

平成28年度
食品安全確保総合調査

次世代シークエンサーの活用状況等に関する 調査

調査報告書

平成 29 年 3 月

一般財団法人 化学物質評価研究機構

目次

調査の概要	1
1. 調査の背景及び目的	2
2. 調査方法等	3
2.1. 有識者から構成される検討会の開催	3
2.2. 調査方法	4
3. 調査結果	7
3.1. 情報検索結果（検索ヒット数）	7
3.2. 海外における NGS の活用状況の把握	12
3.2.1. 遺伝子組換え植物等の分野における NGS データに基づく評価機関の評価状況	12
3.2.2. 文献ベースでの NGS の活用状況	30
3.2.3. 行政等における NGS 活用に向けた取り組み状況	41
3.3. 遺伝子組換え植物の安全性評価の際に一般的に用いられている分析法によるデータと NGS により作成されたデータの比較・検討	44
3.3.1. 次世代シーケンサー (NGS) の機器メーカー毎の特性	44
3.3.2. 遺伝子組換え植物の安全性評価に使用される従来法によるデータと NGS により作成されたデータの比較・検討	48
3.3.3. NGS データの遺伝子組換え植物の安全性評価における利用性	56
3.4. NGS データを遺伝子組換え植物の安全性評価に用いる際の課題の明確化	60
3.4.1. 課題 1: 遺伝子組換え植物の安全性評価に NGS データを活用する際のデータの品質や信頼性等を見極める基準がないことへの対応	60
3.4.2. 課題 2: 植物のゲノム構造の特殊性（同質倍数体、異質倍数体、繰り返し配列の多さ）への対応	61
3.4.3. 課題 3: 多サンプル解析を要する場合のコスト（費用及び時間）への対応	62
3.5. 今後期待される対応	67
NGS関連用語集	69
別添資料 3-1 収集・整理候補とした文献等	
別添資料 3-2 105 報の 400 字概要	
別添資料 3-3 重要と考えられた 22 報の詳細概要	
別添資料 3-4 サンガー法及び NGS に使用されている測定原理	
別添資料 3-5 課題解決に向けたヒントとなる事例	

調査の概要

近年、ゲノム全体の網羅的な解析を可能とする次世代シーケンサー（以下「NGS」という。）を用いた塩基配列決定法が急速に普及しているため、遺伝子組換え食品等の安全性評価に際しては、今後、NGSデータを用いた安全性評価の件数が増加することが予想される。そこで、本調査は、日本における遺伝子組換え植物の安全性評価に資するため、関連する科学文献等の収集・分析・整理により、以下を実施することを目的とした。

- 海外におけるNGSの活用状況の把握
- 遺伝子組換え植物の安全性評価の際に一般的に用いられている分析法によるデータとNGSにより作成されたデータの比較・検討
- 上記の調査結果を踏まえた、NGSを安全性評価に用いる際の課題の明確化

本調査では、5人の有識者から構成される検討会を3回開催して調査の実施方針及び調査内容の検討を行い、調査内容の取りまとめを行った。食品等に関する各国あるいは国際的な評価機関等及び文献データベース（BIOSIS Previews 及び Web of Science Core Collection）を検索対象とし、遺伝子組換え植物やNGS等に関わるキーワードを用いて文献等の調査を行い、最終的に105報の文献等を収集し400字程度の概要を作成した。また、その中から重要度が高いと考えられた22報についてA4で1枚程度の詳細な概要を作成するとともに、和訳を行った。

これらの情報に基づき、遺伝子組換え植物等の分野におけるNGSデータに基づく評価機関における評価状況、文献ベースでのNGSの活用状況、行政等におけるNGS活用に向けた取り組み状況を整理した。また、NGSの機器メーカー毎の特性を整理し、遺伝子組換え植物の安全性評価に一般的に用いられている分析法（以下、従来法）あるいはNGSにより取得されるデータの比較、検討を行った。具体的には、従来法としてサザンブロット及びPCR+サンガーシーケンス、NGS解析に用いる解析アプローチとして全ゲノムシーケンシング（WGS）、ターゲットキャプチャシーケンシング（TCS）及びロングリードシーケンシング（Long Read シーケンシング）について、遺伝子組換え植物の分子キャラクタリゼーションにおいて実施できること、検出感度、分析精度、網羅性、メリット、デメリット等について整理した。さらに、NGSデータの遺伝子組換え植物の安全性評価における利用性を文献的な知見に基づき整理した。これらの結果に基づき、NGSデータを遺伝子組換え植物の安全性評価に用いる際の課題を抽出し、今後期待される対応についてまとめた。

1. 調査の背景及び目的

近年、ゲノム全体の網羅的な解析を可能とする次世代シーケンサー（以下「NGS」という。）を用いた塩基配列決定法が急速に普及しているため、遺伝子組換え食品等の安全性評価に際しては、今後、NGSデータを用いた安全性評価の件数が増加することが予想される。

本調査は、日本における遺伝子組換え植物の安全性評価に資するため、関連する科学文献等の収集・分析・整理により、以下を実施することを目的とした。

- 海外におけるNGSの活用状況の把握
- 遺伝子組換え植物の安全性評価の際に一般的に用いられている分析法によるデータとNGSにより作成されたデータの比較・検討
- 上記の調査結果を踏まえた、NGSを安全性評価に用いる際の課題の明確化

2. 調査方法等

2.1. 有識者から構成される検討会の開催

表 2.1-1 に示す遺伝子組換え植物に関する有識者、及び NGS を含む遺伝子組換え植物に関連する分析法の有識者により構成される検討会を計 3 回開催し、表 2.1-2 に示す内容について協議を行った。なお、人選に際しては、有識者の専門領域が一方に偏ることが無いよう、内閣府食品安全委員会事務局と相談の上行った。

表 2.1-1 検討会を構成する有識者

氏名	所属
小関良宏	東京農工大学大学院工学研究院生命機能科学部門 教授
北川正成	タカラバイオ株式会社 執行役員（受託開発部・バイオメディカルセンター担当）
近藤一成	国立医薬品食品衛生研究所 生化学部 部長
田畑哲之	公益財団法人かずさ DNA 研究所 所長
中川英刀	国立研究開発法人理化学研究所統合生命医科学研究センター ゲノムシーケンス解析研究チーム チームリーダー

(五十音順、敬称略)

表 2.1-2 検討会での協議事項

第 1 回検討会 (2016 年 12 月 9 日)	● 本調査事業で収集・翻訳する情報の整理に関する調査方針・調査項目に関する協議
第 2 回検討会 (2017 年 2 月 3 日)	● 調査方針に基づき収集した情報の概要（論文のアブストラクトなど）を基に、日本の遺伝子組換え植物の安全性評価に資する文献のスクリーニングを実施し、より詳細に収集すべき情報（翻訳の有無等）を選定 ● 第1回検討会時に想定していなかった追加の調査項目等あれば、併せて検討
第 3 回検討会 (2017 年 3 月 10 日)	● 第2回検討会までの協議内容をふまえた調査報告書のとりまとめ ● 今後の検討事項の洗い出し等

2.2. 調査方法

本調査では、海外における NGS の活用状況の把握、遺伝子組換え植物の安全性評価の際に一般的に用いられている分析法によるデータと NGS により作成されたデータの比較・検討及び NGS を安全性評価に用いる際の課題の明確化を行うために、文献等の収集を行った。

文献等の収集は、1)「海外における NGS の活用状況の把握」のための諸外国での NGS を用いたデータ解析を含む科学文献、2)「従来法によるデータとの比較・検討」のための NGS とその他の分析法（サザンブロット分析等）との比較を含む科学文献、の 2 つに分けて行った。調査対象は各国あるいは国際的評価機関等における報告書等及びデータベースに収載された科学技術論文とした。キーワードによる検索で文献等を抽出した。なお、評価機関等、データベース及びキーワードの選定においては、検討会における有識者の意見を踏まえた。

調査対象とした評価機関等を表 2.2-1 に、文献等の検索対象とした文献データベースを表 2.2-2 に示す。調査に使用したキーワードを表 2.2-3 に示す。

また、評価機関が作成した評価書等において引用されている文献、検討会有識者から提供を受けた文献等も調査対象の候補とした。

収集した文献等については、検討会有識者との協議により 100 報程度を選定し、1 報あたり 400 字程度の概要（以下、400 字概要）を作成した。さらに、400 字概要に基づき、検討会有識者との協議により重要と考えられる文献等を 20 報程度選定し、目的、方法、結果、考察等を含む A4 で 1 枚程度の詳細な概要を作成した（以下、詳細概要）。なお、詳細概要を作成した文献等については、全文を和訳した。

表 2.2-1 調査対象とした評価機関等

- 国際連合食糧農業機関 (Food and Agriculture Organization of the United Nations (FAO))
- コーデックス委員会 (Codex Alimentarius Commission (CAC))
- 欧州委員会 (European Commission (EC))
- 欧州食品安全機関 (European Food Safety Authority (EFSA))
- 米国食品医薬品庁 (Food and Drug Administration (FDA))
- 米国環境保護庁 (U.S. Environmental Protection Agency (EPA))
- アメリカ合衆国農務省 (United States Department of Agriculture (USDA)) *
- 仏食品環境労働衛生安全庁 (ANSES)
- 独連邦リスク評価研究所 (BfR)
- 独連邦自然保護庁 (German Federal Agency for Nature Conservation (BFN)) *
- カナダ保健省 (Health Canada)
- カナダ食品検査庁 (Canadian Food Inspection Agency (CFIA))
- オーストラリア・ニュージーランド食品基準機関 (Food Standard Australia New Zealand (FSANZ))
- 経済協力開発機構 (OECD)

表 2.2-2 調査対象とした文献データベース

- BIOSIS Previews
1969 年以降の動物学・植物学・遺伝学・バイオテクノロジー・薬学・農学等の生命科学分野全般の文献情報。収録対象には特許データも含まれる。
- Web of Science Core Collection
科学技術分野(1900 年～)、社会科学分野及び人文科学分野の主要な学術雑誌およそ 8,500 誌、1990 年以降の世界の重要会議・シンポジウム・セミナー・専門家会議・ワークショップ・学会及び総会・大会等で発行された文献情報が含まれる。

表 2.2-3 調査に使用した検索キーワード

分野	検索キーワード
NGS 関連	(“next generation sequencing” OR “whole genome sequencing” OR “exome capture NGS” OR “Junction Sequence Analysis” OR JSA OR “T-DNA backbone sequence” OR Illumina OR PacBio OR “Roche 454” OR “Southern-by-Sequencing” OR “De novo sequencing” OR “Sanger sequencing” OR “Target capture sequencing” OR Nanopore OR “Plasmid Backbone”)
遺伝子組換え 関連	AND (“genetically modified” OR “genetically engineered” OR transgenic OR “gene* recombination” OR GMP)
植物関連*	AND (Plant OR crop OR canola OR “oilseed rape” OR “rape oilseed” OR rapeseed OR “rape seed” OR coleseed OR cotton OR maize OR corn OR wheat OR soybean OR Cantaloupe OR gum OR potato OR rice OR “starch potato” OR “sugar beet” OR Flower OR Alfalfa OR apple OR papaya OR flax OR plum OR radicchio OR squash OR tomato OR carnation)
手法の比較に関する情報への絞込みを行う際に使用したキーワード	
分析法関連	AND (PCR OR “polymerase chain reaction” OR “Northern blot*” OR “Southern blot*”) AND/OR Validat*

3. 調査結果

3.1. 情報検索結果（検索ヒット数）

「2.2.調査方法」で示した調査の結果を表 3.1-1 に示す。

○は用いたキーワードを示し、例えば 1) では「NGS 関連（のキーワード）」かつ（and）「遺伝子組換え関連（のキーワード）」で検索した結果、BIOSIS Previews では 380 件、Web of Science Core Collection では 362 件ヒットしたことを示す。

表 3.1-1 機関等及び文献データベースのヒット件数

		1) 海外における NGS 活用状況		2) 従来法によるデータとの比較・検討				
キ ー ワ ー ド	NGS 関連	○	○	○	○	○	○	○
	遺伝子組換え関連	○	○			○	○	
	植物関連		○				○	○
	分析手法関連			○	○	○	○	○
	Validat*				○	○	○	○
機関等			1547			811		
→重複を除くと 432								
文 献 D B	BIOSIS Previews	380	134	6595	1102	20	8	239
	Web of Science Core Collection	362	146	5164	1037	22	12	207
		→太字分 (726) の重複を除くと 299						

単位：件数

機関等については 1) の 1547 件、2) の 811 件の中から重複を除く 432 件を対象とした。

文献等については、キーワードの組み合わせによっては抽出される文献数が過多の場合、過小の場合があった。最終的に 100 報程度を抽出するという目的に照らして、表中の太字を候補とし、重複を除いた 299 件を対象とした。

さらには評価機関が作成した評価書等において引用されている文献、検討会委員会ら提供を受けた文献等も対象とし、それぞれのアブストラクトの内容等を確認して本調査に関連のある文献等を抽出した。

その結果、機関等の文書 54 件、文献 77 件の合計 131 件となった。リストを別添資料 3-1 に示す。

抽出した文献等のリストに関する検討会有識者との協議により、査読対象とし、400 字概要を作成する文献等は、機関等の文書 43 件、文献 62 件の合計 105 件となった（別添資料 3-1）。作成した 400 字概要を別添資料 3-2 に示す。

さらに、400 字概要に基づく検討会有識者との協議により重要度が高いと考えられる 22 報の文献等を決定した（別添資料 3-1）。タイトル、著者名（筆頭著者所属）及び掲載誌等を表 3.1-2 に示す。22 報について目的、方法、結果、考察等を含む A4 で 1 枚程度の詳細概要を作成した（別添資料 3-3）。また、和訳（全訳）を行った。

表 3.1-2 重要と考えられた文献等 22 報

文献 ID	文献/評価書等	タイトル	著者名 [筆頭著者所属]	掲載誌等
ID-002	文献	Identification of Genomic Insertion and Flanking Sequence of G2-EPSPS and GAT Transgenes in Soybean Using Whole Genome Sequencing Method	Guo B, Guo Y, Hong H, Qiu LJ. [The National Key Facility for Crop Gene Resources and Genetic Improvement (NFCRI) and MOA Key Lab of Soybean Biology (Beijing), Institute of Crop Science, Chinese Academy of Agricultural Sciences Beijing, China]	Front Plant Sci. 2016;7:1009. doi: 10.3389/fpls.2016.01009.
ID-003	文献	Statistical framework for detection of genetically modified organisms based on Next Generation Sequencing	Willems S, Fraiture MA, Deforce D, De Keersmaecker SC, De Loose M, Ruttink T, Herman P, Van Nieuwerburgh F, Roosens N. [Scientific Institute of Public Health (WIV-ISP), Platform of Biotechnology and Molecular Biology (PBB), J. Wytsmanstraat 14, 1050 Brussels, Belgium; Scientific Institute of Public Health (WIV-ISP), Biosafety and Biotechnology Unit (SBB), J. Wytsmanstraat 14, 1050 Brussels, Belgium; University of Gent (UGent), Faculty of Pharmaceutical Sciences, Laboratory of Pharmaceutical Biotechnology, Harelbekestraat 72, 9000 Ghent, Belgium.]	Food Chem. 2016;192:788-98. doi: 10.1016/j.foodchem.2015.07.074.
ID-004	文献	Molecular Characterization of Transgenic Events Using Next Generation Sequencing Approach	Guttikonda, SK; Marri, P; Mammadov, J; Ye, L; Soe, K; Richey, K; Cruse, J; Zhuang, MB; Gao, ZF; Evans, C; Rounsley, S; Kumpatla, SP [Dow AgroSci LLC, 9330 Zionsville Rd, Indianapolis, IN 46268 USA]	PLoS One. 2016 ;11(2):e0149515. doi: 10.1371/journal.pone.0149515.
ID-006	文献	CONTRAILS: A tool for rapid identification of transgene integration sites in complex, repetitive genomes using low-coverage paired-end sequencing	Lambirth KC, Whaley AM, Schlueter JA, Bost KL, Piller KJ [Department of Biological Sciences, University of North Carolina at Charlotte, Charlotte, NC 28223, United States]	Genom Data. 2015;6:175-81. doi: 10.1016/j.gdata.2015.09.001.
ID-007	文献	Next-generation sequencing is a robust strategy for the high-throughput detection of zygosity in transgenic maize	Fritsch L, Fischer R, Wambach C, Dudek M, Schillberg S, Schröper F [Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Forckenbeckstrasse 6, 52074, Aachen, Germany]	Transgenic Res. 2015;24(4):615-23. doi: 10.1007/s11248-015-9864-x.

文献 ID	文献/評 価書等	タイトル	著者名 [筆頭著者所属]	掲載誌等
ID-009	文献	HtStuf: High-Throughput Sequencing to Locate Unknown DNA Junction Fragments	Kanizay LB, Jacobs TB, Gillespie K, Newsome JA, Spaid BN, Parrott WA [Center for Applied Genetic Technologies, Univ. of Georgia, Athens, GA 30602]	Plant Genome. 2015;8(1).doi: 10.3835/plantgenome2014.10.0070
ID-010	文献	Southern-by-Sequencing: A Robust Screening Approach for Molecular Characterization of Genetically Modified Crops	Zastrow-Hayes GM, Lin HN, Sigmund AL, Hoffman JL, Alarcon CM, Hayes KR, Richmond TA, Jeddeloh JA, May GD, Beatty MK [DuPont Pioneer, 7300 NW 62nd Ave., Johnston, IA 50131]	Plant Genome. 2015;8(1).doi: 10.3835/plantgenome2014.08.0037
ID-011	文献	Efficiency to Discovery Transgenic Loci in GM Rice Using Next Generation Sequencing Whole Genome Re-sequencing	Park D, Kim D, Jang G, Lim J, Shin YJ, Kim J, Seo MS, Park SH, Kim JK, Kwon TH, Choi IY [Institute of Green Bio Science and Technology, Seoul National University, Pyeongchang 25354, Korea]	Genomics Inform. 2015;13(3):81-5. doi: 10.5808/GI.2015.13.3.81.
ID-015	文献	Detecting authorized and unauthorized genetically modified organisms containing vip3A by real-time PCR and next-generation sequencing	Liang C, van Dijk JP, Scholtens IM, Staats M, Prins TW, Voorhuijzen MM, da Silva AM, Arisi AC, den Dunnen JT, Kok EJ [RIKILT Wageningen UR, P.O. Box 230, 6700 AE, Wageningen, The Netherlands]	Anal Bioanal Chem. 2014;406(11):2603-11. doi: 10.1007/s00216-014-7667-1.
ID-018	文献	A survey of tools for variant analysis of next-generation genome sequencing data	Pabinger S, Dander A, Fischer M, Snajder R, Sperk M, Efremova M, Krabichler B, Speicher MR, Zschocke J, Trajanoski Z [Division for Bioinformatics, Innsbruck Medical University, Innrain 80, 6020 Innsbruck, Austria]	Brief Bioinform. 2014;15(2):256-78. doi: 10.1093/bib/bbs086.
ID-019	文献	The Use of Next Generation Sequencing and Junction Sequence Analysis Bioinformatics to Achieve Molecular Characterization of Crops Improved Through Modern Biotechnology	Kovalic, D; Garnaat, C; Guo, L; Yan, YP; Groat, J; Silvanovich, A; Ralston, L; Huang, MY; Tian, Q; Christian, A; Cheikh, N; Hjelle, J; Padgett, S; Bannon, G [Monsanto Company, 800 N. Lindbergh Blvd., St. Louis, MO 63167]	Plant Genome. 2012; 5(3):149-163. doi: 10.3835/plantgenome2012.10.0026
ID-022	文献	Application of whole genome shotgun sequencing for detection and characterization of genetically modified organisms and derived products	Holst-Jensen A, Spilsberg B, Arulandhu AJ, Kok E, Shi J, Zel J [Norwegian Veterinary Institute, Ullevaalsveien 68, P.O. Box 750, Sentrum, 0106, Oslo, Norway]	Anal Bioanal Chem. 2016;408(17):4595-614. doi: 10.1007/s00216-016-9549-1.
ID-032	文献	High-Throughput Analysis of T-DNA Location and Structure Using Sequence Capture	Inagaki S, Henry IM, Lieberman MC, Comai L [Plant Biology Department and Genome Center, University of California Davis, Davis, California, United States of America; Department of Integrative	PLoS One. 2015;10(10):e0139672. doi:

文献 ID	文献/評 価書等	タイトル	著者名 [筆頭著者所属]	掲載誌等
			Genetics, National Institute of Genetics, Mishima, Japan]	10.1371/journal.pone.0139672.
ID-033	文献	A Comparison of transgenic and wild type soybean seeds: analysis of transcriptome profiles using RNA-Seq	Lambirth KC, Whaley AM, Blakley IC, Schlueter JA, Bost KL, Loraine AE, Piller KJ [Department of Biological Sciences, University of North Carolina at Charlotte, Charlotte, NC, 28223, USA]	BMC Biotechnol. 2015;15:89. doi: 10.1186/s12896-015-0207-z.
ID-034	文献	Next-generation sequencing as a tool for the molecular characterisation and risk assessment of genetically modified plants: Added value or not?	Pauwels K, De Keersmaecker SCJ, De Schrijver A, du Jardin P, Roosens NHC, Herman P [Scientific Institute of Public Health, Biosafety and Biotechnology Unit, J. Wytsmanstraat 14, B-1050 Brussels, Belgium]	Trends in Food Sci & Tech. 2015; 45(2):319–326.
ID-042	文献	Comparative whole-genome analyses of selection marker-free rice-based cholera toxin B-subunit vaccine lines and wild-type lines	Kashima K, Mejima M, Kurokawa S, Kuroda M, Kiyono H, Yuki Y [Division of Mucosal Immunology, Department of Microbiology and Immunology, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan]	BMC Genomics. 2015;16:48. doi: 10.1186/s12864-015-1285-y.
ID-043	文献	Current and New Approaches in GMO Detection: Challenges and Solutions	Fraiture MA, Herman P, Taverniers I, De Loose M, Deforce D, Roosens NH [Platform of Biotechnology and Molecular Biology (PBB) and Biosafety and Biotechnology Unit (SBB), Scientific Institute of Public Health (WIV-ISP), J. Wytsmanstraat 14, 1050 Brussels, Belgium ; Technology and Food Sciences Unit, Institute for Agricultural and Fisheries Research (ILVO), Burg. Van Gansberghelaan 115, Bus 1, 9820 Merelbeke, Belgium ; Laboratory of Pharmaceutical Biotechnology, Faculty of Pharmaceutical Sciences, Ghent University, Ottergemsesteenweg 460, 9000 Ghent, Belgium]	Biomed Res Int. 2015;2015:392872. doi: 10.1155/2015/392872.
ID-068	文献	Effect of the down-regulation of the high Grain Protein Content (GPC) genes on the wheat transcriptome during monocarpic senescence	Cantu D, Pearce SP, Distelfeld A, Christiansen MW, Uauy C, Akhunov E, Fahima T, Dubcovsky J [Department of Plant Sciences, University of California Davis, USA]	BMC Genomics. 2011;12:492. doi: 10.1186/1471-2164-12-492.
ID-075	文献	A proposed regulatory framework for genome-edited crops	Huang S, Weigel D, Beachy RN, Li J. [Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China.]	Nat Genet. 2016;48(2):109-11. doi: 10.1038/ng.3484.

文献 ID	文献/評価書等	タイトル	著者名 [筆頭著者所属]	掲載誌等
ID-130	文献	Characterization of GM events by insert knowledge adapted re-sequencing approaches	Yang L, Wang C, Holst-Jensen A, Morisset D, Lin Y, Zhang D. [Collaborative Innovation center for biosafety of GMOs, National Center for Molecular Characterization of GMOs, School of Life Science and Biotechnology, Shanghai Jiao Tong University. 800 Dongchuan Road, Shanghai 200240. P. R. China]	Sci Rep. 2013;3:2839. doi: 10.1038/srep02839
ID-131	評価書等	HIGH-THROUGHPUT DNA SEQUENCING IN THE SAFETY ASSESSMENT OF GENETICALLY ENGINEERED PLANTS: PROCEEDINGS OF THE OECD WORKSHOP (April 2016)	OECD	OECD Environment, Health and Safety Publications Series on the Safety of Novel Foods and Feeds No. 29.2016
ID-132	文献	Use of microarray hybrid capture and next-generation sequencing to identify the anatomy of a transgene.	Dubose AJ, Lichtenstein ST, Narisu N, Bonnycastle LL, Swift AJ, Chines PS, Collins FS. [Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA]	Nucleic Acids Res. 2013;41(6):e70. doi: 0.1093/nar/gks1463.

