したところ、期待される表現型が得られた。

図1. エクソーム変異データベースの検索画面



以上に述べたように、本研究ではトマトの96plex エクソーム・シーケンシング法を確立すると共に、95系統のマイクロトム変異体集団に存在するエクソン変異をゲノム

ワイドに解読した。さらに、情報データベースとwebアプリケーションを開発して変異体の *in silico* ワンクリック検索を可能にした。特に検索アプリケーションの開発は、幅広い研究者と育種家による変異体リソースの利活用を推進するものである。二度、エクソーム変異データベースを構築してしまえば、遺伝子毎に分子生物学実験やNGS解読を繰り返す必要がないため、研究や品種育成をより効率化することが可能である。当該プラットフォームは、あらゆる農作物にも応用可能であり、誘発変異集団の利活用法の新たなデファクトスタンダードになると期待される。

5. 主な発表論文等
〔雑誌論文〕(計0件)
〔学会発表〕(計0件)
〔図書〕(計0件)

〔産業財産権〕
出願状況(計0件)
取得状況(計0件)

〔その他〕
エクソーム変異データベース [限定公開]
<http://sosai-tools.agbi.tsukuba.ac.jp/cgi-bin/etilling.cgi>

6. 研究組織
(1)研究代表者
矢野 亮一 (YANO, Ryoichi)
筑波大学・生命環境系・助教
研究者番号: 00443044

(2)連携研究者
江面 浩 (EZURA, Hiroshi)
筑波大学・生命環境系・教授
研究者番号: 00332552

有泉 亨 (ARIIZUMI, Toru)
筑波大学・生命環境系・准教授
研究者番号: 70575381

星川 健 (HOSHIKAWA, Ken)
筑波大学・生命環境系・助教
研究者番号: 70634715