3.2. 海外における NGS の活用状況の把握

3.2.1. 遺伝子組換え植物等の分野におけるNGSデータに基づく評価機関の評価状況

① 遺伝子組換え植物分野における NGS データに基づく評価機関の評価状況

本調査の結果、NGS データを含む申請が行われた遺伝子組換え植物として、Monsanto 社による遺伝子組換えトウモロコシ及び遺伝子組換えダイズ、Okanagan Specialty Fruits 社による遺伝子組換えリンゴがあり、本調査の対象国における、これらの遺伝子組換え植物の食品の認可状況を ISAAA¹（Information Service for the Acquisition of Agri-Biotech Applications）のデータベースにより確認した（表 3.2.1-1）。アメリカ合衆国農務省（United States Department of Agriculture、USDA）及びオーストラリア・ニュージーランド食品基準機関（Food Standard Australia New Zealand、FSANZ）については申請書類及び評価結果が公開されており、申請内容の確認ができたが、カナダ食品検査庁（Canadian Food Inspection Agency、CFIA）については文書等が公表されておらず、内容確認をすることはできなかった。また、仏食品環境労働衛生安全庁（ANSES）では MON87411 について安全性に関する評価が行われていた（ただし、現時点では未認可²）。

評価内容を確認できたいずれの遺伝子組換え植物においても、NGS データは分子キャラクタリゼーションに用いられていた。

以下に、評価時点が比較的古い MON87411 の事例、最新の評価として MON87419 の事例、及び Okanagan Specialty Fruits 社による遺伝子組換えリンゴ NF872 の事例を以下に示す。

なお、ISAAA で確認した結果、MON87411 及び MON87751 については、韓国及び台湾においても既に承認されていた。

¹ <http://www.isaaa.org/gmaprovaldatabase/event/default.asp?EventID=409>

² ANSES で認可されていないのは、安全性に関する結論が出せなかったためで、NGS データに基づくものではないが、分子キャラクタリゼーションが不完全であったことが理由である。

表 3.2.1-1 遺伝子組換え植物分野における NGS データに基づく評価が行われた事例

		USDA	CFIA	FSANZ	ANSES	備考
トウモロコシ	MON 87403	2015 [ID-082]	2015※	2016 [ID-081]	-	
	MON87411	2014 [ID-079]	-	2015 [ID-080]	2015# 認可されず [ID-078]	食品安全委員会 では 2016 年に評 価済み ^{\$}
	MON87419	2016 [ID-077]	2016※	2016 [ID-076]	-	食品安全委員会 では 2017 年に評 価済み ^{\$}
ダイズ	MON87751	2014 [ID-084]	2014※	2016 [ID-083]	-	食品安全委員会 では 2016 年に評 価済み ^{\$}
リンゴ	NF872	2016 [ID-087]	-	-	-	-

表中の数字は認可された年

※評価結果は閲覧できず³、

#評価された年を記載している、

^{\$}食品安全委員会では、「ヒトの健康を損なうおそれはないと判断」している。

[] ID 番号は本調査で収集した文献等の番号（別添資料 3-1 参照）

② MON87411（コウチュウ目害虫に対する抵抗性及び除草剤グリホサートに対する耐性トウモロコシ）

MON87411 はコウチュウ目害虫に対する抵抗性及び除草剤グリホサートに対する耐性トウモロコシであり、USDA（2014, [文献 ID-079]）、FSANZ（2015, [文献 ID-080]）及び ANSES（2015, [文献 ID-078]）において評価が行われている。表 3.2.1-2 に MON87411 トウモロコシの基礎情報を示す。

³ <http://www.hc-sc.gc.ca/fn-an/gmf-agm/appro/index-eng.php>

表 3.2.1-2 MON87411 の基礎情報

宿主	トウモロコシ品種 LH244
導入遺伝子及びその由来	Agrobacterium sp. strain CP4 由来の <i>cp4 epsps</i> 遺伝子 (1,368 nt) 除草剤グリホサートに対する耐性を付与するタンパク質をコードする <i>cp4 epsps</i> 遺伝子
	<i>Bt</i> subsp. <i>Kumamotoensis</i> 由来の <i>cry3Bb1</i> 遺伝子 (1,962 nt) 昆虫の腸に直接的な悪影響をもたらす Cry3Bb1 タンパク質の発現を通じて、corn rootworm (CRW) に対する耐性を与える
	western corn rootworm (<i>Diabrotica virgifera virgifera</i>)由来の <i>DvSnf7</i> 遺伝子 (240 nt) 導入カセットに逆方向の繰り返し配列で 2 つの <i>DvSnf7</i> 断片が含まれており、二本鎖 RNA (dsRNA) 転写物が生成する。dsRNA がトウモロコシ根虫 (CRW) 昆虫に摂取されると、CRW <i>DvSnf7</i> 遺伝子の RNA 干渉 (RNAi) を引き起こし、昆虫は死に至る。
遺伝子の導入方法	LH244 の未成熟胚に 3 種のカセットが入った PV-ZMIR10871 をアグロバクテリウム媒介形質転換法により導入

申請者が MON87411 の分子キャラクタリゼーションに用いた方法のフローを図 3.2.1-1 に示す。

MON87411 では、分子キャラクタリゼーションのインサート数の確認及び遺伝子の安定性評価に、NGS 及び Junction Sequence Analysis (JSA) (NGS/JSA) のデータが用いられている。インサートの配列決定、挿入遺伝子の構成・配列変化の有無 (intactness)・コピー数、バックボーンシーケンスの有無、インサートサイトの完全性・構成については、直接シーケンシング法が用いられている。なお、NGS/JSA により確認されたインサート数については、直接シーケンシング法によるインサートの構成及び配列の解析からも確認されている。

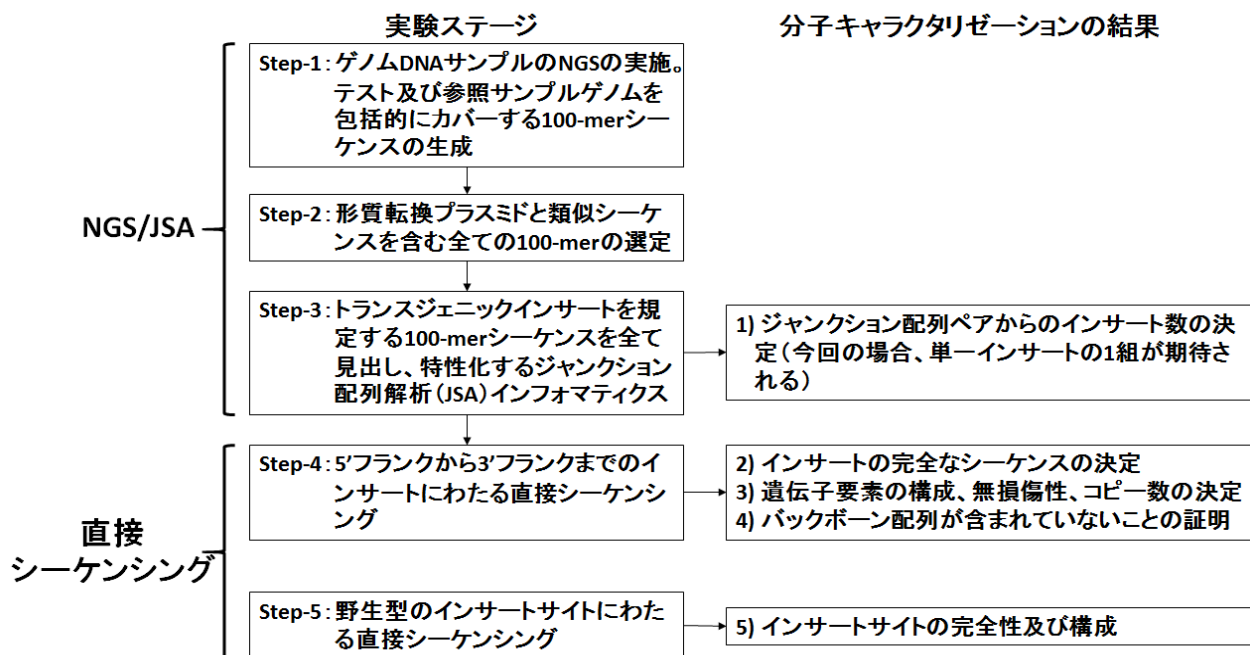


図 3.2.1-1 申請者が使用した MON87411 の分子キャラクタリゼーションのフロー
(FSANZ (2015) の図を改変)

(注；図中に遺伝子の安定性評価に関する記載がないが、遺伝子の安定性評価も
NGS/JSA により実施されている。)

FSANZ (2015) の評価書に基づき、NGS の実験及びデータ解析ならびに JSA の結果詳細
を以下に示す。

■NGS 実験

- ・ MON87411 の第 4 世代及び親株である LH244 (コントロールとして) の全ゲノム DNA を使用。リファレンス DNA には PV-ZMIR10871 を使用。
- ・ 陽性対象として、プラスミド DNA を LH244 の DNA にスパイク (感度確認のために、コピーゲノム等量比で 1 及び 1/10 をスパイク)。
- ・ DNA は約 325bp に切断し、アダプターを連結後、10 サイクルの PCR を行い、Illumina 社の Hiseq を用いた 100 bp のショートリードによるシーケンスを実施。

■JSA を含むデータ解析

- ・ 100 mer のリード長のみを解析。
- ・ 全サンプルの 100 mer のリードを既知の単一コピーの内因性コーンピルビン酸デカルボキシラーゼ遺伝子 (pdc3) にマッピングすることで、効果的なカバレッジの深さを確認した。→この結果 pdc3 はサンプル毎に $> \times 10^7$ で 100-mer によってカバーされており、網羅的なカバレッジであると考えられた (Kovalic ら、

2012⁴)。

- ・ さらに、BLAST アルゴリズムにより PV-ZMIR10871 プラスミドと類似性のあ
る配列を含む 100-mer の解析を実施（この解析により挿入プラスミド配列に完
全に一致する、あるいは、プラスミド配列及び接合配列の両方を含む 100mer
のリードをすべて見出した）。
- ・ $1e^{-5}$ 未満の e 値であり、形質転換プラスミドとの 96.7%以上の同一性がある少
なくとも 30 塩基の一致を有するリードを収集（Kovalic ら, 2012）。
- ・ Bowtie ショートシーケンスアラインメントソフトウェアを使用して、30nt 以上
の重複しないリードを収集。
- ・ Novoalign ソフトウェアによりアダプター配列を除去し、Phred スコアが 12 未
満の低品質のリードを除去
- ・ ジャンクション領域の配列を見つけるために、PV-ZMIR10871 プラスミドに残
りのリードをアライメントし、内因性相同体由来のリードを除くために、コン
トロールゲノムに対してもリードのアライメントを実施

■JSA の結果（図 3.2.1-2）

- ・ 図 3.2.1-2 では各ジャンクション配列のリードは 30 nt のプラスミド配列のみを
含むようトリミングされている。
- ・ T-DNA 及びジャンクション配列の両方を含む 2 つのユニークなジャンクション
配列クラスが検出されていることから、単一の DNA 挿入であることが示され
た。なお、LH244 由来の DNA ではジャンクション配列は見出されなかった。

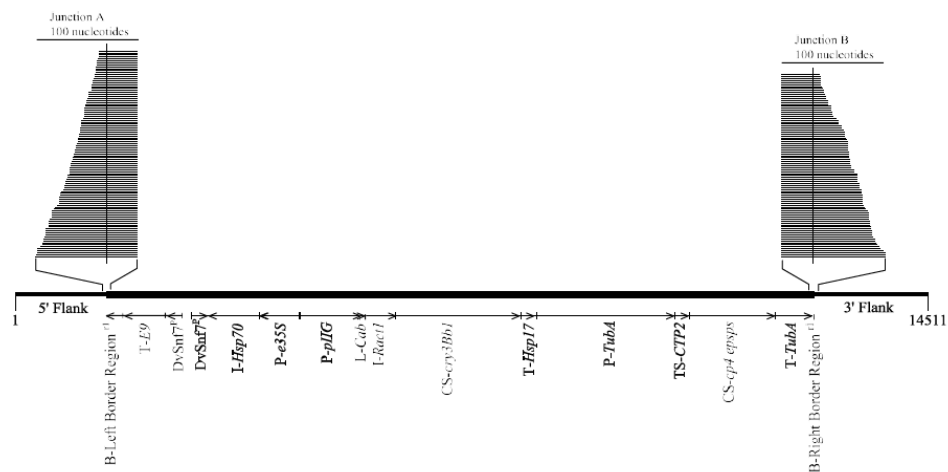


図 3.2.1-2 MON87411 において検出されたジャンクション配列

⁴ 本調査で収集した ID-019 の文献

■遺伝的安定性の確認

- 以下に示すゲノム DNA を用いて、上述の方法で実施した NGS/JSA 解析を実施

DNA の由来	世代	備考
MON87411	R_4	
	R_5	
	R_6	
	R_4F_1	
	R_5F_1	
LH244	-	親株
NL6169	-	従来種
MPA640B	-	従来種

■遺伝的安定性の結果

- コントロールである LH244、NL6169 及び MPA640B においてジャンクション配列は検出されていない。
- MON87411 のすべての世代の DNA において、図 3.2.1-2 で特定された二つの同一のジャンクション配列のみの存在が確認された。
- 全世代においてジャンクション配列のデータに一貫性があることから、MON87411 において、単一のインサートが安定的に維持されていることが示された。

MON87411 についての USDA (2014)、FSANZ (2015) 及び ANSES (2015) による NGS/JSA の解析結果に関する見解を表 3.2.1-3 に示す。

挿入遺伝子の数については、いずれの機関においても、NGS/JSA の結果に基づき評価が行われていた。プラスミドバックボーンの有無については USDA 及び ANSES とも NGS/JSA の結果を採用しているが、FSANZ では直接シーケンシングの結果に基づき評価されている。また、挿入遺伝子の世代間安定性については、USDA 及び FSANZ において NGS/JSA の結果に基づく評価を行っているが、ANSES では分離比解析の結果に基づく評価のみが評価書には記載されている。なお、NGS/JSA の結果に基づく評価を実施していない理由などは評価書には記載されていない。

表 3.2.1-3 MON87411 の NGS/JSA の解析結果に関する 3 つの評価機関の見解

項目	USDA (2014, [文献 ID-079])	FSANZ (2015, [文献 ID-080])	ANSES (2015, [文献 ID-078])
挿入遺伝子の数	■ <u>NGS/JSAの結果から</u> ；単一の完全なT-DNA PVZMIR 10871がMON 87411に挿入されている。	■ <u>NGS/JSAの結果から</u> ； MON 87411 には単一の DNA インサートが含まれている。	■ <u>NGS/JSAの結果から</u> ； 3 つのカセットの intact コピーを含む T-DNA の単一挿入であった。
プラスミドバックボーンの有無	■ <u>NGS/JSAの結果から</u> ； PVZMIR 10871のバックボーン配列は挿入されていない。	■ <u>直接シーケンシング結果から</u> ； MON 87411 にプラスミドバックボーンがないことを示している。 <u>また、この解析は、T-DNA の単一コピーが挿入されているとの NGS/JSA の結論を確認した。</u>	■ <u>NGS/JSAの結果から</u> ； プラスミドベクターに由来する配列は無い
挿入遺伝子の世代間安定性	■ <u>NGS/JSAの結果から</u> ；導入遺伝子の安定性が確認されている（※）。	■ <u>NGS/JSAの結果から</u> ；試験した全ての世代にわたってジャンクション配列の結果が一致していたことから、単一インサートは MON 87411 に安定的に維持されている。	■ <u>分離比解析の結果から</u> ；インサートは安定であり、メンデルの法則に従って、遺伝している。
全体の結論	食品・飼料としての安全性の懸念は指摘されなかった。	公衆衛生上及び安全上の懸念事項は確認されていない。MON87411 由来の食物は、人による消費に関して従来のトウモロコシ品種由来の食物と同様に安全であると考えられる。	安全性についてコメントできない。

※評価本文中において、導入遺伝子の安定性について、「PCR 及び DNA シーケンシング解析」によるデータと記載されているが、参照しているデータは NGS/JSA のものであったことから、NGS/JSA のデータにより導入遺伝子の安定性は判断しているものと考えられる。

USDA 及び FSANZ は NGS データに基づく評価以外の分子キャラクタリゼーションの結果や、その他の項目の評価結果から、いずれも安全性の懸念はないとしている。一方、ANSES では NGS データに基づく評価結果によるものではないが、分子キャラクタリゼーションが以下の点で不十分であるとして、安全性についてコメントできないと結論している (ANSES, 2015)。

- ・ トウモロコシ及び MON87411 トウモロコシの DvSnf7 遺伝子に対する Cry3Bb1 及び CP4 EPSPS のタンパク質及び RNA の量の解析において比較に用いられる NL6169 遺伝子型に関する情報の欠如がある。
- ・ GMP 及びこの GMP を消費する動物における副作用（「標的外効果」）のリスクの評価が、遺伝子 DvSnf7 に対して向けられた RNA 配列を用いて特異的に行われていない。

③ MON87419 (グルホシネート除草剤耐性及びジカンバ除草剤耐性トウモロコシ)

MON87419 はグルホシネート除草剤耐性及びジカンバ除草剤耐性を目的に遺伝子組換えが行われたトウモロコシであり、USDA (2016、[文献 ID-077])、CFIA (2016) 及び FSANZ (2016、[文献 ID-076]) による評価が行われている。CFIA (2016) の評価内容については入手できなかったことから、以下は USDA 及び FSANZ の評価に関する内容を示す。

MON87419 トウモロコシの基礎情報を表 3.2.1-4 に示す。

表 3.2.1-4 MON87419 の基礎情報

宿主	Corn (<i>Zea mays</i>) : LH244
導入遺伝子及びその由来	細菌 <i>Stenotrophomonas maltophilia</i> 株 DI-6 由来 <i>dmo</i> 遺伝子 (ジカンバ除草剤耐性)
	細菌種 <i>Streptomyces viridochromogenes</i> 由来 <i>pat</i> 遺伝子 (グルホシネート除草剤耐性)
遺伝子の導入方法	PV-ZMHT507801 プラスミド (T-DNA I にセレクション用の <i>cp4epsps</i> 、T-DNA II に 2 つの除草剤耐性遺伝子を含む) を用いたアグロバクテリウムによる形質転換

申請者が分子キャラクタリゼーションに用いた方法のフローを図 3.2.1-3 に示す。

NGS データの利用は、前述の MON87411 ではインサート及びコピー数の確認のみであった。しかし、MON87419 では、インサート及びコピー数の確認に NGS/JSA が用いられるだけでなく、バックボーンシーケンスの有無確認に NGS で取得したリードのマッピング手法を適用している。

インサートの配列決定、挿入遺伝子の構成・配列変化の有無 (intactness)、インサートサイトの完全性・構成については、直接シーケンシング法が用いられている。

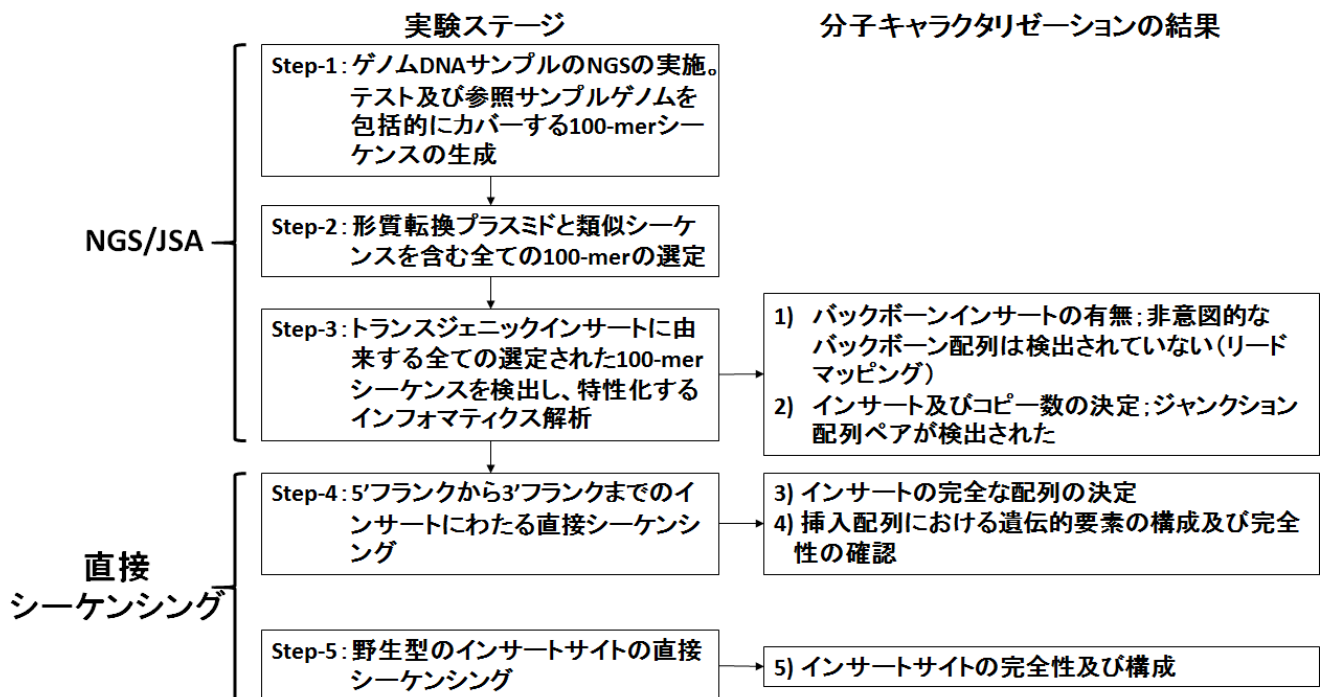


図 3.2.1-3 申請者が使用した MON87419 の分子キャラクタリゼーションのフロー (FSANZ (2016) の図を改変)

FSANZ (2016) の評価書に基づき、NGS の実験及びデータ解析ならびに JSA の結果詳細を以下に示す。

■NGS/JSA の方法

- ・ MON87419 の 5 つの世代 (R₃, R₄, R₅, R₃F₁ 及び R₄F₁) 及び親株 (LH244) の種子からゲノム DNA を単離。
- ・ 陽性対象として、プラスミドベクターPV-ZMHT507801 からの参照 DNA を使用。
- ・ NGS 解析における感度解析のために、LH244 のゲノム DNA をプラスミド DNA にスパイク。
- ・ インサート確認は R₃ 世代のゲノム DNA を用いた NGS/JSA により評価した。
- ・ インサートサイトの数の決定には、全プラスミドベクター配列 (T-DNA I、T-DNA II 及びベクターバックボーン) をクエリーとした。

■NGS/JSA 解析の結果

- ・ NGS/JSA の解析により 2 つのユニークなジャンクション配列が特定され、いずれのジャンクション配列クラスも、ゲノム隣接配列に連結された T-DNA I 境界配列を含み、PV-ZMHT507801 からの T-DNA I の挿入を示している。
- ・ MON87419 及び対照から得られたリードを PV-ZMHT507801 配列にマッピングした結果から、MON87419 に存在するプラスミドバックボーン配列がないことも示

された。

- ・ ジャンクション配列クラスの挿入/連結配列への完全なアラインメントにより、MON87419 に単一のゲノム遺伝子座にひとつの T-DNA I が挿入されていることが確認された（図 3.2.1-4）。

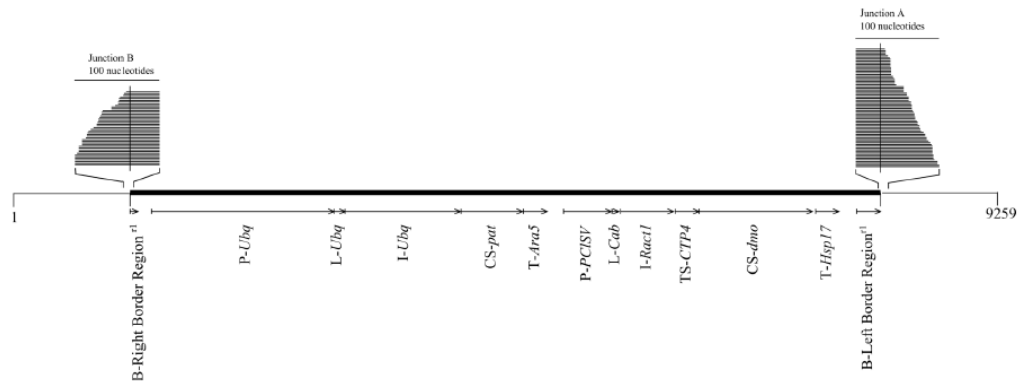


図 3.2.1-4 MON87419 において検出されたジャンクション配列

■ 遺伝的安定性の方法

- ・ 4 つの世代（R₃F₁、R₄、R₄F₁ 及び R₄）の安定性を完全にキャラクターゼーションされた R₃ と比較。
- ・ コントロールには、LH244 及び HCL645 x LH244 を使用。

使用した MON87419	コントロール
R ₃ (LH244) R ₃ F ₁ (HCL645 x LH244) R ₄ (LH244) R ₄ F ₁ (HCL645 x LH244) R ₅ (LH244)	LH244, HCL645 x LH244

■ 遺伝的安定性解析の結果

- ・ コントロール（LH244 及び HCL645 x LH244）ではジャンクション配列は検出されていない。
- ・ MON87419 の全ての世代からの DNA には同一のふたつのジャンクション配列があり、全ての世代にわたって JSA の結果が一致したことから、単一の T-DNA I インサートが MON87419 に安定的に維持されていることが示された。

USDA 及び FSANZ における MON87419 トウモロコシの NGS/JSA の解析結果に関する見解を表 3.2.1-5 に示す。

USDA 及び FSANZ のいずれも、挿入遺伝子の数、プラスミドバックボーンの有無及び挿入遺伝子の安定性について NGS/JSA の結果に基づく評価を行っている。なお、FSANZ (2016) では、挿入遺伝子の世代間安定性について、「サザンブロットあるいは NGS/JSA のような手法が最適である」と記載している。この表現は 2015 年に公表された前述の MON87411 の評価書にはなかったものである。

表 3.2.1-5 MON87419 の NGS/JSA の解析結果に関する 2 つの評価機関の見解

項目	USDA (2016、[文献 ID-077])	FSANZ (2016、[文献 ID-076])
挿入遺伝子の数	■ <u>NGS/JSA の結果から</u> ； 単一の T-DNA が存在している。	■ <u>NGS/JSA の結果から</u> ；MON87419 に 単一のゲノム遺伝子座にひとつの T-DNA I が挿入されていることが確 認された (※)。
プラスミド バックボ ーンの有無	■ <u>NGS/JSA の結果から</u> プラスミド PV-ZMHT507801 骨格または T-DNA II 由来の配列を含まない。	■ <u>NGS/JSA の結果から</u> ；MON87419 にプラスミドバックボーンの配列 はないことが示されている (※)。
挿入遺伝子の世代間安定性	■ <u>NGS/JSA の結果から</u> ；MON87419 に 挿入された T-DNA 配列は、5 つの繁殖 世代にわたって単一の遺伝子座で安 定して遺伝した。	■ <u>NGS/JSA の結果から</u> ；JSA の結果 が一致することから、単一の T-DNA I インサートは 5 世代にわたって安定 していた。
全体の結論	MON 87419 トウモロコシは植物病害 虫のリスクをもつ可能性は低い。	公衆衛生上及び安全上の懸念事項は 確認されていない。MON87419 由来 の食物は、人による消費に関して従 来のトウモロコシ品種由来の食物と 同様に安全であると考えられる。

(※) T-DNA I の単一コピーの挿入及び T-DNA II のベクターバックボーンがないとする NGS/JSA に基づく結論が、直接シーケンシングによる解析の結果によっても確認されていることが記載されている。

④ NF872（非褐変 Arctic®フジリンゴ）

NF872（非褐変 Arctic®フジリンゴ）は4種のポリフェノールオキシダーゼ（PPO）の抑制により褐変しないリンゴであり、現時点では USDA においてのみ認可されている（USDA, 2016, [文献 ID-087]）。NG872 は、2015 年にすでに認可されている2種類の遺伝子組換えリンゴ（GS784（栽培品種ゴールデン・デリシャスに由来する Arctic®ゴールデン）及び GD743（栽培品種グラニー・スミスに由来する Arctic®グラニー）と全く同じコンストラクト（pGEN-03）を用いて開発されており、これらの遺伝子組換えリンゴの認可の拡張申請に基づき、USDA が評価を実施した。

表 3.2.1-6 に NF872 の基礎情報を示す。

表 3.2.1-6 FN872 の基礎情報

宿主	<i>Malus domestica</i> , cv Fuji（フジリンゴ）	
導入遺伝子及びその由来	第1カセット	アグロバクテリウム由来のノパリン合成酵素プロモーター配列（PNOS）
		大腸菌Tn5由来のネオマイシンリン酸転移酵素II遺伝子（ <i>nptII</i> ）
		アグロバクテリウム由来のノパリン合成酵素ターミネーター配列（TNOS）
	第2カセット	アルファルファモザイクウイルスRNA4由来の非翻訳リーダー配列及びカリフラワーモザイクウイルス由来のCaMV 35Sプロモーターの重複エンハンサー（P70）
		リンゴ由来の4つのポリフェノールオキシダーゼ（PPO）遺伝子の4つの断片のキメラセンス抑制配列（PPO2、GPO3、APO5、 <i>pSR7</i> （PGAS）） PPO遺伝子ファミリーの発現を抑制することで非褐変となる
		アグロバクテリウム由来のノパリン合成酵素ターミネーター配列（TNOS）
遺伝子の導入方法	バイナリートランスフォーメーションベクターpGEN-03をアグロバクテリウム媒介形質転換法により導入	

申請者は、インサートの確認及びインサートマップの作成に全ゲノムシーケンシング（WGS）を使用しており、この解析により染色体3、13及び17にT-DNAの挿入が確認されている。なお、申請者はインサートマップを分子キャラクタリゼーションにおけるサザンブロットの代わりとも位置づけている（実際にサザンブロットの結果は提出されていない）。また、挿入遺伝子の世代間安定性については、GS784及びGD743の申請で確認されているとして、評価対象となっていない。

申請資料に基づく WGS の実験及びデータ解析ならびに WGS に基づく結果詳細を以下に示す。

■全ゲノムシーケンシング (WGS)

- ・ NF872 の葉試料及びその親 NF 対照から DNeasy Plant Mini Kit (Qiagen) を用いてゲノム DNA を抽出。
- ・ TruSeq Rapid SBS Kit-HS Kit (Illumina) を用いて、HiSeq2500 (Illumina) により WGS を実施 (2 x 150 サイクル、ペアエンド)。
- ・ 低品質の読み取りは、Q20 のしきい値を持つカスタム PERL スクリプトにより除去。

■T-DNA/リンゴジャンクションの特定

- ・ bbDuk (オープンソースソフト) を用いて、PPO 配列を除いた GEN-03 ベクターと完全に一致する 21 bp を含むリードを検索し、6,069 リード (458 kb) を特定。
- ・ Multiple BLAST 検索により、ベクター及びりんごの配列をもつリードを特定し、インハウスの R スクリプトソフトウェアを用いて T-DNA/リンゴの配列を評価。
- ・ T-DNA 及びリンゴのゲノム配列と完全に一致する 16 リード (2,416 bp) を同定。
- ・ さらに、6,069 リードのうち、20 bp T-DNA 及び 20 bp リンゴの計 40 bp のジャンクションをもつリードを特定。
- ・ 3 つの T-DNA の挿入を示す 6 つの T-DNA/リンゴジャンクションを特定。
- ・ ゴールデン・デリシャスの v1.0 コンティグの Phytozome v9.1 BLAST サーバーを用いてインサート箇所を評価した結果、4 つの異なるコンティグ上の 3 つのゲノム位置にマッピングされた。染色体 17 の挿入は、2 つのマッピングされたコンティグ間で起こり、挿入の各末端は別々のデータベースエントリーにマッピングされた。

T-DNA が挿入された染色体の位置 (USDA (2016、[文献 ID-087]) より抜粋)

Chromosome	Contig Size (Mb)	Contig Name
3	24.01	MDC009179.475
13	31.60	MDC006811.297
17	5.76	MDC019498.104 / MDC016794.218

■内部 T-DNA ジャンクションの特定

- ・ もとの 6,069 個の GEN-03 特異的なリードと GEN-03 バイナリーベクターのマトリックス分析を用いて、ジャンクションをまたぐ個別のリードを同定した結果、GEN-03 と部分的な同一性しかなく、T-DNA/リンゴジャンクションに含まれない 150 bp のリードを特定 (なお、第 1 の GEN-03 部位に対する同一性のないリードは、第 2 の GEN-03 位置に対する同一性を有していた)。

■染色体毎のインサートの概要

【染色体 3 への挿入】（図 3.2.1-5）

- 染色体 3 への挿入は、3 つの GEN-03 T-DNA 断片から構成されており、PPO 抑制遺伝子を有する逆方向反復構造を含んでいる（図 3.2.1-5）。2 つの逆方向挿入遺伝子の組換えサイトはシーケンシングリードの中には見出されず、組換えサイトが PPO 配列内にあることを示している（つまり、NOS ターミネーター及び RB 配列が欠落している）。
- 機能性配列を含まない、1409bp のベクター骨格（13872 - 67）が含まれている。

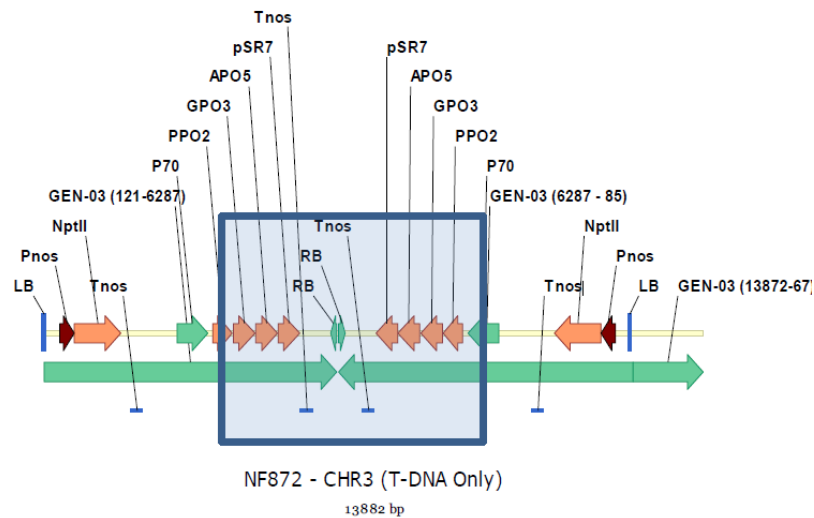


図 3.2.1-5 染色体 3 へのインサート

【染色体 13 への挿入】（図 3.2.1-6）

- 染色体 13 への挿入は、2 つの GEN-03 T-DNA フラグメントで構成されており、PPO 抑制遺伝子を有する逆方向反復構造を含んでいる（図 3.2.1-6）。また、2 つの逆向き導入遺伝子間の組換えサイトは RB 配列内にある。
- 染色体 13 の挿入には、バイナリーベクターのバックボーン配列は含まれていない。

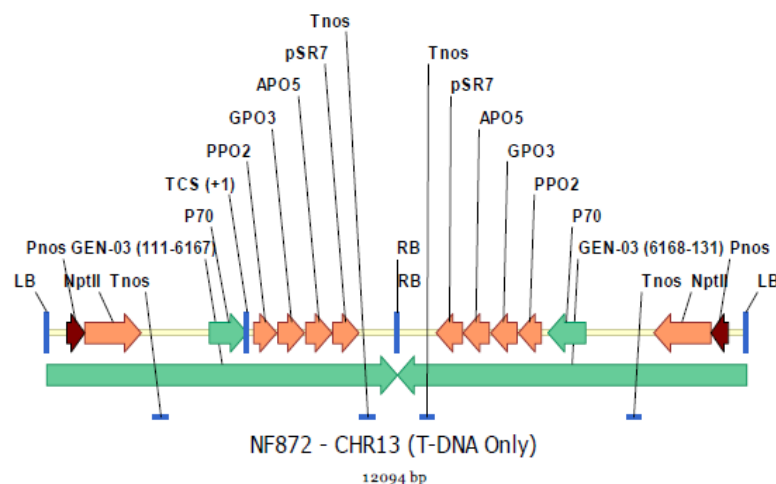


図 3.2.1-6 染色体 13 へのインサート

【染色体 17 へのインサート】（図 3.2.1-7）

- 染色体 13 への挿入は、2 つの GEN-03 T-DNA フラグメントで構成されており、ひとつの部分的な T-DNA 断片（完全な PPO 抑制遺伝子を含む）が含まれている（図 3.2.1-7）。また、部分的な CaMV 35S プロモーター（P70）が残っている。
- 染色体 17 の挿入には、バイナリーベクターのバックボーン配列は含まれていない。

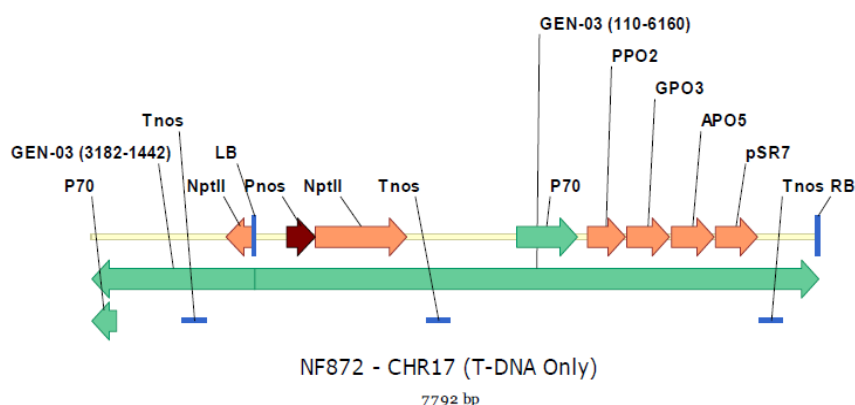


図 3.2.1-7 染色体 17 へのインサート

■意図しない結果の可能性

- マッピングアプローチにより、NF872 における 3 つの導入遺伝子挿入部位の正確なマッピングが可能になり、挿入部位の染色体位置及び各位置に挿入された T-DNA の正確なリミット（precise limit）が得られた。
- 染色体 3 を除き、導入遺伝子挿入の内部構造を明らかにした。
- 染色体 3 への挿入については、内部組換え部位が PPO 配列内で生じたため、今回使用したバイオインフォマティクスアプローチではクロスオーバーの正確な位置を明らかに出来なかった。
- 染色体 3 及び 13 の挿入については、オーファンプロモーター配列による未知の内在性配列の発現の可能性がないような導入遺伝子の構造となっている。
- 染色体 13 に挿入された GEN-03 バイナリーベクター断片は、機能性配列を含んでおらず、安全性の懸念はない。
- 染色体 17 への挿入には、部分的な CaMV 35S プロモーターが含まれているが、PPO 抑制挿入遺伝子には連結しておらず、リンゴゲノムに配向されている（The insertion in CHR17 includes a partial CaMV 35S promoter that is no longer attached to our PPO suppression transgene and which is oriented such that it is pointing into the apple genome.）。配列を解析した結果、プロモーター部分の機能エレメント（近位及び中領域全体、第 1 エンハンサーの遠位領域全体、第 2 エンハンサーの遠位領域の 3' 末端を含む）のほとんどは欠損していた。このことから、部分的な CaMV 35S プロモーターが未知の内因性配列の発現に関与する可能性は無く、安全性の懸念はない。

NF872 について、WGS のデータに基づく USDA の評価結果を表 3.2.1-7 に示す。

表 3.2.1-7 NF872 の WGS の解析結果に関する USDA の見解

項目	USDA (2016、[文献 ID-087])
挿入遺伝子の数	6 つの T-DNA/リンゴジャンクションが同定され、3 つの染色体染色体 3、13 及び 17 に複数の GEN-03 断片を有する 3 つの複合 T-DNA 挿入が存在することが示された。
プラスミドバックボーンの有無	- 染色体 3 における挿入には、機能性配列を持たない約 1400bp のベクターバックボーン配列が含まれ、他の 2 つの染色体における挿入では、ベクターのバックボーン配列は含まれていない。 - 染色体 17 中の挿入は部分的 CaMV 35S プロモーターを含むが、プロモーターの機能的要素の大部分は欠失している。この部分的 CaMV 35S 挿入が未知の内因性配列の発現を駆動する可能性はないため、安全性の問題ではない。 - 染色体 3 及び 13 における挿入遺伝子の構造は、孤立したプロモーター配列による未知の内因性配列の発現の可能性はほとんどない。
全体の結論	NF872 リンゴは、個別または集合的に、人間の環境の質に重大な影響を与えず、連邦政府がリストアップしている脅威または絶滅のおそれのある種、上場提案された種、またはその指定もしくは提案された重要な生息地に影響を与えない

(2) 植物以外の遺伝子組換え生物における NGS データに基づく評価機関の評価状況

米国食品医薬品庁 (FDA) では食品添加物について、一般的に安全と認められる (GRAS ; Generally Recognized As Safe) 物質について、企業からの届出 (Notice) が行われている。本調査において、FDA に提出された GRAS 届出において、NGS のデータが使用された事例が 2 件あった。乳酸菌 (*Bacillus coagulans*) 及びパン酵母に関するもので、いずれも遺伝子組換えは行われていない。以下に詳細を示す。

① 乳酸菌 (*Bacillus coagulans*) の事例 [文献 ID-111]

届出番号 GRN 601 の事例で、乳酸菌 *Bacillus coagulans* スポア製剤 (LactoSpore® : Microbial Type Culture Collection (MTCC) では MTCC5856) に関する Sabinsa Corporation による届出である (2015 年 9 月 15 日付け)。

GRAS 対象の MTCC5856 を特徴付ける試みとして、Genotypic Technology [P] Ltd. (インド) で実施された Illumina 社の Miseq を用いた全ゲノムシーケンシング (WGS) が行われ (最大 301 bp のペアエンド)、以下の解析をしている。

- ・ アセンブルした配列を RAST (Rapid Annotation using Subsystem Technology) ⁵ にアップロードし、*Bacillus coagulans* ATCC 7050 をリファレンスとしたアノテーション付けによりタンパク質機能を推定した。また、KAAS (KEGG Automatic Annotation Server) ⁶ にアセンブルした配列を入力し、パスウェイの特定を行った。
→ 98 コンティグから 3995 タンパク質機能及び 40 の機能パスウェイが推定された。
- ・ tRNAscan-SE ⁷ (version. 1.3.1) により 82 の tRNA 遺伝子を、RNAmmer 1.2 ⁸ により 10 の rRNA 遺伝子を同定した。
- ・ アセンブルした scaffold を antiSMASH 2.0 ⁹ に入力して、MTCC 5856 のゲノム中のバクテリオシン遺伝子クラスターの同定を行った。その結果、4 つのバクテリオシン遺伝子クラスターが特定された (バクテリオシンの存在/産生はプロバイオティクスとして望ましい性質のひとつ)。
- ・ NCBI データベースから *Bacillus spp.* の嘔吐毒素 (emetic toxin) 遺伝子の配列をダウンロードし、MTCC 5856 のアセンブルしたコンティグをクエリーとしたブラスト解析を実施した。その結果、MTCC 5856 に嘔吐毒素遺伝子がないことが示唆された。
- ・ ゲノム中にプラスミドがないことを確認するため、アセンブルした配列を PlasmidFinder ¹⁰ (version 1.2) で解析した結果 (腸内細菌及びグラム陽性プラスミドのデータベースを選択)、rep21 プラスミドのレファレンス配列に 25% の寄与が

⁵ <http://www.nmpdr.org/FIG/wiki/view.cgi/FIG/RapidAnnotationServer>

⁶ <http://www.genome.jp/tools/kaas/>

⁷ <http://lowelab.ucsc.edu/tRNAscan-SE/>

⁸ <http://www.cbs.dtu.dk/services/RNAmmer/>

⁹ <http://antismash.secondarymetabolites.org/> (Nucl Acids Res (2013) 41 (W1): W204-W212.)

¹⁰ <https://cge.cbs.dtu.dk/services/PlasmidFinder/>

あるとする、1件のヒットのみであったことから、MTCC 5856のゲノムに如何なるプラスミドも存在しないと結論した。

- ・ アミノグリコシド、 β -ラクタム、フルオロキノロン、ホスホマイシン、フシジン酸、MLS-マクロライド-リコサミド-ストレプトグラミB、ニトロイミダゾール、フェニコール、リファンピシン、スルホンアミド、テトラサイクリン、トリメトプリム及び糖ペプチドへの薬物耐性遺伝子を特定するために、アセンブルした配列を ResFinder¹¹ (version 2.1) によりシーケンスアイデンティティ及びカバレッジの閾値をそれぞれ98%及び60%として解析した。その結果、いずれの薬物耐性遺伝子も特定されなかった。

なお、2017年1月時点では、本届出に対するFDAからの意見は「Pending」となっている。

② パン酵母 (*Saccharomyces cerevisiae*) の事例 [文献 ID-112]

届出番号 GRN 604 の事例で、適応性進化により食品及び飲料中におけるアクリルアミド生成を防ぐ酵母 (RBSC AR 酵母) に関する Renaissance BioScience Corp (RBSC) による届出である (2015 年 10 月 14 日付け)。表 3.2.1-8 に対象酵母に関する基礎情報を示す。

表 3.2.1-8 GRN 604 における酵母に関する基礎情報

親酵母株	パン産業において広く使用されている工業用 <i>S.cerevisiae</i> のパン酵母株
目的機能獲得のために必要な遺伝子	細胞外の細胞壁関連アスパラギン分解酵素アスパラギナーゼ II をコードする天然 <i>ASP3</i> 遺伝子 (4 連タンデムの繰り返し配列) <i>ASP3</i> は、酵母での窒素利用を選択的に制御する機構 (窒素代謝抑制 (NCR) として知られている。
適応性進化の方法	突然変異誘発による適応的進化法 (DNA 組換え技術は使用していない)

親酵母株及び RBSC AR 酵母のリボソーム DNA の内部転写スペーサー (ITS) 領域、25S 領域及び 18S 領域の配列について、S288C *S.cerevisiae* のリファレンスゲノムとの比較により *S.cerevisiae* との同一性を示すために、Miseq (Illumina 社) による解析が行われた。親酵母株及び RBSC AR 酵母のゲノム DNA から、Nextera XT を用いてシーケンスライブラリを調製し、Miseq により 2x300 bp のペアエンドでのシーケンスを行い、S288C *S.cerevisiae* の

¹¹ <https://cge.cbs.dtu.dk/services/ResFinder/>

リファレンスゲノムとのマッチングが行われている。この結果、3 者の ITS 領域、25S 領域、18S 領域が完全に一致したことから、親酵母株と RBSC AR 酵母の同一性を示している。

本届出において、申請者は RBSC AR 酵母が GRAS に該当すると結論している。

なお、FDA は 2016 年 4 月 27 日に「意図された使用条件下においてはアスパラギン分解酵母が GRAS であるという RBSC の結論に関して、現時点では疑問がない」との意見を公表している¹²。

3.2.2. 文献ベースでのNGSの活用状況

遺伝子組換え植物の分野における文献ベースでの研究における NGS の活用状況について整理した結果についてまとめた。関連する個別論文の概要を表 3.2.2-1 に示す。また、植物のゲノムサイズ及び染色体構造を表 3.2.2-2 に示す。

以下に示すように、表 3.2.2-1 の結果をまとめると、遺伝子組換え植物の分野において NGS は主に分子キャラクタリゼーションの実施あるいは分子キャラクタリゼーションのための手法開発を目的とした研究に活用されており、次いで RNA-seq による導入遺伝子の機能/効果解析、NGS と他法を組み合わせた解析手法の提案等であった。主な解析対象となっていた植物はダイズ、イネ、シロイヌナズナ、トウモロコシであった。

① NGS を用いた分子キャラクタリゼーション

- ・ SNPs、欠損（イネ）
- ・ T-DNA の挿入（ダイズ）
- ・ (T-DNA を含む) DNA 挿入部位／数の特定（ダイズ、亜麻、イネ、シロイヌナズナ、トウモロコシ）
- ・ 配列確認（イネ）
- ・ 挿入遺伝子の再編成（ダイズ、シロイヌナズナ）
- ・ CRISP/Cas によるトランスジェニック植物（シロイヌナズナ）

② NGS データの新規取得法

- ・ マルチコピー配列の解析（イネ）
- ・ ライブラリ調製法（シロイヌナズナ、トウモロコシ）

③ データ解析方法の開発（条件検討など）

- ・ 挿入部位の特定方法（ダイズ）
- ・ 解析条件検討（イネ、テンサイ、ダイズ、アブラナ、ワタ、トウモロコシ、コムギ）

④ 未確認遺伝子組換え植物の検出に関する手法開発（トウモロコシ、ワタ、ダイズ、イネ）

⑤ mRNA シークエンシング（RNA-seq）を用いた挿入遺伝子の影響解析

¹² <http://www.fda.gov/Food/IngredientsPackagingLabeling/GRAS/NoticeInventory/ucm511407.htm>

- ・挿入 siRNA の確認（バナナ）
 - ・RNAi の効果の確認（イネ）
 - ・導入遺伝子の機能/効果解析（ワタ、ダイズ、イチゴ、トウモロコシ、ナタネ、タバコ、シロイヌナズナ、セイヨウアブラナ、トマト）
- ⑥ 導入遺伝子に由来する siRNA の検出
- ⑦ 新規機能遺伝子の同定（コムギ）

表 3.2.2-1 遺伝子組換え植物に NGS を適用した事例 (概要)

*1 データクリーニング: NGS の生データ (測定直後) からのノイズデータの除去、マッピング/アライメント: 既知の参照ゲノムに対してシーケンスリードをマップすること。
*2 Raw reads: 生データ (測定直後) のリード数、clean reads: データクリーニング後のリード数、Reads mapped to genome: マッピング/アライメントされたリード数、Average depth/ Coverage: 平均のシーケンス深度 (冗長度) もしくはカバレッジ

文献 ID	タイトル	著者	掲載誌	測定機器	アルゴリズム*1	解析対象	測定したリード数*2	実施内容
ID-001	Global Transcriptomic Analysis of Targeted Silencing of Two Paralogous ACC Oxidase Genes in Banana	Xia Y et al.	Int J Mol Sci. 2016; 17(10).	Illumina Hiseq2500	1) データクリーニング ・信頼性の低いリード、アダプター配列の排除 2) マッピング/アライメント ・配列データを Blastx にかけて遺伝子を同定 3) 定量解析 ・ Transcriptome データを Trinity で再構成 ・ RSEM で定量化	GM バナナ バナナにアグロバクテリウムによる形質転換法により Mh-ACO1 及び Mh-ACO2 の siRNA ベクターを導入	〔clean reads〕 ・ WT: 2,068,587 (96.42%) ・ As1: 16,046,988 (97.51%) ・ As2: 16,279,746 (97%)	<u>導入遺伝子による影響/機能解析</u> Mh-ACO1 及び Mh-ACO2 の機能解析の為に RNA-seq データからユニジーンを同定し、GO 解析を実施。
ID-002	Identification of Genomic Insertion and Flanking Sequence of G2-EPSPS and GAT Transgenes in Soybean Using Whole Genome Sequencing Method	Guo B et al.	Front Plant Sci. 2016; 7:1009.	Illumina Hiseq2500 ・ 125bp ペアエンド	1) データクリーニング ・信頼性の低いリード (Q<20)、アダプター配列の排除 2) マッピング/アライメント ・ BWA (デフォルトパラメータ) 〔参照ゲノム〕 Phytozome 由来の Glycine max Wm82.a2.v1 〔外来遺伝子〕 pKT-rGE ベクター配列	GM ダイズ ダイズにアグロバクテリウムによる形質転換によりグリホサート耐性遺伝子を導入	〔clean reads〕 ・ GE-J16 : 179,326,462 (22.41Gb) ・ ZH10-6 : 210,087,270 (26.26Gb) 〔Average depth/ Coverage〕 ・ GE-J16 : 20× ・ ZH10-6 : 22×	<u>分子キャラクタリゼーション/手法開発</u> 挿入位置の同定、ホストゲノムの置換 (13~24bp) を NGS による WGS によりを実施。また、Co-segragation 解析により及び経世代挿入遺伝子の安定性を確認。なお、挿入位置及び隣接配列については PCR 増幅後にサンガー法により確認した。
ID-003	Statistical framework for detection of genetically modified organisms based on Next Generation Sequencing	Willems S et al.	Food Chem. 2016; 192:788-798.	Illumina Hiseq 2000 Illumina Hiseq 1500 ・ 100bp ペアエンド	1) データクリーニング ・ カスタム Perl スクリプトで、Q20 を下回ったときにシーケンスされたペアエンドリードをトリミング(30bp 以上のリードを使用) 2) マッピング/アライメント ・ BWA and Bowtie2 combined with Python and Perl scripts 3) NGS データの比較解析 ・ CLC Genomics Workbench 7	・ イネ ・ ヌートル (コムギ) ・ テンサイ ・ ダイズ ・ ナタネ ・ ワタ ・ トウモロコシ ・ コムギ アグロバクテリウムによる形質転換により Cry1B 遺伝子を導入	〔clean reads〕 ・ 100% Bt rice : 107,455,990 ・ 10% Bt rice : 81,491,366	<u>分子キャラクタリゼーション/解析条件の検討</u> 遺伝子組換え植物を同定するために、Detection、Proof 及び Identification の 3 段階の検出レベルにおけるリード数を検討。また検出感度の為の検討として、GM 作物と野生型作物の混合サンプルにおけるリード数の検討も実施。。
ID-004	Molecular Characterization of Transgenic Events Using Next Generation Sequencing Approach	Guttikonda, SK et al.	PLoS One. 2016; 11(2):e0149515.	■WGS Illumina Hiseq 2000 ・ 100bp ペアエンド ■targeted capture sequencing (TCS) Illumina Miseq ・ 250bp ペアエンド	1) データクリーニング ・ CASAVA software (Illumina) -アダプター配列のトリミング -Q30 を採用リード 2) マッピング/アライメント ・ Burrows-Wheeler Aligner と Samtools ソフト 〔参照ゲノム〕 Soybean Williams82 reference genome 3) 高次解析 ・ BEDTools ソフトでゲノム及びトランスジーンの配列カバレッジを算出	GM ダイズ ダイズにアグロバクテリウムによる形質転換により外来遺伝子を導入	〔Reads mapped to genome〕 ・ TE 1 (F2-1) : 93,762,230 ・ TE 1 (T3-3) : 110,949,911 ・ TE 2 (F2-10) : 102,156,087 ・ TE 2 (T3-1) : 91,367,370 ・ TE 1×TE 2 : 140,544,330 〔Average depth/ Coverage〕 ・ TE 1 (F2-1) : 10× ・ TE 1 (T3-3) : 11× ・ TE 2 (F2-10) : 10× ・ TE 2 (T3-1) : 9× ・ TE 1×TE 2 : 14×	<u>分子キャラクタリゼーション/手法比較</u> 推定挿入隣接配列(3'と 5'の各 1kB)を PCR 増幅してサンガー配列決定した。 サザンブロットで挿入ベクターを検出した。 短いリードは、ゲノムの複雑な反復配列または T-DNA 内の反復領域内の接合領域を解明するには課題がある。

*1 データクリーニング: NGS の生データ (測定直後) からのノイズデータの除去、マッピング/アライメント: 既知の参照ゲノムに対してシーケンスリードをマップすること。

*2 Raw reads: 生データ (測定直後) のリード数、clean reads: データクリーニング後のリード数、Reads mapped to genome: マッピング/アライメントされたリード数、Average depth/ Coverage: 平均のシーケンス深度 (冗長度) もしくはカバレッジ

文献 ID	タイトル	著者	掲載誌	測定機器	アルゴリズム*1	解析対象	測定したリード数*2	実施内容
ID-006	CONTRAILS: A tool for rapid identification of transgene integration sites in complex, repetitive genomes using low-coverage paired-end sequencing	Lambirth KC et al.	Genom Data. 2015; 6:175-81.	Illumina Hiseq2000 ・ 1-5ug gDNA ・ コバリスで 350bp(平均)に断片化 ・ 100bp ペアエンド	1) データクリーニング ・ Q30 を採用リード 2) マッピング/アライメント ・ Bowtie 〔参照ゲノム〕 soybean reference genome sequence version 2.75 (Phytozome)	GM ダイズ ダイズにアグロバクテリウムによる形質転換により human thyroglobulin gene を導入	〔clean reads〕 ・ ST77-KP2 hTG : 27,983,663 〔Average depth/ Coverage〕 ・ ST77-KP2 hTG : ~5×	<u>分子キャラクタリゼーション/手法開発</u> 非常に複雑で反復性のある c 植物ゲノムにおけるトランスジェン挿入部位の同定するために NGS 解析(低カバレッジ、ペアエンド)と PCR を使用した CONTRAILS (Characterization of Transgene Insertion Locations with Sequencing) 法を開発し、挿入位置、配列を確認。
ID-007	Next-generation sequencing is a robust strategy for the high-throughput detection of zygosity in transgenic maize	Fritsch L et al.	Transgenic Res. 2015; 24(4):615-23.	Ion Torrent Sequencer (Life Technology) ・ トランスジェン 特異的 PCR 増幅産物の NGS 解析	1) データクリーニング ・ 記載なし 2) マッピング/アライメント ・ Lasergene Genomic Suite software ・ SeqMan Ngen program	GM トウモロコシ アグロバクテリウム由来 CP4 EPSPS 及び Bt 由来 Cry34/35Ab を導入した	〔Raw reads〕 ・ 約 8,000,000 〔Average depth/ Coverage〕 ・ 約 100×	<u>分子キャラクタリゼーション/手法開発</u> トランスジェンと内在ゲノム接合体データを取得するために、qPCR と NGS の 2 手法による検討を実施 (効率的な GM 植物検出系の開発)。
ID-008	Genetics, structure, and prevalence of FP967 (CDC Triffid) T-DNA in flax	Young L et al.	Springerplus. 2015; 4:146.	Illumina HiSeq ・ mate-paired/ paired-end ・ トランスジェン 特異的 PCR 増幅産物の NGS 解析	1) データクリーニング ・ 記載なし 2) マッピング/アライメント ・ Bowtie2、Ray、SOAPdenovo 〔参照ゲノム〕 flax reference genome (ver 1.0 at Phytozome.net, accessed 8 Feb 2015)	GM 亜麻 (FP967) (種子を NGS 解析した)	〔alignment reads〕 ・ 約 32,000 reads	<u>分子キャラクタリゼーション/</u> GM 亜麻における NGS を用いた分子キャラクタリゼーション (遺伝的特徴及び構造、配列決定)。
ID-009	HtStuf: High-Throughput Sequencing to Locate Unknown DNA Junction Fragments	Kanizay LB et al.	Plant Genome. 2015; 8(1).	Illumina Miseq	1) データクリーニング ・ Geneious 7 (市販ソフトウェア) 2) マッピング/アライメント ・ Geneious 7 (市販ソフトウェア) 〔参照ゲノム〕 soybean reference genome	GM ダイズ (Williams82)	〔Raw reads〕 ・ 2,364,016 リード	<u>分子キャラクタリゼーション/手法開発</u> ダイズにおける長鎖の導入遺伝子や短鎖のマルチコピー配列の解析に有用な、改変 TA クローニングによるライブラリ調製法とイルミナシーケンスを組み合わせた手法 (high-throughput sequencing to locate unmapped DNA fragments (HtStuf)) を開発。

*1 データクリーニング: NGS の生データ (測定直後) からのノイズデータの除去、マッピング/アライメント: 既知の参照ゲノムに対してシーケンスリードをマップすること。
*2 Raw reads: 生データ (測定直後) のリード数、clean reads: データクリーニング後のリード数、Reads mapped to genome: マッピング/アライメントされたリード数、Average depth/ Coverage: 平均のシーケンス深度 (冗長度) もしくはカバレッジ

文献 ID	タイトル	著者	掲載誌	測定機器	アルゴリズム*1	解析対象	測定したリード数*2	実施内容
ID-010	Southern-by-Sequencing: A Robust Screening Approach for Molecular Characterization of Genetically Modified Crops	Zastrow-Hayes GM et al.	Plant Genome. 2015; 8(1).	Illumina Miseq Illumina Hiseq 2500 ・ 100bp ペアエンド	1) データクリーニング ・信頼性の低いリード (Q<20)、アダプター配列の排除 ・ K-mer 解析 : JELLYFISH, version 1.1.4 2) マッピング/アライメント ・ Bowtie ver. 1.0.0 〔参照ゲノム〕 maize reference genome (2009): 2 ミスマッチを許容 〔外来遺伝子〕 T-DNA: ミスマッチ許容なし →アライメントされなかった配列 : Bowtie 2 ver 2.1.0 で再解析 ・ BWA ver. 0.5.9-r16: ジャンクション解析でのアライメント 3) 高次解析 ・ SSAKE (Short Sequence Assembly by progressive K-mer search and 3' read Extension) version 3.8.2: 長鎖コンテイクの作成 ・ BLAT version 35×1: 長鎖コンテイクの相同性検索	GM トウモロコシ	〔reads〕 ・ 記載なし	<u>分子キャラクタリゼーション/手法開発</u> サザンプロット解析の問題を解決するために、シーケンスキャプチャーと NGS を組み合わせた Southern-by-Sequencing (Sbs)を開発。プラスミド骨格の検出及びコピー数の確認、ジャンクション配列の検出が可能で、サザンプロットから置き換えもできるとの記載あり。
ID-011	Efficiency to Discovery Transgenic Loci in GM Rice Using Next Generation Sequencing Whole Genome Re-sequencing	Park D et al.	Genomics Inform. 2015; 13(3):81-85.	Illumina Hiseq 2500 ・ 150bp ペアエンド	1) データクリーニング ・ BGI-Shenzhen (Shenzhen, China) -短いリード及び低品質のリード、アダプター配列のトリミング 2) マッピング/アライメント ・ Burrows-Wheeler Aligner (BWA) ver.0.6.2 ・ Assembly by Short Sequences (ABYSS) アルゴリズム ver. 1.3.4.2 〔参照ゲノム〕 TIGR7.0 〔外来遺伝子〕 pBar-1c、pFHBT1 及び pGL2RC7 3) 高次解析 ・ De novo 解析/BLAST 解析	GM イネ EGF を組み込んだイネ (event PJKS131-2)	〔Raw reads〕 ・ 194,965,440 〔Average depth/ Coverage〕 ・ 72×	<u>分子キャラクタリゼーション/</u> ゲノムとベクター配列に類似性がある場合に解析エラーが生じたため、形質転換ベクターに基づいたより正確なアルゴリズムの開発が必要である。 NGS を用いる場合、サンプルの品質、データ量、パイオインフォマティクスに大きく依存するため、NGS データと PCR による遺伝子導入部位検出のための適切なガイドラインが求められる。
ID-014	Host-induced post-transcriptional hairpin RNA-mediated gene silencing of vital fungal genes confers efficient resistance against Fusarium wilt in banana	Ghag SB et al.	Plant Biotechnol J. 2014; 12(5):541-553.	Illumina GAIIX	1)データクリーニング ・ 記載なし 2)マッピング/アライメント ・ Bowtie	GM バナナ バナナにアグロバクテリウムによる形質転換により ihpRNA 導入遺伝子を導入	〔Raw reads〕 ・ Line V2 : 5,125,712 reads ・ Line F2 : 13,556,873 reads 〔Aligned reads〕 ・ Line V2 : 39,7594 reads ・ Line F2 : 11,5252 read	<u>siRNA 解析</u> トランスジェニックバナナに導入した挿入遺伝子 (ihpRNA) による siRNA の検出のため、NGS による siRNA 解析を実施。

*1 データクリーニング: NGS の生データ (測定直後) からのノイズデータの除去、マッピング/アライメント: 既知の参照ゲノムに対してシーケンスリードをマップすること。
*2 Raw reads: 生データ (測定直後) のリード数、clean reads: データクリーニング後のリード数、Reads mapped to genome: マッピング/アライメントされたリード数、Average depth/ Coverage: 平均のシーケンス深度 (冗長度) もしくはカバレッジ

文献 ID	タイトル	著者	掲載誌	測定機器	アルゴリズム*1	解析対象	測定したリード数*2	実施内容
ID-015	Detecting authorized and unauthorized genetically modified organisms containing vip3A by real-time PCR and next-generation sequencing	Liang C et al.	Anal Bioanal Chem. 2014; 406(11):2603-2611.	・ Illumina HiSeq2000 -100 bp ペアエンド ・ PacBio RS * Targeted Sequencing	■HiSeq2000 1) データクリーニング ・ PRINSEQ ver.0.5.1 2) De novo 解析 ・ Velvet ver.1.2.06 3) マッピング/アライメント ・ Bowtie2 ver. 2.02 〔参照配列〕 MIR162 (ID: HI203339) 4) 平均 depth の算出 ・ BEDtools 2.16.2 5) コンセンサス配列の作成 ・ Geneious (Biomatters Ltd) ■PacBio RS 1) データクリーニング ・ 800 塩基長以上のリードを採用 2) De novo 解析 ・ CLC Genomics Workbench (EMEA) ・ BLAST 3) マッピング/アライメント ・ Bowtie2 ver.2.02 〔参照配列〕 MIR162 (ID: HI203339)	〔Illumina HiSeq2000〕 vip3A を導入した ・ ト ウ モ ロ コ シ (MIR162) ・ 綿 (COT102) 〔PacBio RS〕 vip3A を導入した ・ ト ウ モ ロ コ シ (MIR162)	〔Mapped reads〕 ■HiSeq2000 ・ VIPup: 1,417,334 ・ VIPdown: 1,642,933 ■PacBio RS ・ VIPup: 8,525 ・ VIPdown: 5,423 〔Read depth〕 ■HiSeq2000 ・ VIPup: 56,857 ・ VIPdown: 59,361 ■PacBio RS ・ VIPup: 4,660 ・ VIPdown: 2,449	<u>未知 GMO 検出法の開発</u> 未認可 GMO を検出するために定量 PCR 法及び NGS を用いた接合領域の検出法を開発した。 より長く正確なコンティグを生成できるよう de novo 解析の更なる改善が必要で、PacBio を用いてデータ補完することでこの点が改善できた。
ID-016	Next-Generation Sequencing as a Tool for Detailed Molecular Characterisation of Genomic Insertions and Flanking Regions in Genetically Modified Plants: a Pilot Study Using a Rice Event Unauthorised in the EU	Wahler D et al.	Food Anal. Methods. 2013; 6: 1718.	Illumina GenomeAnalyzer2 ・ 100bp ペアエンド	1) データクリーニング、2) マッピング ・ Illumina Genome Analysis System (Genome analyzer and CASAVA 1.8.2) 〔参照ゲノム〕 <i>O. sativa ssp. japonica</i> (cultivar Nipponbare MSU6) 3) 変異解析 ・ MEGABLAST 〔外来遺伝子〕 pCAMBIA-1300	GM イネ (LLRice62)	〔Raw reads〕 ・ 343,314,310 〔Read depth〕 ・ 平均 65×	<u>分子キャラクタリゼーション</u> 挿入位置及び配列特定に NGS を使用。
ID-017	Analysis of differentially expressed genes in response to endogenous cytokinins during cotton leaf senescence	Zhao P et al.	Biol Plant. 2013; 57: 425.	Illumina Solexa	1) データクリーニング ・ Illumina pipeline により、低品質の排除 2) マッピング/アライメント 〔参照ゲノム〕 ワタゲノム (TIGR データベース) 3) 高次構造 ・ KEGG pathway 解析	GM ワタ アグロバクテリウムによる形質転換によりイソペンテニルトランスフェラーゼ遺伝子 (ipt) を導入	〔clean reads〕 ・ Transgenic cotton lines : 224617 reads ・ azygous cotton lines : 200071 reads	<u>導入遺伝子による影響/機能解析</u> GM ワタの mRNA 解析を Illumina 社の Solexa 技術を用いて実施した事例。

*1 データクリーニング: NGS の生データ (測定直後) からのノイズデータの除去、マッピング/アライメント: 既知の参照ゲノムに対してシーケンスリードをマップすること。
*2 Raw reads: 生データ (測定直後) のリード数、clean reads: データクリーニング後のリード数、Reads mapped to genome: マッピング/アライメントされたリード数、Average depth/ Coverage: 平均のシーケンス深度 (冗長度) もしくはカバレッジ

文献 ID	タイトル	著者	掲載誌	測定機器	アルゴリズム*1	解析対象	測定したリード数*2	実施内容
ID-019	The Use of Next Generation Sequencing and Junction Sequence Analysis Bioinformatics to Achieve Molecular Characterization of Crops Improved Through Modern Biotechnology	Kovalic et al.	Plant Genome. 2012; 5(3):149-163	Illumina HiSeq ・ 100bp ペアエンド	1) データクリーニング ・ 天然大豆由来リードは BlastAll 2.2.21 ソフトウェアで除外 ・ Bowtie 0.12.3 : 重複したリードの検索 ・ Novoalign 2.06.09 : アダプター配列の除去 2) マッピング/アライメント ・ Bowtie 0.12.3 3) ジャンクション検出 ・ Perl script でカスタムプログラム。 〔外来遺伝子〕 PVGMGOX20 もしくは PV-GMPQ/HT4404	GM ダイズ (MON17903、MON87704) *従来種: A3244、A3525	〔Raw reads〕 ・ MON17903: 1,139,538,382 ・ A3244: 1,322,787,632 〔Effective Coverage Depth〕 ・ MON17903: 97× ・ A3244 : 102×	<u>分子キャラクタリゼーション/手法開発</u> 評価書に引用が多い Junction Sequence Analysis の元論文。遺伝子組換え大豆 MON17903 や MON87704 (複数の T-DNA を有する) の分子キャラクタリゼーションを NGS/JSA 及びサングロッド法と比較 (結果は同等)。
ID-027	RNAi-mediated transgenic rice resistance to Rice stripe virus	Li L et al.	J Integrative Agriculture. 2016;15(11); 2539–2549	Illumina Hiseq: Chaina BGI	1) データクリーニング ・ 18–30 塩基をデータ解析に用いた 2) マッピング/アライメント ・ GSNAP (Wu and Nacu 2010) 〔参照ゲノム〕 イネゲノム (Osativa-193) 〔外来遺伝子〕 RSV reference genome (ID: NC-003755.1, NC-003754.1, NC-003776.1, NC-003753.1)	GM イネ イネにウィルスに RSV スクレオカプシドタンパク質 (NCP) 遺伝子に対する siRNA 配列を導入	〔Mapped reads〕 ・ 131 852	<u>small RNA-Seq 解析</u> siRNA 配列を導入したことによるトランスジェニックイネ (T4-B1) における RNAi の影響評価のための small RNA-Seq 解析。
ID-032	High-Throughput Analysis of T-DNA Location and Structure Using Sequence Capture	Inagaki S et al.	PLoS One. 2015; 10(10):e0139672.	Illumina Hiseq2500	1) データクリーニング ・ 信頼性の低いリード、アダプター配列の排除 2) マッピング/アライメント ・ Bowtie2 ・ SAMtools、 〔参照ゲノム〕 <i>Arabidopsis thaliana</i> (TAIR10 reference sequence) 3) ピークコール ・ MACS2 program	GM シロイヌナズナ シロイヌナズナにアグロバクテリウムによる形質転換により Columbia-0 (Col-0) もしくは Landsberg erecta (Ler) を導入	〔Clean reads〕 ・ 116,221～358,798	<u>分子キャラクタリゼーション</u> T-DNA 挿入の位置及び接合構造を決定するための費用効果の高い方法の開発として解析のための独自のパイプラインを作成。
ID-033	A Comparison of transgenic and wild type soybean seeds: analysis of transcriptome profiles using RNA-Seq	Lambirth KC et al.	BMC Biotechnol. 2015; 15:89.	Illumina Hiseq2000 ・ 100bp ペアエンド	1) データクリーニング ・ FastQC で確認 2) マッピング/アライメント ・ Tophat ver. 2.0.13 〔参照ゲノム〕 大豆ゲノム (version 2.75) 3) 発現量解析 ・ featureCounts program : リードカウント ・ edgeR ver 3.8.5 と cufflinks ver. 2.2.1 : 差分解析 ・ AgriGO : GO 解析	GM ダイズ(3 系統) バイナリーベクター法による導入 ・ 764 : mSEB タンパクを発現 ・ ST77 : hTG タンパクを発現 ・ ST111 : hMB タンパクを発現	〔Raw Reads〕 7,000,000-12,000,000	<u>導入遺伝子による影響/機能解析</u> 遺伝子組換え大豆タンパク質の生理機能に高レベルで発現させることの効果に関する研究として、HiSeq による遺伝子発現量データを取得 (RNA-Seq) 。
ID-035	Exploiting genome variation to improve next-generation sequencing data analysis and genome editing efficiency in Populus tremulaxalba 717-1B4	Xue LJ et al.	Tree Genetics & Genomes. 2015; 11: 82.	Illumina HiSeq2500 ・ 100bp ペアエンド ・ シングルリード 50bp	1) de novo 解析 ・ Trinity r20140717 2) マッピング ・ TopHat2 v2.0.12 3) 変異解析 ・ SHORE v0.9.3 ・ GATK v3.2.2 ・ SAMtools v0.1.19.0 4) 変異置換ゲノム ・ VCFtools v0.1.12b 5) アノテーション変換 ・ liftOver 6) 発現解析	カロリナポプラ (<i>Populus tremuloides</i>) とウラジロハコヤナギ (<i>Populus alba</i>) の交雑種 (717-1B4)	〔Raw reads〕 ・ 10～15 milion	<u>バリエント置換カスタムゲノムの開発</u> カロリナポプラとウラジロハコヤナギの交雑種 (717-1B4) をテストケースとして、WGS と RNA-seq 解析を実施し、両データからバリエント検出のパイプラインを構築した。さらに得られたデータから、バリエント置換カスタムゲノム: sPta717 を作成した。この sPta717 は CRISPR によるゲノム編集のためのガイ

*1 データクリーニング: NGS の生データ (測定直後) からのノイズデータの除去、マッピング/アライメント: 既知の参照ゲノムに対してシーケンスリードをマップすること。
*2 Raw reads: 生データ (測定直後) のリード数、clean reads: データクリーニング後のリード数、Reads mapped to genome: マッピング/アライメントされたリード数、Average depth/ Coverage: 平均のシーケンス深度 (冗長度) もしくはカバレッジ

文献 ID	タイトル	著者	掲載誌	測定機器	アルゴリズム*1	解析対象	測定したリード数*2	実施内容
					• DESeq2 v1.6.2 • SLIM ソフトウェア			ド RNA (gRNA) デザインにも役立つことが期待される。
ID-038	Transcription factors and anthocyanin genes related to low-temperature tolerance in rd29A:RdreB1BI transgenic strawberry	Gu X et al.	Plant Physiol Biochem. 2015; 89:31-43.	Illumina HiSeq2000	1) データクリーニング • BGI-Illumina pipeline -低クオリティの配列を除去 -3'側アダプター配列のみを除去 -Unknown'N'を含むリードを除去 -リード (21 bp) が 1 コピーのみを除去 2) 発現量解析 • TPM (number of transcripts per million clean tags) • False Discovery Rate (FDR) で。 3) パスウェイ解析 • KEGG 4) モチーフ解析 • Phytozome website	GM イチゴ イチゴに rd29A promoter を導入	[Clean reads] • NT-C: 3,433,754 • NT-LT: 3,443,081 • T-C: 3,662,143 • T-LT: 3,502,436 [Mapped tags] • NT-C: 1,921,591 • NT-LT: 1,839,631 • T-C: 2,260,024 • T-LT: 1,962,720	<u>導入遺伝子による影響/機能解析</u> 非 TG 及び RdreB1BI イチゴの RNA-seq 解析から低温耐性に関連して発現変動する多数の遺伝子を同定。
ID-040	Transcriptome differences between Cry1Ab resistant and susceptible strains of Asian corn borer	Xu LN et al.	BMC Genomics. 2015; 16:173.	Illumina Hiseq2000	1) データクリーニング •信頼性の低いリード (Q<20)、アダプター配列の排除 2) マッピング/アライメント • Trinity 3) 遺伝子の同定 • uBLAST、NR, Swiss-prot protein database,KEGG	GM トウモロコシ(4系統) 耐虫植物 Bt トウモロコシ(Cry1Ab)を強制発現 • ACB-BtS-1 • ACB-BtS-2 • ACB-AbR-1 • ACB-AbR-	[Clean reads] • 51,462,903～54,667,526 [Total number of unigene] • 53,710～63,032 種類	<u>導入遺伝子による影響/機能解析</u> 耐虫植物 Bt トウモロコシ (Cry1Ab を強制発現)に対して抵抗性を示すトウモロコシ害虫 (Asian corn borer (ACB)) のメカニズムを調べるために RNA-seq 解析を実施。Bt 感受性 (ACB-BtS) 及び Cry1Ab 耐性 (ACB-AbR) の個体で発現量に差のあった 3,793 遺伝子の遺伝子群の機能解析により、成長調節、解毒及び代謝/異化プロセス関連酵素に Cry1Ab 耐性に関連している可能性が示唆された。
ID-041	Transcriptome Profiling of hrf2 Transgenic Rapeseed Revealed Genes Related to the Defense Responses and Signaling Pathways	Wang Y et al.	Crop Sci. 2015; 55(2):800-810.	Illumina Solexa • 100bp ペアエンド	1) データクリーニング • アダプター配列、Quality が低いリードをトリミング 2) 遺伝子の同定 〔参照 DB〕NCBI の Express sequence Tag (EST) 3) 発現量解析 • ノーマライズ : TPM (number of transcripts per million clean tags) • FDR ($p \leq 0.01$ かつ $\log_2 \text{ratio} > 1$)	GM ナタネ ナタネにアグロバクテリウムによる形質転換により hrf2 を導入 • Untransformed Y4 • Transgenic T-4	[clean reads] • Y4 : 5,645,091 • T-4 : 6,077,037	<u>導入遺伝子による影響/機能解析</u> ハルピン媒介植物応答の分子機構に関する研究として、RNA-seq 解析を実施。

*1 データクリーニング: NGS の生データ (測定直後) からのノイズデータの除去、マッピング/アライメント: 既知の参照ゲノムに対してシーケンスリードをマップすること。
*2 Raw reads: 生データ (測定直後) のリード数、clean reads: データクリーニング後のリード数、Reads mapped to genome: マッピング/アライメントされたリード数、Average depth/ Coverage: 平均のシーケンス深度 (冗長度) もしくはカバレッジ

文献 ID	タイトル	著者	掲載誌	測定機器	アルゴリズム*1	解析対象	測定したリード数*2	実施内容
ID-042	Comparative whole-genome analyses of selection marker-free rice-based cholera toxin B-subunit vaccine lines and wild-type lines	Kashima K et al.	BMC Genomics. 2015; 16:48.	Illumina Hiseq2000 ・100bp ペアエンド	1) データクリーニング ・CASAVA software 2) マッピング/アライメント ・Burrows-Wheeler Aligner (BWA ver. 0.5.9) 〔参照ゲノム〕 Os-Nipponbare: Reference -IRGSP-1.0 build 5 ・SAMtools ・BEDTools 3) SNP/Indel 検出 ・SNPEff	GM イネ ・MucoRice-CTB : アグロバクテリウムによる形質転換により CTB (コレラトキシン B サブユニット) を導入 *コレラに対する経口ワクチンとして	〔Raw Reads〕 ・323,436,829～384,384,291 〔Average depth/ Coverage〕 ・84.3～101.3×	<u>分子キャラクタリゼーション</u> MucoRice-CTB (コレラトキシン B サブユニット) を導入したイネゲノムについて、ゲノム変異 (SNP や InDel) を野生型イネと比較した。
ID-043	Current and New Approaches in GMO Detection: Challenges and Solutions	Fraiture MA et al.	Biomed Res Int. 2015; 2015:392872.	■Targeted Sequencing ・454 システム ・Illumina ・PacBio ■Whole Genome Sequencing. ・Illumina (HiSeq 含む) ・PacBio	*レビューのため、個々の情報については記載なし	GMP ■Targeted Sequencing (ID015 を引用) ・トウモロコシ ・ダイズ ・vip3Aa20 ・maize endogen gene : Bt11 gene, Bt176 gene ・soybean endogene gene : 35S/CTP4 construct, CP4-EPSPS element, p35S promoter, ■Whole Genome Sequencing ・イネ	*レビューのため、個々の情報については記載なし	<u>検出法レビュー</u> qPCR の技術的課題に対応するため、代替 GMO 検出法が開発されている。迅速な GM ターゲットの検出正確な GM 標的の定量化、既知/未知の GMO のキャラクタリゼーションが可能であり、ゲノム、導入した塩基配列の情報が入手できない場合、こうした GMO とキャラクタリゼーションができるのは WGS だけである。
ID-045	The wheat Sr50 gene reveals rich diversity at a cereal disease resistance locus	Mago R et al.	Nat Plants. 2015; 1:15186.	Roche 454	1) データクリーニング ・記載なし 2) マッピング/アライメント ・Newbler v2.3 3) 遺伝子同定/予測 ・FGENESH ・GENSCAN	コムギの黒さび病抵抗性 Sr50 遺伝子と、CC-NB-LRR をコードするオオムギ Mla 遺伝子	* Brief Communication のため、記載なし	<u>分子キャラクタリゼーション</u> コムギの黒さび病抵抗性 Sr50 遺伝子の配列と、CC-NB-LRR をコードするオオムギ Mla 遺伝子の配列の相同性を、物理地図、変異、相補性から同定し、Sr50 はライムギ染色体 1RS 上の Sr31 等とは異なる抵抗特異性を示した。
ID-050	Cell-autonomous-like silencing of GFP-partitioned transgenic Nicotiana benthamiana	Sohn SH et al.	J Exp Bot. 2014; 65(15): 4271-4283.	Illumina Hiseq2000	1) データクリーニング ・信頼性の低いリード (Q<20)、アダプター配列の排除 2) マッピング/アライメント ・CLC Genomics Workbench ver.5.1.5 ・R-3.02 3) 遺伝子発現解析 ・FDR ($p \leq 0.01$) 4) 遺伝子同定 ・BLASTx (E<10 ⁻⁵ , NCBI) ・Blast2GO	GM タバコ ベンサミアナタバコに 35S-GFP を導入	〔clean reads〕 ・GFP 発現株 : 87,122,428 ・サイレンシング株 : 88,841,280	<u>導入遺伝子による影響/機能解析</u> トランスジェニックなベンサミアナタバコの葉の分割サイレンシングの仕組みの解明を目的とし、35S-GFP 発現トランスジェニック植物の葉における RNA-seq 解析を行った。

*1 データクリーニング: NGS の生データ (測定直後) からのノイズデータの除去、マッピング/アライメント: 既知の参照ゲノムに対してシーケンスリードをマップすること。
*2 Raw reads: 生データ (測定直後) のリード数、clean reads: データクリーニング後のリード数、Reads mapped to genome: マッピング/アライメントされたリード数、Average depth/ Coverage: 平均のシーケンス深度 (冗長度) もしくはカバレッジ

文献 ID	タイトル	著者	掲載誌	測定機器	アルゴリズム*1	解析対象	測定したリード数*2	実施内容
ID-054	Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis	Feng Z et al.	Proc Natl Acad Sci U S A. 2014; 111(12):4632-7.	Illumina Hiseq2500	1) データクリーニング ・記載なし 2) SNP 同定 ・SAMtools 〔参照ゲノム〕 <i>Arabidopsis thaliana</i> TAIR 10 ・Blastn アルゴリズムで検出し、IGV software で off-targets を確認	GM シロイヌナズナ シロイヌナズナに CRISPR-CAS により CRISPR-Cas T-DNA を導入	〔clean reads〕 ・ 87,180,454～98,180,740 〔Average depth/ Coverage〕 ・ 58.70～66.11×	<u>分子キャラクタリゼーション</u> シロイヌナズナに CRISP/Cas により 12 の異なる標的部位に 7 つの遺伝子を導入し、遺伝子変異パターン、導入効率、特異性及び遺伝率について、継世代 (T0~T3)での影響を、NGS (WGS) で調べている。
ID-130	Characterization of GM events by insert knowledge adapted re-sequencing approaches	Yang L et al.	Sci Rep. 2013; 3:2839.	Illumina HiSeq2000 ・ Paired-End library preparation kits 及び Oligos for adaptor ligation DNA でライブラリ調製 (平均 500bp) ・ 45bp×2	1) データクリーニング ・ BGI-Shenzhen (Shenzhen, China) -短いリード及び低品質のリード、アダプター配列のトリミング 2) マッピング/アライメント ・ Burrows-Wheeler Aligner (BWA) ver.0.6.2 ・ Assembly by Short Sequences (ABYSS) アルゴリズム ver. 1.3.4.2 〔参照ゲノム〕 TIGR7.0 〔外来遺伝子〕 pBar-1c、pFHBT1 及び pGL2RC7 3) 高次解析 ・ De novo 解析/BLAST 解析	イネ (組換え前 ; Minghui 63) アグロバクテリウムによる形質転換で Cry1C 及びホスフィノトリシンアセチルトランスフェラーゼ [Bar]を導入 (T1c-19) パーティクル・ガン法によるコトランスフォーマー ショ ン により、Cry1Ab(c), bla 及び hpt (pFHBT1) rice chitinase 2 (Cht-2) 及び hpt を導入 (TT51-1)	〔clean reads〕 ・ T1c-19 : 99,491,940 ・ TT51-1 : 110,276,382 〔Average depth/Coverage〕 ・ T1c-19 : 23.8× ・ TT51-1 : 26.4×	<u>分子キャラクタリゼーション/解析手法開発</u> ①形質転換ベクターの完全な DNA 配列が利用可能な場合、②形質転換ベクターの DNA 配列は利用できないが、挿入遺伝子の配列は導入遺伝子エレメント配列ライブラリに含まれる(少なくとも 1 つの遺伝子エレメントを含むことが予想される)場合、③導入遺伝子挿入物に関する DNA 配列情報が利用できず、導入遺伝子配列ライブラリの利用が限定される場合、の 3 種のケースにおける分子キャラクタリゼーションの方法が提案されている。②及び③については、de novo 解析及びコンティグ解析が用いられている。

表 3.2.2-2 植物のゲノムサイズ及び染色体構造

	和名	英名	(代表的な) 学名	ゲノムサイズ (Mb)	染色体の構造
1	イチゴ	strawberry	<i>Fragaria x ananassa</i>	240	八倍体 (Octoploid)
2	亜麻	flax	<i>Linum usitatissimum</i>	373	二倍体 (diploid)
3	イネ	Rice	<i>Oryza sativa</i>	385	二倍体 (diploid)
4	バナナ	banana	<i>Musa acuminata</i>	523	[同質] 三倍体 ([auto-] triploid)
5	テンサイ	Sugar beet	<i>Beta vulgaris</i>	758	二倍体 (diploid)
6	ジャガイモ	potato	<i>Solanum tuberosum</i> L	860	[同質] 四倍体 ([auto-] tetraploid)
7	タバコ	tobacco	<i>Nicotiana tabacum</i>	900	複二倍体 (amphidiploid)
8	大豆	Soybean	<i>Glycine max</i>	1,115	二倍体 (diploid)
9	クロガラシ	black mustard	<i>Brassica nigra</i>	468～760	異質倍数体 (allopolyploid) 複二倍体 (amphidiploid)*
	(学名)	(学名)	<i>Brassica rapa</i>	468～784	
	ヤセイカンラン	wild cabbage	<i>Brassica oleracea</i>	599～868	
	カラシナ	Mustards	<i>Brassica juncea</i>	922～1495	
	ナタネ	Oilseed	<i>Brassica napus</i>	1,127～1,235	
	アビシニアガラシ	Abyssinia mustard	<i>Brassica carinata</i>	1,284～1,544	
10	ワタ (綿)	Cotton	<i>Gossypium hirsutum</i>	2,250	異質四倍体 (allotetraploid)
11	トウモロコシ	Maize	<i>Zea mays</i>	2,300	二倍体 (diploid)* ¹
12	コムギ	Wheat	<i>Triticum aestivum</i>	17,000	異質四倍体 (allotetraploid)* ²
(参考)	シロイヌナズナ	(学名)	<i>Arabidopsis thaliana</i>	130	二倍体 (diploid)

*¹ 同質四倍体 (autotetraploid) の報告もある。*² 六倍体 (hexaploid) の報告もある。

- ・同質倍数体 (autopolyploid): 一つの倍数体が同一種類のゲノムからだけなる。二倍体 (diploid)、[同質] 四倍体 ([auto-] tetraploid) など。
- ・異質倍数体 (allopolyploid): 異なる 2 種類以上のゲノムで構成されている倍数体。異なるゲノムを持つ植物間の雑種に由来すると考えられている。異質四倍体 (allotetraploid) など。

* 複二倍体 (amphidiploid): 異質倍数体の中でゲノム構成が AABB や AABBCC のように、2 種類以上のゲノムがそれぞれ 2 個ずつ存在すると、減数分裂において二価染色体のみが形成され、二倍体と区別できない。この型の異質倍数体をとくに複二倍体 (amphidiploid) という。

3.2.3. 行政等におけるNGS活用に向けた取り組み状況

(1) OECD

OECD は 2016 年 4 月 18 日に遺伝子組換え植物の安全性評価におけるハイスループット DNA シーケンスに関する OECD ワークショップを開催した。各国の取り組み状況を共有し、ワークショップでの発表内容をまとめた文書を公表している (OECD, 2016、[文献 ID-131])。

(2) カナダ

カナダでは、GMO に由来する製品は販売前に安全性評価を行う必要があり、従来は分子キャラクタリゼーションには古典的な分子生物学的技術が用いられていた。しかし、2013 年に、いくつかの企業から分子キャラクタリゼーションのサポートとして全ゲノムシーケンシング (WGS) のデータが提出されるようになり、カナダ保健省 (HC) やカナダ食品検査庁 (CFIA) の分子キャラクタリゼーションのデータ評価者に新規のバイオインフォマティクスの知識が必要とされるようになった。

そこで、最初の取り組みとして、2015 年 3 月にカナダ、米国及びメキシコの行政関係者、並びに工業会及びアカデミックの専門家を参集したワークショップを CFIA が開催した。その結果として、ピアレビュー文献ならびに DNA シーケンシング及びバイオインフォマティクスに関するカナダ政府の専門家の協力により、内部向けのガイダンス文書を作成することが結論された。その後、HC 及び CFIA の評価者を対象とした 2015 年 11 月に開催されたワークショップにおいて、内部向けガイダンス案が示された。この文書は、詳細な技術情報、ゲノムシーケンス解析における専門用語及び基本的概念、更には WGS の利点と限界性に関する情報を提供することが目的とされ、文書の中には仮想データに基づくケーススタディも含まれている。

このガイダンス案はワーキンググループにより最近確定され、今後、内部向けガイダンスから抽出した重要な原則に基づき、企業向けに重要ポイントを示したガイダンスを作成予定である (OECD, 2016、[文献 ID-131])。

(3) 欧州

従来使用される技術 (例 ; サザンブロット、サンガーシーケンシング) と比較して、このような技術が遺伝子組換え植物の分子キャラクタリゼーションにおいてどのような価値を提供しうるかを検討するために、2013 年 11 月に国際ワークショップが開催されている (OECD, 2016、[文献 ID-131])。

また、欧州では、遺伝子改変による非意図的な副作用をスクリーニングするために RNA-Seq を用いた研究を実施した GRACE プロジェクト¹³ (GMO Risk Assessment and Communication Evidence) があつた (2012 年 7 月～2015 年 11 月)。さらに、DECATHLON プロジェクト¹⁴ (2013 年 12 月～

¹³ <http://www.grace-fp7.eu/>

¹⁴ <http://www.decathlon-project.eu/>

2016 年 12 月) では、食品病原体、GMO のトレーサビリティ、関税問題に利用できる費用対効果の高い高度な DNA ベースの方法の開発することを目的として、NGS を含む DNA ベースの関連技術についての最小性能パラメータ (MPPs ; Minimal Performance Parameters) の開発も行われている。

さらに、2016 年 2 月に欧州委員会共同研究センター (JCR) から「Guideline for the submission of DNA sequences and associated annotations within the framework of Directive 2001/18/EC and Regulation (EC) No 1829/2003 (文献 ID-133)」が公表されており、指令 2001/18 / EC および 1829/200 規則に基づく義務の一貫として、申請者の提出情報に含まれるインサート及び隣接領域の配列の取得及び提出に関する最小要件及び推奨事項が記載されている(報告様式含む)。このガイドラインでは、申請者が提出するインサート及び隣接領域の配列について、NGS を使用した場合に必要な情報として、最適なシーケンシングの深さを要求しており、通常 40 未満ではない、と記載されている。また、NGS を用いた際のインフォマティクスに関する報告要求事項として、①ベースコール手順の詳細(使用ツール(名称、バージョン)、使用したオプション及びパラメータ、ベースコールの作業時に認められた不確実性を含む)、②FASTQ 形式での NGS リード生データ(フィルターされ、アダプター配列が除かれたもの)及びフィルター及びトリミングに使用したソフト及びパラメータ、③アライメント/マッピングに使用する配列(SAM、BAM または CRAM 形式)等が挙げられている(文献 ID-133)。

(4) オーストラリア・ニュージーランド (FSANZ)

調査した範囲では、NGS の評価に関わる取り組みがわかる資料は無かった。しかしながら、2016 年に公表された MON87419 トウモロコシの評価書において「サザンブロットあるいは NGS/JSA のような分子技術が遺伝的安定性研究には最も適している (Molecular techniques, such as Southern blot analysis or NGS/JSA, are considered the most appropriate techniques for studying genetic stability.)」と述べており (FSANZ 2016、[文献 ID-076])、独自に NGS データの活用に関する知見を収集していると考えられる。

(5) 工業界等

2016 年 4 月 18 日に開催された遺伝子組み換え植物の安全性評価におけるハイスループット DNA シーケンスに関する OECD ワークショップにおいて、工業会からは以下の見解が示されている(文献 ID-132)。なお、Dow AgroSciences LLC、Syngenta Crop Protection LLC、BASF Plant Science LLP、Bayer CropScience、DuPont Pioneer、Monsanto Company、CropLife International がワークショップに参加している。Monsanto 社は、全ゲノムシーケンス (WGS) アプローチへの適用例を示し、T-DNA と境界ゲノムシーケンスの間のジャンクション領域をシーケンシングすることで、ダイズの T-DNA 挿入及びコピー数の決定に成功し、挿入及びコピー数の評価を通して、経世代の挿入の完全性及び安定性を決定することができることを提唱した (Kovalic et al.,2012, 文献 ID-19)。DuPont Pioneer 社は、ハイスループットの遺伝子組換え系統の選定において NGS とターゲットシ

ーケンスキャプチャー（TSC）の適用に成功した（Zastrow-Hayes et al., 2015, 文献 ID-010）。さらに、Dow AgroSciences は単一及び積層（stacked）遺伝子改変種のキャラクタリゼーションに対する WGS 及び TSC アプローチの比較解析に関連するデータを発表している（Guttikonda et al., 2016, 文献 ID-004）（文献 ID-132）。

T-DNA の挿入、コピー数及びバックボーンの有無の確認は Junction Sequence Analysis により実施が可能であり、10X（Guttikonda et al., 2016）から 75X（Kovalic et al., 2012）に及ぶ NGS カバレッジにより、シングルコピーの挿入あるいはゲノムへの複数の挿入を確認できる。また、WGS データをベクターバックボーンにマッピングすることでベクターバックボーンが存在が確認できる。プラスミド及びリファレンスゲノムにリードをマッピングすることで、T-DNA の完全性及びフランキンク境界の確認も行える。しかし、T-DNA の完全性及びフランキンク境界に NGS を適用する場合には、T-DNA 内に内因性の制御エレメント及び長い反復配列がない系統に限られる。経世代の T-DNA の安定性の判定に、WGS による Junction Sequence Analysis が利用できる。NGS はサザンブロットとは異なり、自動化が行い易く、労働集約型にならない。また、挿入/T-DNA コピー数、バックボーンの有無及び世代間安定性においてサザンブロットと同等の結果を示す（文献 ID-132）。

CropLife International は、農薬等に関する企業による国際的な組織である。2016 年 2 月にウェブページにおいて「CROP LIFE INTERNATIONAL POSITION PAPER –METHODOLOGIES TO INFORM RISK ASSESSMENT¹⁵」を公表している。過去 20 年間の GM 作物のリスク評価における方法論、例えば分子キャラクタリゼーション、タンパク質発現（protein expression）、毒性評価等の分野において、新しい方法が開発され、評価に利用されてきていることを概説している。分子キャラクタリゼーションにおける NGS の利用について、従来使用されてきたサザンブロットと同様に正確な分子キャラクタリゼーションが行える、としている。

¹⁵https://croplife.org/wp-content/uploads/2016/01/CLI-Methods-Flexibility-Position-Paper_CLI-Final_09_16_2015.pdf

3.3. 遺伝子組換え植物の安全性評価の際に一般的に用いられている分析法によるデータと NGS により作成されたデータの比較・検討

3.3.1. 次世代シーケンサー (NGS) の機器メーカー毎の特性

NGS はプラットフォームに使用されるメーカーや機種によって測定原理が異なる等により、性能や実施できる内容が大きく異なることから、調査に先立ち、次世代シーケンサー (NGS) の機器メーカー毎の特性について整理した。

NGS 機種の性能比較の概要として、1 リードの長さ及び測定あたりのデータ量の関係を図 3.3.1-1 に示す。また、NGS の各世代における特長を表 3.3.1-1 に示す。

一般に 1 リードの長さが 1000 bp 未満の場合をショートリード型、1000 bp 以上の場合をロングリード型に分類しており、それぞれの型に分類される機種の性能をそれぞれ表 3.3.1-2 及び表 3.3.1-3 に示す。

また、各メーカーが採用している測定原理を把握しておくことも重要と考えられることから、従来使用されているサンガー法を含めた測定原理を別添資料 3-5 にまとめた。

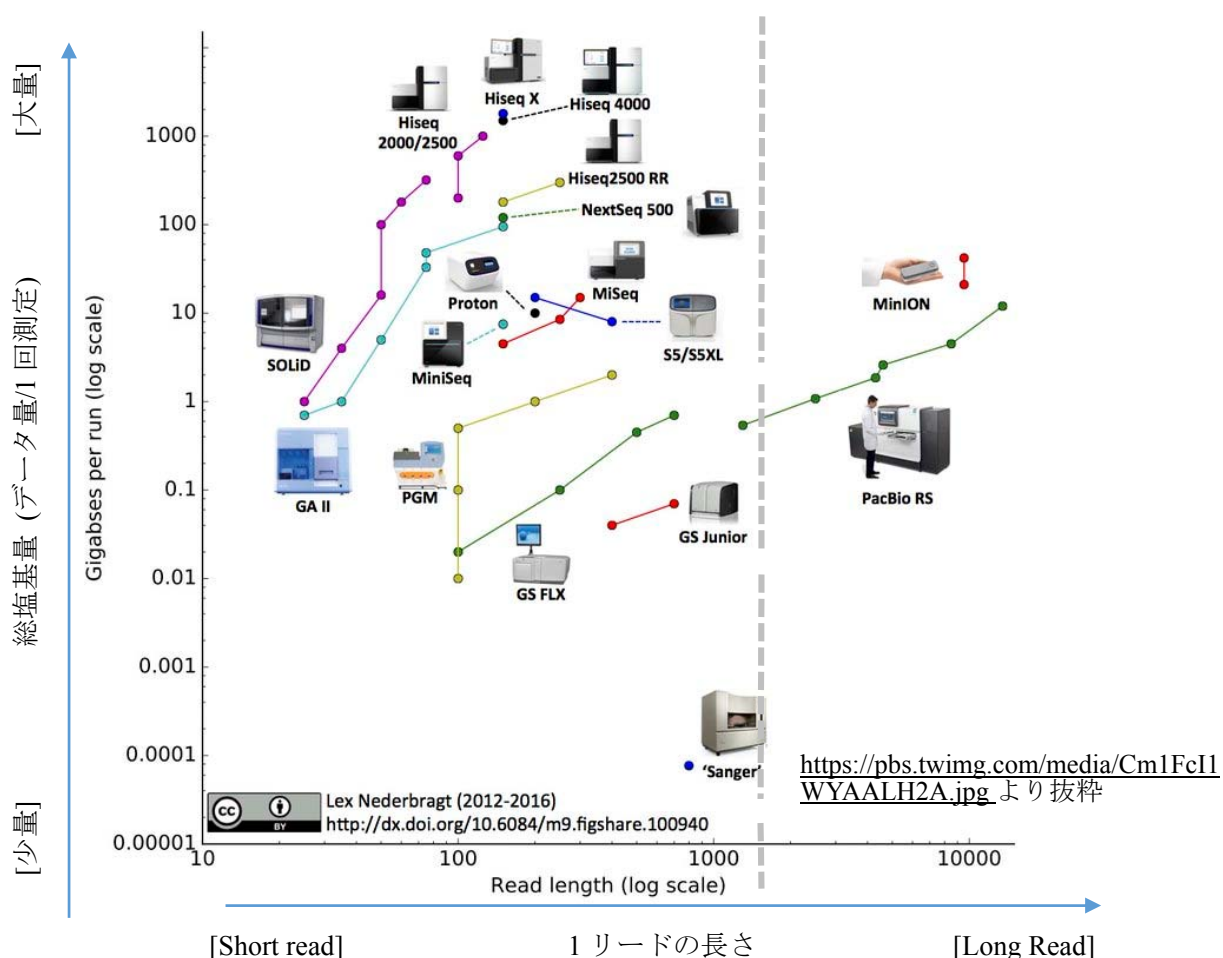


図 3.3.1-1 次世代シーケンサー (NGS) の測定量の比較 [機種別]

表 3.3.1-1 NGS の各世代における特長

分類	NGS (=第 2 世代)	3 rd Generation Sequencing (=第 3 世代)	
	DNA 合成・光検出法を用いた超並列シーケンシング	1 分子リアルタイム ・シーケンシング	半導体シーケンシング
機種	・ GS FLX、GS Jounior ・ MiniSeq/MiSeq/NextSeq/HiSeq ・ SOLiD	・ PacBio RS	・ Ion PGM ・ Ion Torrent (S5/S5XL) ・ MinION (Nanopore)
特長	DNA 合成法を用いて、蛍光・発光など光検出により、超並列的に塩基配列を決定する	1 分子を鋳型として DNA 合成を行い、1 塩基ごとの反応を蛍光・発光などで検出→リアルタイムに塩基配列を決定する	専用の半導体チップ上に並んだ数 μm のウェルの中でシーケンシング反応→リアルタイムに塩基配列を決定する
1 リード [※] の長さ	25～1kb 程度	10 kb～40kb	100 b (Ion Torrent) 2kb (Nanopore)
総塩基量	35M～2Tb	0.5G～9 Gb	30M～8Gb
精度	優れている	比較的優れている	比較的優れている (Ion Torrent) エラーが多い (Nanopore)

表 3.3.1-2 ショートリード型シーケンサー

メーカー	機器名称	解析法 ^{*1}	リード長 (bp)	リード数	総塩基量 (bp)	エラー率 ³	精度 ^{*4}	その他
Illumina	MiniSeq	SBS	[最小] 75×1	2,500 万	1.6～1.9 Gb	～0.1%	Q30 以上が80%以上	中出力キット
			[最大] 150×2	5,000 万	6.6～7.5 Gb		Q30 以上が80%以上	高出力キット
	MiSeq	SBS	[最小] 36×1	1,200 万～1,500 万	0.5～0.6 Gb		Q30 以上が90%以上	V2 試薬
			[最大] 300×2	4,400 万～5,000 万	13.2～15 Gb		Q30 以上が70%以上	V3 試薬
	NextSeq	SBS	[最小] 75×2	1 億～1.2 億	16～19 Gb		Q30 以上が80%以上	中出力フローセル
			[最大] 150×2	3 億～4 億	100～120 Gb		Q30 以上が75%以上	高出力フローセル
	HiSeq 2000/2500	SBS	[最小] 36×1	13 億～14 億	47～52 Gb		Q30 以上が85%以上	TRUSEQ SBS V3 (1 枚)
			[最大] 125×2	36 億～40 億	900～1,000 Gb		Q30 以上が80%以上	TRUSEQ SBS V3 (2 枚)
	HiSeq 2500 RR ^{*2}	SBS	[最小] 36×1	2.5 億～3 億	9～11 Gb		Q30 以上が85%以上	HISEQRAPIDSBSKITV2(1枚)
			[最大] 250×2	5 億～6 億	250～300 Gb		Q30 以上が75%以上	HISEQRAPIDSBSKITV2(2枚)
	HiSeq 4000	SBS	[最小] 50×1	21 億～25 億	105～125 Gb		Q30 以上が85%以上	フローセル 1 枚
			[最大] 150×2	43 億～50 億	1300～1500 Gb		Q30 以上が75%以上	フローセル 2 枚
	HiSeq X	SBS	[最小] 150×2	26 億～30 億	800～900 Gb		Q30 以上が75%以上	フローセル 1 枚
			[最大] 150×2	53 億～60 億	1,600～1,800 Gb		Q30 以上が75%以上	フローセル 2 枚
Thermo Fisher	SOLiD	SBS	[最小] 50×1	16 億～32 億	80～160 Gb	<0.1%	—	フローセル 1～2 枚
			[最大] 150×2	5 億～10 億	160～320 Gb		—	フローセル 1～2 枚
Roche	GS FLX	PS	500～700	100 万	0.7 Gb	1%	400bp のQ20 リード長	—
	GS Junior	PS	500～700	10 万	0.035 Gb		400bp のQ20 リード長	—

メーカー	機器名称	解析法 ^{*1}	リード長 (bp)	リード数	総塩基量 (bp)	エラー率 ^{*3}	精度 ^{*4}	その他
Thermo Fisher	Iontorrent S5/ S5XL	HID	[最小] 200	300～500 万	0.6～1 Gb	～1%	—	Ion 520 チップ
			[最大] 400	1,500 万～2,000 万	6～8 Gb		—	Ion 530 チップ
	Ion Proton ^{*5}	HID	～200	6,000 万～8,000 万	～10 Gb	～1%	—	—
Thermo Fisher	Ion PGM ^{*6}	HID	[最小] 200	40 万～55 万	0.03～0.05 Gb	～1%	—	Ion 314 Chip v2
			[最大] 400	400～550 万	1.2～2 Gb		—	Ion 318 Chip v2

^{*1} SBS (Sequencing By Synthesis)、PS (PyroSequencing)、HID (Hydrogen Ion Detection).

^{*2} RR: Rapid Run モード.

^{*3} 数字は 2014 年データ. なお、Sanger 法 (capillary sequencing) では 0.1～1%のエラー率.

^{*4} Q scores はそれぞれの塩基が不正確に BaseCall される可能性を示す (Q30: 1 in 1000、Q20: 1 in 100).

^{*5} 価格は 26,900,000 円～ (<https://www.wakenyaku.co.jp/ctg/ls.php?i=46> より)

^{*6} 価格は 7,700,000 円～ (<https://www.wakenyaku.co.jp/ctg/ls.php?i=46> より)

表 3.3.1-3 ロングリード型型シーケンサー

メーカー	機器名称	解析法 ^{*1}	リード長 (bp)	リード数	総塩基量 (bp)	エラー率 ^{*2}	精度	その他
Pacific Biosciences	PacBio RSII ^{*3} / Sequel System ^{*4}	SMRT	10k～40k	～80 万	0.5～9 Gb	<1%	1 本リードあたり86%	—
NANOPORE	MinION	ONS	～200kb	?	?	4%	—	—

^{*1} SMRT (Single Molecule, Real-Time sequencing)、ONS (Oxford Nanopore system).

^{*2} 数字は 2014 年データ. なお、Sanger 法 (capillary sequencing) では 0.1～1%のエラー率.

^{*3} 価格は 128,000,000 円～ (<https://www.wakenyaku.co.jp/ctg/ls.php?i=46> より)

^{*4} 価格は 78,000,000 円～ (<https://www.wakenyaku.co.jp/ctg/ls.php?i=46> より)

3.3.2. 遺伝子組換え植物の安全性評価に使用される従来法によるデータとNGSにより作成されたデータの比較・検討

比較対象とする従来法をサザンブロット及び PCR とサンガーシーケンスとし、NGS 解析として全ゲノムシーケンシング (WGS)、ターゲットシーケンシング及びロングリードシーケンシングについて、主要な機器、必要サンプル量、原理/方法、分かること/できること、塩基長/リード、検出感度、分析精度、網羅性、メリット、デメリットを整理した結果を表 3.3.2-1 に示す。また、従来法と新手法 (NGS 解析) のそれぞれの利点をまとめたものを図 3.3.2-1 に示す。

「分かること/できること」については、表 3.3.2-1 に示すように NGS 解析は T-DNA の挿入コピー数についてはサザンブロットに若干劣るものの、WGS、ターゲットシーケンシング及びロングリードシーケンシングのいずれの NGS 解析においても、バックボーン配列の有無、T-DNA 挿入の安定性、T-DNA の完全性 (低深度の場合定性的に、高深度の場合 1 塩基レベルの精度で)、T-DNA/ホストゲノムの接合部分の DNA 配列が可能であり、サザンブロット及び PCR とサンガーシーケンスを組み合わせることなく、これらの情報を取得することができる。さらに、ホストゲノム側の欠損/挿入については、表 3.3.2-1 に示すいずれの NGS 解析方法においても、従来法である PCR とサンガーシーケンスの組合せに比べ、高い網羅性、高い精度で知見を得ることができる。さらに、ターゲットシーケンスでは多検体解析が可能であり、ロングリードシーケンスでは繰り返し配列の同定が可能となる。ただし、NGS 解析で得られる知見の信頼性については、シーケンス深度 (Depth) が影響するため留意が必要である。また、WGS の場合、PCR バイアスの影響、繰り返し配列に対する分析精度の低さ等のデメリットがある。ターゲットシーケンスの場合には、PCR バイアスの影響の他に、煩雑な前処理及びゲノム上の挿入位置が同定できない場合がある (WGS と異なり、ゲノム全体の配列情報が得られないため) といったデメリットがある。ロングリードシーケンスでは、現時点では高コストであるために多検体解析に向かないといったデメリットが考えられた。

ここで、各 NGS 解析の代表的な GMO 解析法を図 3.3.2-2~3.3.2-6 に示した。

表 3.3.2-1 従来法と NGS 解析との比較 (ゲノム DNA: ゲノム上への T-DNA の挿入確認)

比較項目	従来法		NGS (Next generation Sequencing)		
	サザンブロット	PCR 法+サンガーシーケンス	WGS (Whole genome sequencing)	Target sequencing (Capture sequencing)	Long Read sequencing
主な測定機器	—	Capillary sequencer	Illumina Hiseq	Illumina Miseq	PacBio RSII MINIon
必要サンプル量	1～10 µg	1 ng～1 µg	3～10 µg	1～10 µg	3～10 µg
原理/方法	・ゲノム DNA を制限酵素処理し、電気泳動した後、メンブランに転写し、DIG/放射ラベルした T-DNA 配列由来のプローブとハイブリダイズさせる	(・3′/5′RACE でゲノム上の挿入位置を同定) ・T-DNA 配列特異的プライマーとホストゲノム特異的プライマーで、PCR 反応で増幅 ・サンガー法で DNA 配列を解読	・断片化したゲノム DNA を PCR 増幅によりシーケンスライブラリを作成し、蛍光・発光など光検出により、超並列的に塩基配列を決定する	(同左)	・SMRT (Single Molecule, Real-Time sequencing) 異なる蛍光色素を付加した dNTP を流し、DNA ポリメラーゼによる伸長反応で塩基が取り込まれた時だけ発する蛍光を一分子単位でリアルタイムに検出する。
分かること/できること	○T-DNA の挿入コピー数 ○バックボーン配列の有無 ○T-DNA 挿入の安定性 (継世代) ○T-DNA の完全性 (定性的) × × × × ×	× × × × ○T-DNA の完全性 (1 塩基レベル) ○T-DNA/ホストゲノムの接合部分の DNA 配列 △ホストゲノム側の欠損/挿入 (部分配列) × ×	△T-DNA の挿入コピー数 ○バックボーン配列の有無 ○T-DNA 挿入の安定性 (継世代) ○T-DNA の完全性 (定性的) :低深度 ○T-DNA の完全性 (1 塩基レベル) :高深度 ○T-DNA/ホストゲノムの接合部分の DNA 配列 ◎ホストゲノム側の欠損/挿入 (全ゲノム) × ×	△T-DNA の挿入コピー数 ○バックボーン配列の有無 ○T-DNA 挿入の安定性 (継世代) ○T-DNA の完全性 (定性的) :低深度 ○T-DNA の完全性 (1 塩基レベル) :高深度 ○T-DNA/ホストゲノムの接合部分の DNA 配列 ○ホストゲノム側の欠損/挿入 (部分的) × ◎多検体の解析	△T-DNA の挿入コピー数 ○バックボーン配列の有無 ○T-DNA 挿入の安定性 (継世代) ○T-DNA の完全性 (定性的) :低深度 ○T-DNA の完全性 (1 塩基レベル) :高深度 ○T-DNA/ホストゲノムの接合部分の DNA 配列 ◎ホストゲノム側の欠損/挿入 (全ゲノム) ◎繰り返し配列の同定 ×
塩基長/リード	—	500～700 bp	36～500 bp	36～500 bp	～40,000 bp
検出感度	低い (数百～数 Kb)	高い (1 塩基レベル)	高い (1 塩基レベル)	高い (1 塩基レベル)	高い (1 塩基レベル)
分析精度	—	高い (0.1～1%のエラー率)	高い (～0.1%のエラー率)	高い (～0.1%のエラー率)	他の NGS 機器に比べて若干劣る (<1%)
網羅性	低い (数百～数 Kb)	低い (500bp 前後)	高い (全ゲノムを対象)	中程度 (限定したゲノム領域)	高い (全ゲノムを対象)
メリット	・挿入コピー数の確認に用いられる	・DNA 配列の確認に用いられる	・「PCR 法+サンガーシーケンス」と「サザンブロット」で得られる情報が、NGS 解析のみで得られる。 ・ゲノム全体の配列情報が得られる。	・多検体の解析ができる	・PCR バイアスの影響が少ない ・繰り返し配列に対する分析精度が高い ・ゲノム全体の配列情報が得られる。 ・DNA メチル化も検出可能 (PacBio のみ)
デメリット	・DNA 配列情報が全く得られない ・ゲノム DNA の限られた部分の結果しか得られない	・T-DNA の挿入位置が同定されていないとホストゲノム特異的プライマーが設計できない (ゲノム位置同定に必要な RACE 法は時間がかかる) ・ゲノム DNA のごく一部の配列情報しか得られない	・シーケンス深度 (Depth) が低い場合、NGS データの信頼性が低くなる ・PCR バイアスの影響 ・繰り返し配列に対する分析精度が低い ・多検体の解析には不向き (高コスト)	・シーケンス深度 (Depth) が低い場合、NGS データの信頼性が低くなる ・PCR バイアスの影響 ・前処理が必要 (DNA キャプチャー、クローニング、PCR 増幅など) ・ゲノム上の挿入位置が同定できない場合がある (ゲノム全体の配列情報が得られない)	・シーケンス深度 (Depth) が低い場合、NGS データの信頼性が低くなる ・多検体の解析には不向き (高コスト)
代表的な GMO 解析法	—	—	・Junction Sequence Analysis (JSA) 法：図 3.3.2-2 及び図 3.3.2-3	・Microarray hybrid capture and next-generation sequencing：図 3.3.2-4 ・Southern-by-Sequencing (SBS) 法：図 3.3.2-5 ・HitStuf 法：図 3.3.2-6	特になし

■外来遺伝子導入における分子キャラクタリゼーション (従来法と NGS 解析との比較)

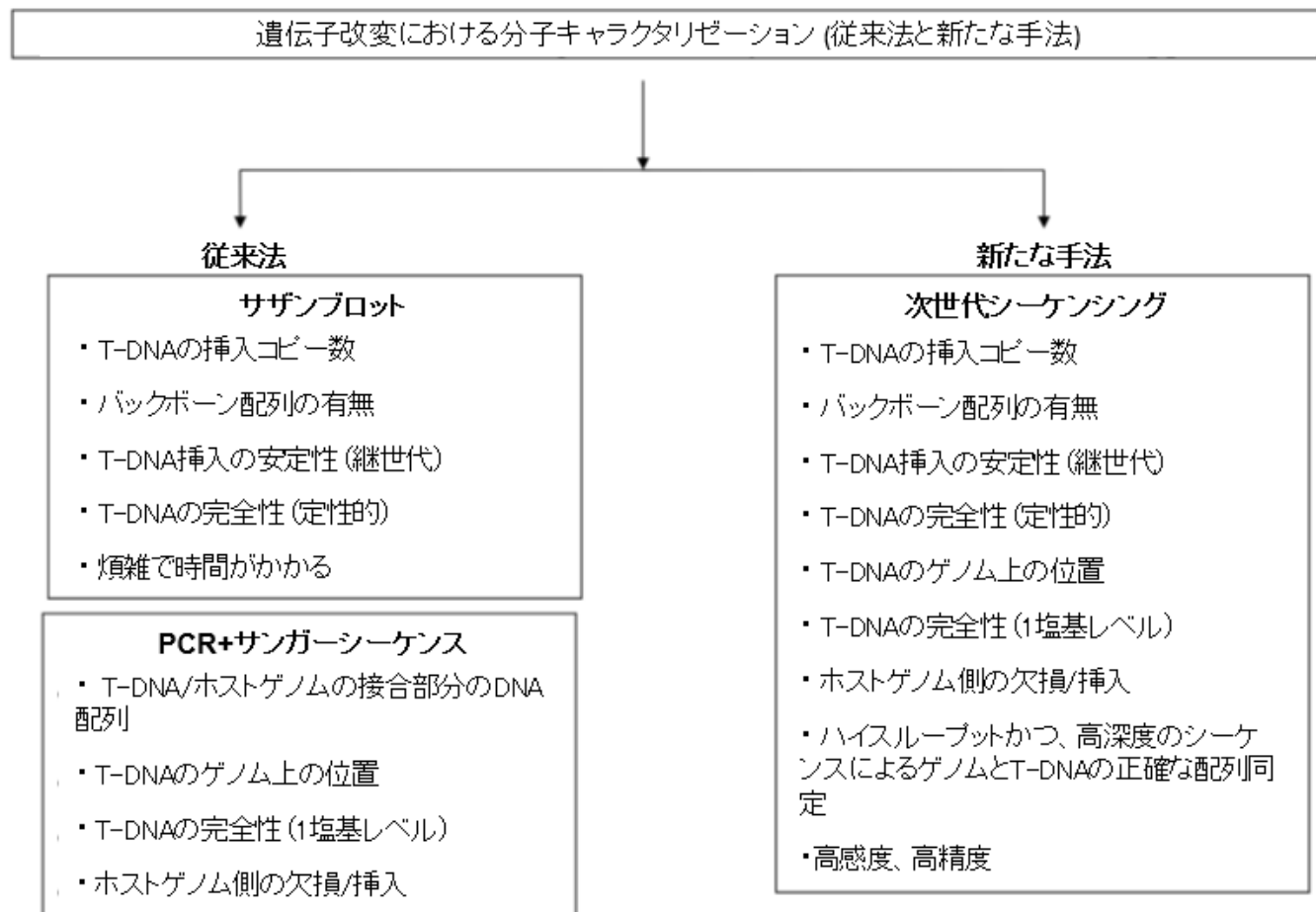


図 3.3.2-1. 従来法と新手法 (NGS 解析) のそれぞれの利点をまとめたもの (文献 ID-004) 2016 年

*一部を表 3.3.2-1 の「分かること」に記載

■具体的な NGS 解析法①：Junction Sequence Analysis (JSA) (文献 ID-019) 2012 年

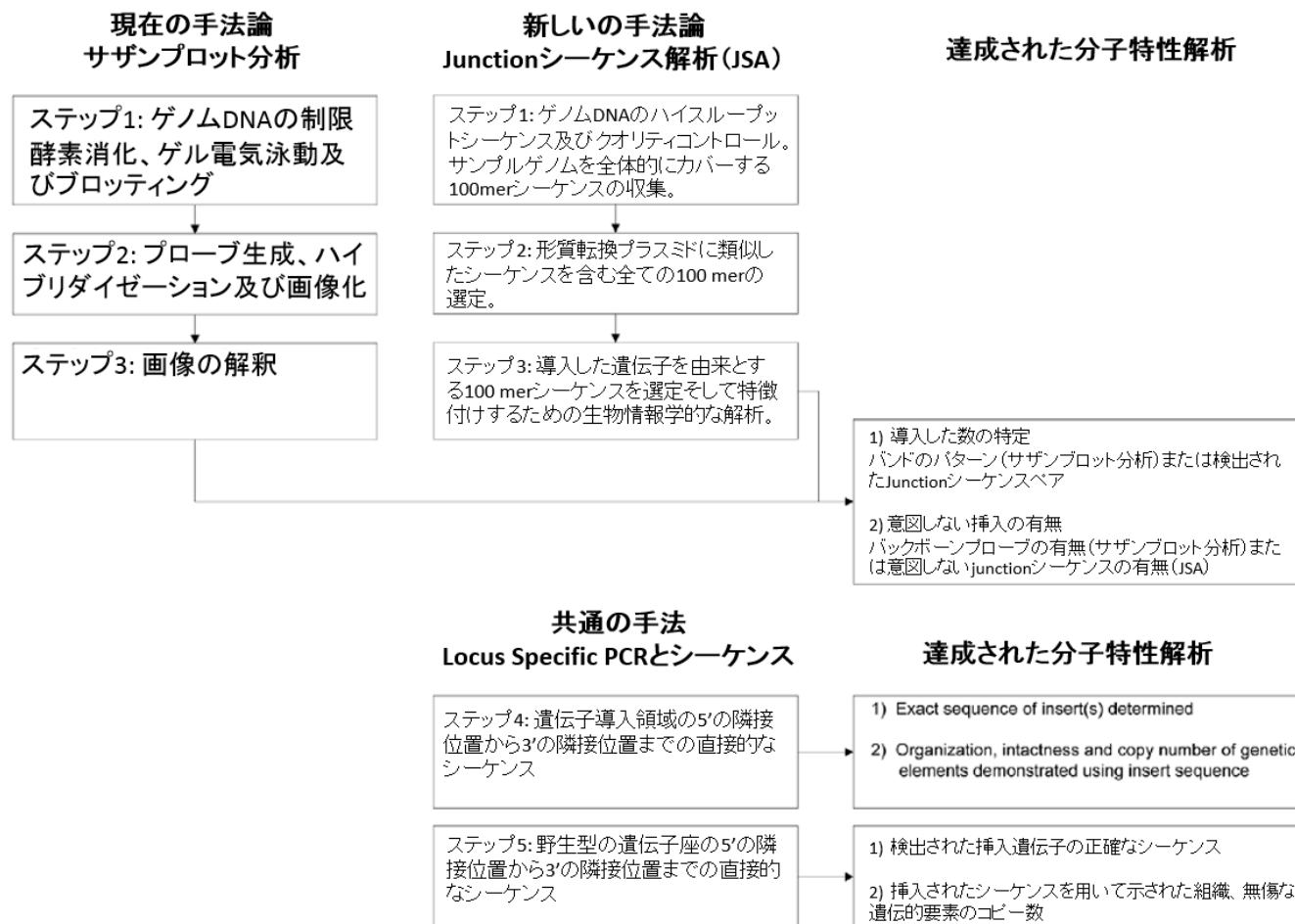


図 3.3.2-2. JSA 法と従来法との比較

■具体的な NGS 解析法①：Junction Sequence Analysis (JSA) (文献 ID-019) 2012 年〔続き〕

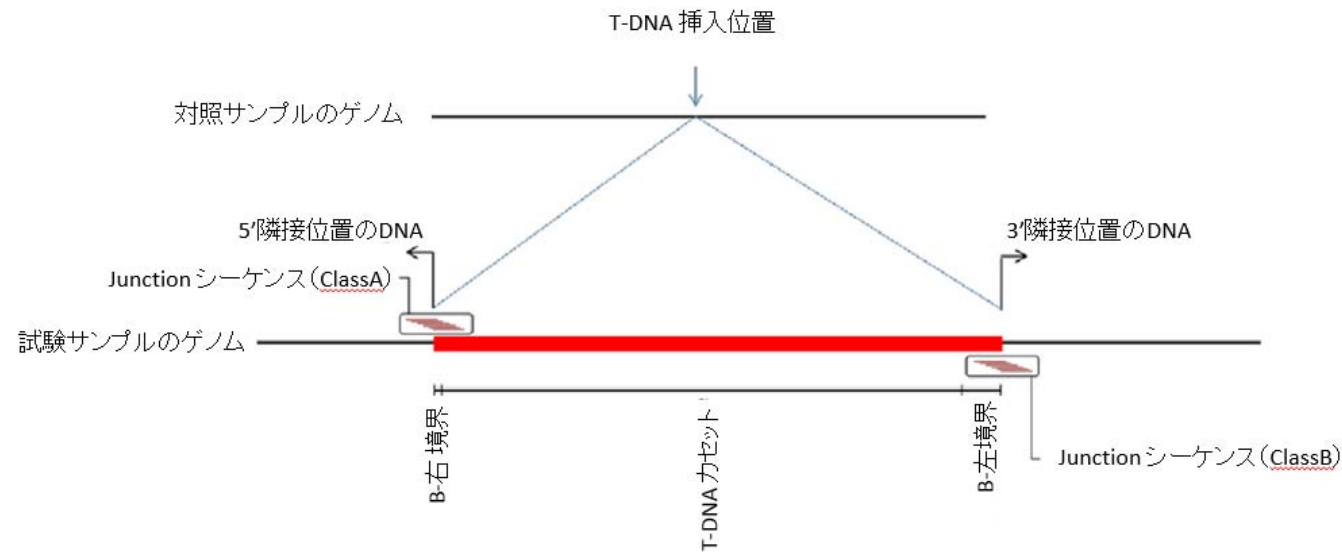


図 3.3.2-3. 外来 DNA の挿入位置/新たな接合配列同定のモデル図

〔Junction Sequence Analysis (JSA)について〕

- ステップ 1) 外来 DNA 挿入サンプルと WT サンプルについて、100 塩基長の包括的なハイスループットシーケンス (NGS 解析) を行う
- ステップ 2) 100 塩基長の NGS データから外来 DNA (形質転換したプラスミド) 配列を抽出する
- ステップ 3) バイオインフォマティクス解析により外来 DNA 配列の宿主ゲノム上の接合部分配列を同定する (Junction Sequence Detection)

(挿入された外来 DNA のコピー数確認：Junction sequence pairs を検出する) ⇔ サザンブロットではバンドパターン(本数)を調べる

(意図しない配列の挿入：導入した外来 DNA 以外の DNA 配列が検出される) ⇔ サザンブロットではバックボーン配列プローブのバンドが全く検出されない

- ステップ 4) 外来 DNA 配列挿入部位の 5'側及び 3'側の直接配列決定 (⇔ サザンブロットでも共通)
- ステップ 5) WT サンプルの (外来 DNA の挿入部位付近の) 直接配列決定 (⇔ サザンブロットでも共通)

■具体的な NGS 解析法②：Microarray hybrid capture and next-generation sequencing (文献 ID-132) 2013 年

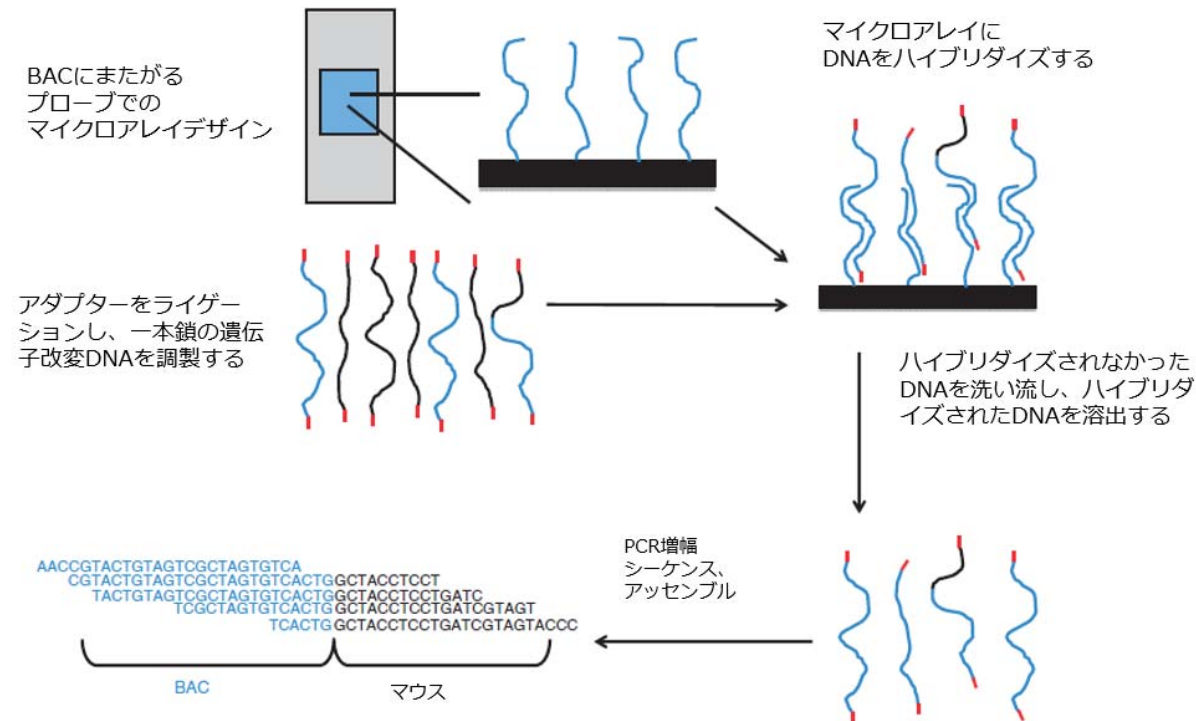


図 3.3.2-4. マイクロアレイ法による概要

- マイクロアレイ上に BAC(外来遺伝子)由来の DNA 配列に相補的なプローブをデザインして、配置した
- トランスジェニックマウス由来のゲノム DNA を断片化し、両末端にアダプター配列をライゲーションさせる (一本鎖)
- マイクロアレイとハイブリダイズさせ、非結合断片を Wash した後、ハイブリした DNA 断片を溶出させる (*ここで、BAC 配列をもつ DNA 断片がセレクトされる)
- 溶出された DNA をアダプターに特異的なプライマーを用いて PCR 増幅し、配列決定し、BAC /マウスまたは非連続的な BAC / BAC 接合部の位置を決定した

■具体的な NGS 解析法③：Southern-by-Sequencing (SBS) (文献 ID-010) 2015 年

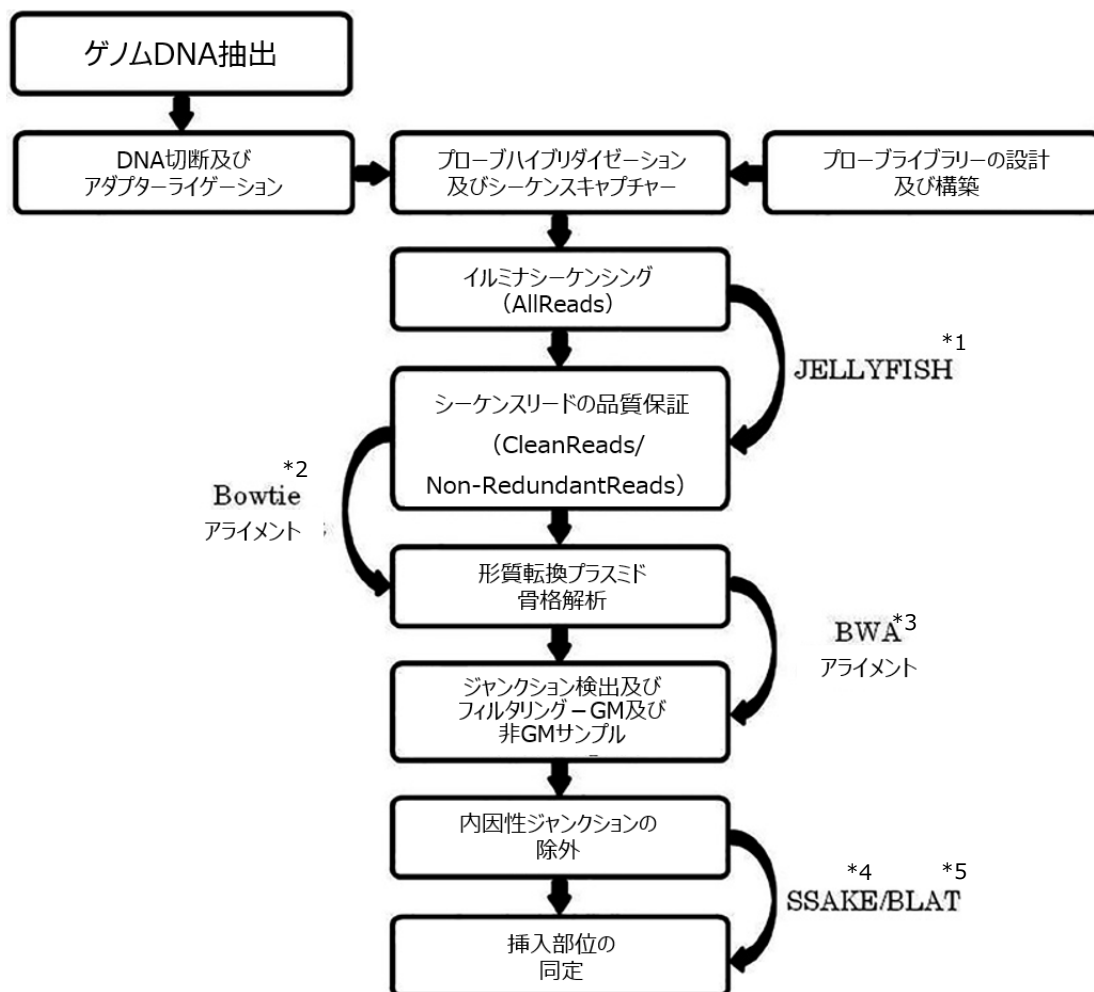


図 3.3.2-5. Southern-by-Sequencing の概要

- GM 植物/食物からのゲノム DNA の抽出
- ゲノム DNA の断片化及びアダプター配列のライゲーション (プローブライブラリーのデザイン及び構築 / Roche NimbleGen)
- シーケンスキャプチャー (プローブライブラリーとのハイブリ)
- イルミナシーケンス (HiSeq 2500)
- 一定品質以上のシーケンスリードを選定、リードのクリーニング (トリミング)
- 外来 DNA/バックボーン配列を解析する
- GM 由来データをリファレンス配列と比較し、接合配列を同定する
- 固有の接合部配列を除去する
- リファレンスゲノム及び形質転換プラスミドと整列させて接合配列の起点を決定し、挿入部位を同定する。

*1 JELLYFISH : DNA 中の k-mers を高速でメモリ効率よくカウントするためのツール。k-mer とは長さ k の部分配列であり、配列の出現を数えることは、DNA 配列の多くの分析の中心的なステップである。

*2 Bowtie : マッピング/アライメントツール

*3 BWA : 参照ゲノムに対するマッピング/アライメントツール

*4 SSAKE : アセンブルツール

*5 BLAT : Blast に類似したアライメントツール

■具体的な NGS 解析法④：HtStuf: High-Throughput Sequencing to Locate Unknown DNA Junction Fragments (文献 ID-009) 2015 年

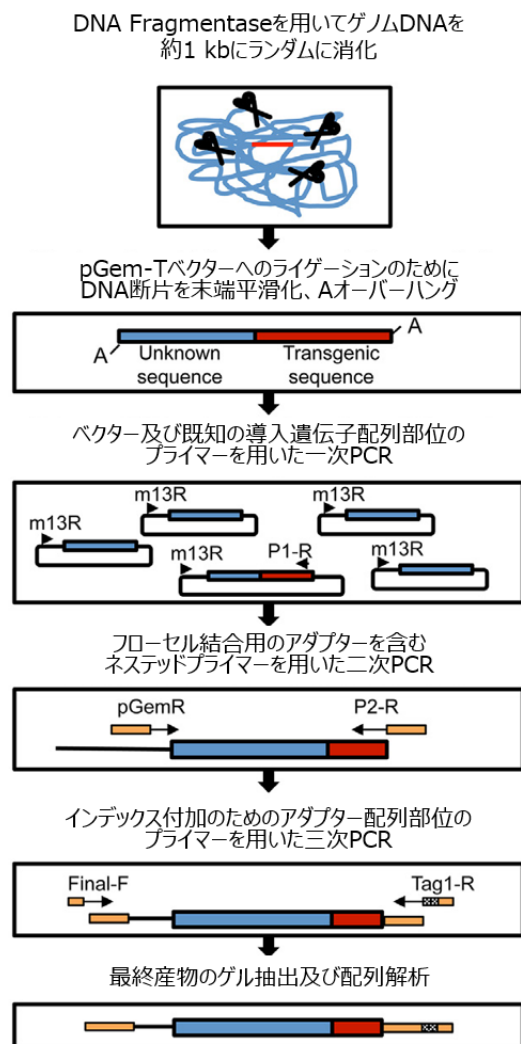


図 3.3.2-6. ライブラリ構築の概要

- GM 植物/食物ゲノム DNA を断片化酵素で約 1kb に断片化
- DNA 断片の平滑末端にA塩基をオーバーハングで付加し、pGem-T vector にライゲーション
- 1 回目の PCR 反応 (ベクター配列と既知の外来 DNA 配列で実施)
↓*ここで外来 DNA 配列を持たないものは増幅されない (シーケンス解析する配列をセレクトできる)
- 2 回目の PCR 反応 (Nested PCR: さらに内側のベクター配列と既知の外来 DNA 配列で実施)
- 3 回目の PCR 反応 (アダプター配列+インデックス配列)
- 目的の増幅産物をゲル精製し、シーケンス解析 (Illumina Miseq) を行う

3.3.3. NGS データの遺伝子組換え植物の安全性評価における利用性

既に海外の評価機関等の評価に利用されている、もしくは文献的な利用例に基づき、我が国の遺伝子組換え植物（GMP）の安全性評価の分野において NGS データが使用される項目について、NGS を用いる解析のうち全ゲノムシーケンシング（WGS）、ターゲットキャプチャシーケンシング（TCS）及びロングリード（Long Read）のアプローチによる利用性を表 3.3.3-1 にまとめた。なお、実施できる段階を以下に示す 3 つの段階に分けて整理した。

検出レベル 1：従来法（サザンブロット及び PCR+サンガー法）と同等レベルの情報取得

検出レベル 2：検出レベル 1+ゲノム全体の配列情報の取得

検出レベル 3：・検出レベル 2 + 詳細なゲノム全体の配列情報の取得 または
・これまでにない新たな評価として GMP の機能性評価

(1) 検出レベル 1：従来法（サザンブロット及び PCR+サンガー法）と同等レベルの知見を得る

表 3.3.3-1 に示すように、WGS 法、TCS 法及び Long Read のいずれの方法でも、T-DNA の挿入部位近傍を解析することで、従来法と同等レベルで 1) T-DNA の挿入の有無、2) T-DNA の配列の確認、3) T-DNA の挿入コピー数の確認及び 4) 世代間におけるジャンクション配列の一貫性に基づく経世代における T-DNA の遺伝的安定性の評価、の知見が得られる。ただし、1)及び 2)については、NGS のいずれの方法でも、ある一定以上のシーケンス深度を必要とする点に留意が必要である。また、TCS 法は、プローブや PCR プライマーにより特定の DNA 領域のみを解析する手法であるため、2) T-DNA 配列の確認の際には、プローブや PCR プライマーのデザインによっては T-DNA 領域全体のシーケンスを取得できない場合がある。TCS 法の場合、3)の挿入コピー数の確認では、ホストゲノムのシーケンス深度と T-DNA のシーケンス深度が同等であること、また、ホストゲノムの配列がゲノム上の単一配列由来であることに基づきコピー数が 1 コピーであることを確認するため、WGS 法に比べコピー数確認の正確性が劣る場合がある。さらに、TCS 法では 4) 経世代における T-DNA の遺伝的安定性評価では、1)～3)を実施する必要があるため、2)及び 3)に付随する TCS 手法の制約への留意が必要である。

ここで、WGS、TCS あるいは Long Read による NGS 解析結果については原則としてバリデーションを行う必要はないと考えられる。しかし、NGS データのシーケンス深度が低い場合や明確なデータが得られなかった場合等にサンガー法や PCR 法などの対応が必要と考えられる。

(2) 検出レベル 2：検出レベル 1 + ゲノム全体の配列情報の知見を得る

WGS 法あるいは Long Read を用いた場合には、シーケンス深度が低いゲノム領域のデータを許容するケースがあるものの、検出レベル 1 で取得する情報に加え、5) バックボーン配列の有無の確認が可能である（表 3.3.3-1）。また、ゲノム全体についてバックボーン配列の有無確認ができることにより、検出レベル 1 の 3) T-DNA の挿入コピー数に関するデータの精度の向上（T-DNA 挿入の False negative の軽減）が図れる（表 3.3.3-1）。なお、TCS 法については、ゲノム全体の DNA 配列データが得られないため、バックボーン配列に関する知見を得ることはできず、検出レベル 1 で取得する 3) T-DNA の挿入コピー数の情報の精度を向上させることはできない。

(3) 検出レベル 3

① 検出レベル 2 + 詳細なゲノム全体の配列情報の知見を得る

WGS 法あるいは Long Read を用いる場合、ゲノム全体を解析できることから、6) T-DNA のホストゲノム上の挿入位置の同定、及び 7) インサートマップの作成が可能である（表 3.3.3-1）。これらは、従来法では得ることが出来なかった詳細かつ網羅的なゲノム DNA 及び T-DNA に関わる情報である。ただし、ゲノム全体である一定程度のシーケンス深度が必要である点及び解析結果が解析に使用するリファレンスゲノムの精度に依存する点に留意が必要である。なお、TCS 法はゲノム全体を解析できないため、6)及び 7)の解析には適用できない。

② GMP の機能性評価（これまででない評価）

NGS が登場したことにより、RNA を対象とした網羅的な遺伝子/miRNA/siRNA 発現量解析が可能となり、8) T-DNA 導入による影響/効果（機能獲得型（Gain of function））の評価及び 9) 内在遺伝子のノックダウン（RNAi）による影響/効果（機能喪失型（loss of function））の評価が可能となった（表 3.3.3-1）。RNA を対象とした解析については、ショートリードの場合、深いシーケンス深度での解析が行える点で利点があるが、スプライスバリエーションの検出に弱い。一方、ロングリードの場合、シーケンス深度は浅くなるが、スプライスバリエーションの検出に強い利点がある。

これらの評価は GMP の機能に関する新たな知見を与える可能性があり、今後 GMP の新たな評価項目として重要になる可能性がある。ただし、RNA コピー数の定量評価を行うためには非常に多くのシーケンス深度が必要（1～100,000 倍）であり、解析結果が使用するリファレンスゲノムの精度に依存することの留意が必要である。

表 3.3.3-1 NGS 解析法の遺伝子組換え植物の安全性評価における利用性

検出レベル	評価項目	解析対象	必要な NGS 解析の条件	WGS 法	TCS 法	Long Read
検出レベル 1 : 従来法 (サザン ブロット及び PCR+サンガー 法) と同等	1) T-DNA の挿入の有無 (接合配列の確認)	・ T-DNA の挿入部位 近傍 (例: 10Kb)	・ ある一定以上のシーケンス深度が必要 (例 : 100 以上)	○	○	○
	2) T-DNA 配列の確認 (T-DNA 構造が維持されているか?)			○	△*1	○
	3) T-DNA の挿入コピー数の確認	・ 接合配列部分の T-DNA とホストゲノム DNA のシーケンス深度 がほぼ同じレベルことを確認する		○	△*2	○
	4) 継世代における T-DNA の遺伝的安定性の評価 (世代間におけるジャンクション配列の一貫性)	・ 1)～3)の項目について F2、F3...と世代間で実施する		○	×～○	○
検出レベル 2 : 検出レベル 1 + ゲノム全体の配 列情報	3') T-DNA の挿入コピー数の確認	・ 検出レベル 1 に 5)のデータを加えることで、挿入コピー数のデ ータ精度が向上する		○	△*3	○
	5) バックボーン配列の有無の確認	・ ゲノム全体	・ シーケンス深度が低いゲノム領域のデー タを許容するケースがある	○	×	○
検出レベル 3 : 検出レベル 2 + 詳細なゲノム全 体の配列情報	6) T-DNA のホストゲノム上での挿入位置の同定	・ ゲノム全体	・ ゲノム全体である一定以上のシーケンス 深度が必要 ・ 正確なリファレンスゲノム情報 (ドラフ トゲノムの精度に結果が依存する)	○	×	○
	7) インサートマップの作成			○	×	○
検出レベル 3 : GMP の機能性 評価 (これまで にない評価)	8) T-DNA 導入による影響/効果の評価 = Gain of function =	・ RNA	・ RNA コピー数の定量評価を行うため非 常に多くのシーケンス深度が必要 (1～ 100,000 倍) ・ 正確なリファレンスゲノム情報 (ドラフ トゲノムの精度に結果が依存する)	Short Read の場合: シーケンス深度が 深い スプライスバリア ントの検出に弱い		シーケンス深 度が浅 い スプライ スバリア ントの検出 に強い
	9) 内在遺伝子のノックダウン (RNAi) による影 響/効果の評価 = Loss of function =					

WGS：全ゲノムシーケンシング、TCS：ターゲットキャプチャシーケンシング、

○：適用できる、×：適用できない

*¹ TCS 法の場合、プローブや PCR プライマーのデザインによっては、T-DNA 領域全体のシーケンスを取得できない場合がある。

*² TCS 法では、ゲノム全体の DNA 配列データが得られないため、ホストゲノムのシーケンス深度と T-DNA のシーケンス深度が同等であること、また、ホストゲノムの配列がゲノム上の単一配列由来であることに基づきコピー数が 1 コピーであることを確認するため、WGS 法に比べコピー数確認の正確性が劣る場合がある。

*³ TCS 法では、ゲノム全体の DNA 配列データが得られないため、挿入コピー数に関するデータ精度の向上は期待できない。

3.4. NGS データを遺伝子組換え植物の安全性評価に用いる際の課題の明確化

これまで調査した評価機関等での評価や文献情報をもとに、遺伝子組換え植物（GMP）の安全性評価に NGS データを用いる際の課題について、以下の 3 つの課題が考えられた。

- 課題 1：GMP の安全性評価に NGS データを活用する際のデータの品質や信頼性等を見極める明確な基準がないことへの対応
- 課題 2：植物のゲノム構造の特殊性（同質倍数体、異質倍数体、繰り返し配列の多さ）への対応
- 課題 3：多サンプル解析を要する場合のコスト（費用及び時間）への対応

各課題における具体的な項目、内容及び課題解決のヒントを表 3.4-1 に示す。また、課題解決に向けたヒントになると考えられた事例を別添資料 3-5 に示した。

3.4.1. 課題 1: 遺伝子組換え植物の安全性評価に NGS データを活用する際のデータの品質や信頼性等を見極める基準がないことへの対応

NGS データを GMP の安全性評価に用いる際の課題として、NGS データの品質や信頼性等を見極める基準がないことが挙げられる。具体的には、①NGS データの検出限界の程度（分子キャラクタリゼーションにどの程度の NGS データ量があれば、評価できるのか）、②植物毎に大幅に異なるゲノムサイズへの対応（植物の種類によってゲノムサイズには数百～17,000 Mb 程度の違いがあり、必要となる NGS データ量が異なる）及び③野生型ゲノムが混入した場合の検出感度が挙げられる（表 3.4-1）。これら①～③項目については、GMP において NGS データを用いる際に必要とされるデータ量/シーケンス深度の検証が行われた事例（別添資料 3-5 の事例-1/ 文献 ID-003）や、異なる NGS 測定装置を組み合わせることなどで欠点を補い、データの精度を上げることなどが解決に向けたヒントになると考えられた。なお、異なる NGS 測定装置の組合せは、単一機器によるシーケンス深度を確保するよりも有効である可能性がある。

NGS データから必要な知見を得るためにはバイオインフォマティクス等によるデータ解析が必要であるが、NGS 解析ソフトウェアやアルゴリズムにはバリエーションがあり、GMP の安全性評価に適したソフト/アルゴリズムかの判断が必要であり、④NGS 解析ソフトウェアの選定も重要な項目となる（表 3.4-1）。GMP ではアルゴリズムの違いによるデータ解析への影響に関する知見は本調査では得られていないが、NGS によるヒト疾患のための解析において異なるソフトウェア（32 種）を用いた場合のデータ解析結果への影響について検証した論文があり（別添資料 3-5 の事例-2/ 文献 ID-018）が課題解決に向けて参考になると考えられる。

また、NGS 解析による繰り返し測定の精度や GMP の個体間差が NGS データに基づく解析結

果に与える影響が考えられるため、どの程度のサンプルを解析すればよいかという⑤解析すべきサンプル数も具体的な課題として挙げられる（表 3.4-1）。しかし、現時点では本課題に関する具体的な知見は得られていない。

GMP の外来遺伝子を特定する様々な方法が論文発表されているが、今後はそれらの方法の有用性や適用範囲などを明確化し、⑥解析方法の標準化/調和が必要になると考えられる（表 3.4-1）。これまでに GMP の申請に使用されている NGS 解析方法としては、分子キャラクタリゼーションの解析に特化した Junction Sequence Analysis (JSA) 法や Southern-by-Sequencing (SBS) 法などが開発されている（本報告書 3.3.2 項の図 3.3.2-2、図 3.3.2-3 及び図 3.3.2-5 を参照）。また、JSA は既に申請に使用されている。それぞれの方法に利点と欠点を整理し、GMP の安全性評価に適した解析方法を標準化していく必要があると考えられる。

3.4.2. 課題 2：植物のゲノム構造の特殊性（同質倍数体、異質倍数体、繰り返し配列の多さ）への対応

遺伝子組換えのホストとなる植物は、ゲノムの組換え反応が起こりやすいために同質倍数体や異質倍数体と染色体構造が非常に複雑である上、ゲノムに繰り返し配列が多いなど、ゲノム配列の解読には困難さがある。現在の NGS 解析の主流は、ショートリードタイプ（100～300 bp のリード長）のデータが多いが、リード長が短い測定は、繰り返し配列の検出に弱く、複雑な構造を持つ植物ゲノム全体の配列決定には不向きな面がある。また、現在主流のショートリードタイプの NGS の検出原理は「DNA 合成・光検出法を用いた超並列シーケンシング法」をとっており、PCR 反応を必要とするため、GC リッチな配列の検出が弱い。そのため、① これまで実施されてきた NGS 解析による植物ゲノムへの外来 DNA の挿入の評価の正確性について検討する必要があると考えられた（表 3.4-1）。課題解決のヒントとしては、近年開発が進んでいる、ロングリードタイプ（数 Kb 以上）のシーケンサーの活用がある。ロングリードタイプは、繰り返し配列に強い利点があり、さらに PCR 反応が不要であることから、GC リッチな配列の検出も行える。

ロングリードタイプのシーケンサーとしては、1 分子リアルタイムシーケンシング法を採用した PacBio（パシフィック・バイオ・サイエンシズ社）がある。PacBio は、1 分子単位で DNA ポリメラーゼを使って DNA 合成し、そのときに発せられる蛍光色素のシグナルをリアルタイムに検出することでシーケンス解析ができる（別添資料 3-4）。PCR 増幅が不要でロングリード（20Kb 以上）かつ低いエラー率で、GC 含有量に関わらず幅広いゲノム領域をカバーできる特性を持っている（PacBio の測定原理については別添資料 3-4 を参照）。このように非常に高いスペックを有する機器ではあるが、高額であるためまだ一般にはそれほど広まっていない。

ロングシーケンスの別の手法としては、ナノポアシーケンシング法を採用した MinION があり、手のひらサイズと非常にコンパクトで、他の NGS 機器に比べると格段に安価（数十

万円)であることも大きな特徴である。測定原理としては、膜たんぱく質中に形成させたナノサイズの穴(ナノポア)に一本鎖 DNA を通し、下方へ移動させる中で生じるイオン電流の変化で塩基配列を読み出す方法を取っている(MinION の測定原理については別添資料 3-4 を参照)。ただし、現時点では、イオン電流の違いで塩基(A, C, G, T)の種類を同定するため、エラー率が高く、塩基の読み取り精度が低い。

植物のゲノム構造の複雑さのために、②リファレンスゲノムとして登録されているゲノム配列はドラフト段階のものが多く、信頼性が乏しい場合があることが次の課題として挙げられる(表 3.4-1)。この課題解決のヒントとしては、異質倍数体のカラシナ(*Brassica juncea*)やナタネゲノム(*Brassica napus*)について、上述のロングリードタイプとショートリードタイプの NGS データを組み合わせるシーケンスデータを取得し、さらに長鎖 DNA フラグメント解析ツール(例: BioNano 社 Irys System)を用いて、複雑なゲノム構造についてシーケンスギャップの位置を明確にし、コンティグをつなげてさらに長いシーケンス配列を生成する効果的な scaffolding を行った研究例があった(添付資料 3-5 の実例-3)。この研究ではショートリードタイプの NGS で深いシーケンス深度を確保し、ロングリードタイプの NGS で繰り返し配列や GC リッチな配列の解読を行い、かつ長いコンティグを作成した。さらに長鎖 DNA フラグメント解析ツールでコンティグを繋ぎ合わせる scaffolding を行い、より長いシーケンスデータを作成した。この方法により、非常に複雑な異質倍数体のゲノム配列を精度高く解読でき、さらに、同属の類似したゲノム配列についても、各品種の違いに応じて、ゲノム配列の違いを見出すことができていた。

また、③新たなゲノム編集技術: CRISPR-Cas9 等への対応が課題として挙げられる(表 3.4-1)。例えば、CRISPR-Cas9 法はターゲットとするゲノム DNA に相補的なガイド RNA をデザインし、Cas9 の働きで DNA 二本鎖を切断してゲノム配列の任意の場所を削除、置換、挿入することができる新しい遺伝子改変技術である。この CRISPR-Cas9 法は原理的にオフターゲット変異(ターゲットのゲノム DNA 以外で変異が生じる)が生じる可能性があるが、ヒト、マウス、ショウジョウバエ等ではガイド RNA 配列のオフターゲットを予測できるソフトウェア(解析ツール)がある。しかし、植物ゲノムに対しては、同様のツール開発はまだ行われていない。また、上述のとおり、植物ゲノムは非常に複雑であるために、リファレンスゲノムの DNA 配列決定の精度が低い部分が存在するために、オフターゲット変異と自然発生的な変異との区別が困難になることが予想される。そこで、CRISPR-Cas9 のようなゲノム編集技術で作成された GMP を評価する場合は、これらの問題点を考慮する必要がある。

3.4.3. 課題 3 : 多サンプル解析を要する場合のコスト(費用及び時間)への対応

今後、GMP の外来遺伝子の分子キャラクタリゼーションに NGS 解析がますます適用されるようになると予想されるが、全ゲノムシーケンシング(WGS)を行うには 1 サンプル当たりの費用及び時間がかかり、課題 1 の⑤で示した解析すべきサンプル数が多くなる場合

には①NGS 解析のハイスループット化が課題として挙げられる。課題解決のヒントとして、Southern-by-Sequencing (SBS) 法、Microarray hybrid capture 及び next-generation sequencing 法並びに HtStuf 法等(原理は別添資料 3-4 を参照)のターゲットキャプチャシーケンス(TCS)法のような多検体に対応できる解析方法の開発が考えられる。ただし、TCS 法は効率的に GMP の NGS データを取得できる一方で、ゲノム全体の配列情報は得られないため、分子キャラクター化をする上で T-DNA の配列確認の際にはプローブや PCR プライマーのデザインによって T-DNA 領域全体のシーケンスを取得できない場合があることや、T-DNA の挿入コピー数の確認においてシーケンス深度からコピー数を確認することで WGS に比べ正確度が劣る場合があること等について、評価の際に留意が必要となる(3.3.3 項の表 3.3.3-1 参照)。

表 3.4-1 遺伝子組換え植物（GMP）の安全性評価に NGS データを用いる際の課題

課題 1： GMP の安全性評価に NGS データを活用する際のデータの品質や信頼性等を見極める明確な基準がないことへの対応		
課題（小項目）	内容	課題解決のヒント
① NGS データの検出限界の程度	・ GMP の安全性評価の各項目（分子キャラクタリゼーション）に、どの程度の NGS データ量があれば、評価できるのか	■ GMP の NGS 解析において必要とされるデータ量/シーケンス深度の検証〔事例-1/ 文献 ID-003〕 ・ T-DNA を NGS データから同定するために、①Detection、②Proof、③Identification の 3 段階の検出レベルに分けて、それぞれに必要なリード数を検討 ・ 7 種類の GMP について検討 ・ 野生型と GMP の混合サンプル（GMP が 0.01%、1%、50%混合比率）を用いた検討
② 大幅に異なるゲノムサイズへの対応	・ 植物は種類によってゲノムサイズが大幅に異なるため（数百～17,000 Mb）、必要な NGS データ量をどのように判断するか	
③ 野生型ゲノムが混入した場合の検出感度	・ GMP が一部混在したサンプルの場合、どの程度の NGS データ量があれば、評価できるのか	■ 異なる NGS 測定装置を組み合わせることで、欠点/利点を補い、データの精度を上げる（単一機器でシーケンス深度を確保するよりも有効な可能性あり）
④ NGS 解析ソフトウェアの選定	・ NGS 解析ソフトウェアやアルゴリズムにはバリエーションがあり、どれが GMP 安全性評価に適しているのか	■ アルゴリズムの違いによるデータ解析結果の違いに関する検証〔事例-2/ 文献 ID-018〕 ・ ヒト疾患の NGS データを用い、32 種の変異解析ツールを検討 ・ GMP ではアルゴリズムの違いがデータ解析結果に与える影響については知見なし
⑤ 解析すべきサンプル数	・ NGS 解析の繰り返し測定に対する精度	・ 同一の GMP で繰り返し測定を行った報告はない

課題 1： GMP の安全性評価に NGS データを活用する際のデータの品質や信頼性等を見極める明確な基準がないことへの対応		
課題 (小項目)	内容	課題解決のヒント
	・ GMP の個体間差が NGS データに与える影響 ・ GMP のゲノム配列は部位別 (種子、葉、茎、根など)で同じか	・ 個体間差を小さくするために複数サンプルを用いている例が多い ・ 同一 GMP の部位別の NGS 解析を実施した研究で同様の結果が得られているが事例が少ない
⑥ 解析方法の標準化/調和	・ 様々な方法が論文化されており、メーカーが申請に使用する方法が複数ある	■ 申請に利用されうる方法からの検討 ・ Junction Sequence Analysis (JSA) 法 ・ Southern-by-Sequencing (SBS) 法 (3.3.2 項の図 3.3.2-2、図 3.3.2-3 及び図 3.3.2-5 を参照)

課題 2：植物のゲノム構造の特殊性（同質倍数体、異質倍数体、繰り返し配列の多さ）への対応		
課題 (小項目)	内容	課題解決のヒント
① ショートリードタイプは、繰り返し配列や GC リッチな配列の検出が弱い	・ 現在の NGS 解析はショートリードタイプのデータが多く、これまでに実施してきた評価における植物ゲノムへの外来 DNA の挿入確認は正確か	■ ロングリードタイプ (数 Kb 以上) のシーケンサーの活用 ・ PacBio (1 分子リアルタイムシーケンシング) ・ MinION (ナノポアシーケンシング) (原理は別添資料 3-4 を参照)
② 植物ゲノムのリファレンスゲノムの精度	・ リファレンスゲノムとして登録されている配列はドラフト段階のものが多く、ゲノム配列として信頼性が乏しい場合がある (T-DNA の挿入を評価できる精度があるか)	■ より精度の高いロングリード解析を用いた植物ゲノムの解読 [実例-4] ・ PacBio RS によるカラシナ/ ナタネ (異質倍数体) ゲノムの解読 →さらにショートリードタイプのデータをプラスしてシーケンス深度を上げ、フラグメント解析によってシーケンスギャップを埋めることで、品種の違いによるゲ

課題2：植物のゲノム構造の特殊性（同質倍数体、異質倍数体、繰り返し配列の多さ）への対応		
課題（小項目）	内容	課題解決のヒント
		ノム配列の違いも解析が可能な
③ 新たなゲノム編集技術: CRISPR-Cas9 等への対応	・ CRISPR-Cas9 法等によってオフターゲット変異が生じる可能性がある ・ さらに、植物は信頼性が乏しいドラフト段階のゲノムが多いため、オフターゲット変異と自然発生的な変異との区別が困難になることが予想される	■ CRISPR-Cas9 法のオフターゲットを予測できるソフトウェア（解析ツール）がある ・ ヒト、ショウジョウバエ等では CRISPOR、CCTop 等のソフトウェアがある。 ・ GMP では具体的な知見なし

課題3：多サンプル（n 数の確保、経世代、部位別等）がある場合には費用も時間もかかる		
課題（小項目）	内容	課題解決のヒント
① NGS 解析のハイスループット化	・ WGS 法では 1 サンプルあたりのデータ取得に費用も時間もかかる ・ TCS 法によるデータの場合、安全性評価目的に沿った方法であるのか判断する必要がある。	■ ターゲットキャプチャシーケンス（TCS）法など、GMP の安全性評価に特化した、多検体に対応できる解析方法の開発 ・ Southern-by-Sequencing (SBS) 法 ・ Microarray hybrid capture 及び next-generation sequencing 法 ・ HtStuf 法 (原理は別添資料 3-4 を参照) ■ TCS 法の特性の把握（表 3.3.3-1 参照）

3.5. 今後期待される対応

GMP の安全性評価における NGS データの利用に向けて今後期待される対応として以下が挙げられる。

■ GMP 安全性評価における NGS データ利用に関するガイダンスの作成及び運用

ガイダンスにおいて明確にしておくべき事項として以下が考えられる。

1) GMP 安全性評価のために必要な対象サンプルについて

評価に必要なサンプルセットとして

- ・ 解析対象とする部位（種子、葉、茎、根等）
- ・ 1 サンプルあたりの n 数

2) GMP 安全性評価が可能と判断できる NGS 解析データの（定量的な）基準について

- ・ NGS 解析に供する gDNA の品質基準（分解度、純度、濃度等）
- ・ NGS の生データの品質基準（Q30 のパーセンテージ等）
- ・ データ解析手法及びアルゴリズムの適切性
- ・ リファレンスゲノムのデータソース及びバージョン等
- ・ NGS 解析結果の品質基準（マッピングされたリード数、シーケンス深度、カバレッジ等）

これらについては、表 3.4-1 に示した課題への対応が必要であり、対象とする植物種、使用された NGS 機種、測定条件、解析パイプライン等によって基準が変わる可能性がある。

3) NGS データのバリデーションの必要性に関する判断基準について

- ・ どのような NGS 結果の場合にどのようなバリデーションが必要か？

例えば、十分なシーケンス深度のある Junction Sequence Analysis (JSA) による解析が行われた場合、サンガー法によるバリデーションは不要との意見もあるため、NGS データに基づく結果に応じて、どのようなバリデーションが必要であるかの判断基準が必要と考えられる。

4) 解析結果以外に提出が必要な情報の範囲及び項目について

- ・ NGS 解析条件（例：採用した NGS 解析法、ライブラリの作製方法、使用した機種、シーケンス条件、データ解析のアルゴリズム、リファレンスゲノムの情報等）
- ・ NGS データの精度確認ができる情報（gDNA 品質データ、NGS データの Q30 のパーセンテージ等）

また、客観的評価の観点及び知財の観点から生データの提出の必要性について検討する必要があると考えられ、例えば T-DNA 挿入近傍の○Kb の配列のみ等を求めるという考え方もある。

なお、NGS データの提出時に必要な情報として、JRC が 2016 年 2 月に公表したガイドラインが参考になると考えられる（文献 ID-133）。

■ GMP に関する NGS データの蓄積/NGS 研究の促進

- 1) GMP に最適化された NGS 解析手法のさらなる開発
- 2) リファレンスゲノムの情報更新（より信頼度及びカバー率の高いゲノム情報へ）

NGS 関連用語集

固有名詞あるいは現時点で適切な和訳がない場合には「用語（和訳）」部分には英語表記をそのまま掲載している。

(アルファベット順)

用語 (英語)	用語 (和訳)	説明
ABYSS (Assembly by Short Sequences)	ABYSS	ショートリード用に設計された <i>de novo</i> でペアエンドシーケンスをアセンブルするソフトウェア (http://www.bcgsc.ca/platform/bioinfo/software/abyss)。
adaptor	アダプター	(シーケンスライブラリの) DNA フラグメントに 5'もしくは 3'端に結合させた配列既知のオリゴヌクレオチド (https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf)。
AgriGO	AgriGO	遺伝子オントロジー (GO) 解析に使用するウェブベースのツールであり、データベース (http://bioinfo.cau.edu.cn/agriGO/)。
alignment	アライメント	シーケンスしたリードとリファレンスゲノムの配列を一致させること。
alignment reads	アライメントリード	アライメントされたリード。
analysis pipeline	解析パイプライン	(NGS の)データ解析に必要なツールもしくはアルゴリズムを解析手順に従いフロー形式で定型化したもの。もしくはそれをシステム内でプログラム化したもの。
assemble	アセンブル	シーケンスにより取得した配列同士の相同性を比較してつなぎ合わせて、より長い配列にすること。
BAM (file)	BAM (ファイル)	マッピング結果のファイル形式のひとつで、SAM ファイルを 2 進数変換したもの。SAM ファイルは非常に容量が大きいファイルとなるため、2 進数化し、ファイルサイズの圧縮を行っている。

用語 (英語)	用語 (和訳)	説明
base-call	ベースコール	測定装置(シーケンサー)によって得られた測定値を塩基配列に変換すること。 (base-calling の場合は、そのプロセスを指す)。
BEDtools	BEDtools	NGS データ解析用ツールのひとつ (http://bedtools.readthedocs.io/en/latest/)。
BLAST	BLAST	Basic Local Alignment Search Tool の略。 クエリとなるヌクレオチドまたはタンパク質配列を配列データベースと比較し、統計的有意性を計算することで生物学的配列間の類似性の領域を見出すプログラム(https://blast.ncbi.nlm.nih.gov/Blast.cgi)。
BlastAll	BlastAll	BLAST 中のアルゴリズムのひとつ (http://nebc.nox.ac.uk/bioinformatics/docs/blastall.html)。
Blastn	Blastn	BLAST 中のアルゴリズムのひとつ。
Blastx	Blastx	BLAST 中のアルゴリズムのひとつ。
Bowtie	Bowtie	NGS 解析から得られたリードをリファレンスゲノムにマッピング・アライメントを行うソフトウェアのひとつ。
BWA (Burrows-Wheeler Aligner)	BWA	Bowtie と同じく、マッピング・アライメントを行うソフトウェアのひとつ。
CLC Genomics Workbench	CLC Genomics Workbench	NGS データ解析のソフトウェア (有償) のひとつで、ゲノムアセンブル・マッピング/アライメント、解析結果の可視化等、一連のデータ解析に対応可能。 (https://www.qiagenbioinformatics.com/products/clc-genomics-workbench/)
clean reads	clean reads	NGS 解析で得られた生データからノイズとなる塩基やリード等を取り除いた後のリード。

用語 (英語)	用語 (和訳)	説明
contig	コンティグ	contiguous sequence の略。シーケンスしたリード同士をつなぎ合わせた長い配列のことをさす。
coverage	カバレッジ	リファレンスゲノムにアライメントされる（カバーしている）リードの数の平均値。例）ヒト [ゲノムサイズ: 3Gb] サンプルを 90Gb シーケンスした場合、coverage は 30 倍。
coverage rate	カバレッジ比	カバレッジした程度 (read depth of coverage、read depth と同義)。
CRAM	CARM	NGS データのマッピング・アライメントを行うソフトウェアのひとつ。もしくは CARM ソフトウェアで作成されたマッピング・アライメントファイルの名称 (http://www.ebi.ac.uk/ena/software/cram-toolkit)。
Cufflinks	Cufflinks	トランスクリプトーム解析 (RNA-seq 解析) に用いられるソフトウェアのひとつ。転写産物のアセンブリを行い、発現量を推定し、異なるサンプル間の比較解析等を行う (http://cole-trapnell-lab.github.io/cufflinks/)。
Data cleaning	データクリーニング	NGS 解析で得られた生データからノイズとなる塩基やリード等を取り除くこと。
de novo assembly	De novo 解析	シーケンスした塩基配列を基に、ゲノムの塩基配列の推定を行うまたは新たなゲノム配列を決定すること。
deep sequence	ディープシーケンス	カバレッジ/シーケンス深度が多くなるような条件でシーケンスを行うこと。
depth	シーケンス深度	同じ箇所を何回シーケンスしたかを表す尺度。冗長度とも言われる。「20×」であれば 20 回同じ箇所をシーケンスしたことを意味する。
edgeR	edgeR	トランスクリプトーム解析 (RNA-seq 解析) に用いられるソフトウェアのひとつ。異なるサンプル間の比較解析等を行う (https://bioconductor.org/packages/release/bioc/html/edgeR.html)。
Enrich	エンリッチ	ターゲット配列を濃縮すること。

用語 (英語)	用語 (和訳)	説明
Expressed Sequence Tag (EST)	EST	生体中で発現している遺伝子のうち一部分の配列。
FASTQ file	FASTQ ファイル	NGS データのフォーマットのひとつで、シーケンスした塩基配列と、各塩基のシーケンスクオリティが記載されている。
featureCounts	featureCounts	遺伝子、エクソン、プロモーター、染色体の位置などのゲノム特性にマッピングされたリードをカウントする、リードの要約プログラム。RNA-seq 及びゲノム DNA-seq のリードの両方のカウントに利用できる (http://bioinf.wehi.edu.au/featureCounts/)。
FPKM (fragments per kilobase of exon per million fragments mapped)	FPKM	RNA-seq で得られたシーケンスデータから各遺伝子の発現量を算出する際に行う正規化する方法。ペアエンドリードの時に用いられる方法。
Genome Analyzer IIX (GAIIx)	Genome Analyzer IIX (GAIIx)	Illumina 社が提供するショートリード型の NGS 機器のひとつ。現在は取り扱い終了となっている。
Geneious	Geneious	NGS データ解析ソフトウェアのひとつで、それ以外にもクローニングのシミュレーション、マイクロサテライト解析、16S rRNA を使用した生物多様性解析などが実施可能(http://www.geneious.com/)。
GS FLX+	GS FLX+	Roche 社が提供するの中程度のリード長が測定できる NGS 機器のひとつ。現在は取り扱い終了となっている。
GS Junior	GS Junior	GS FLX+の下位機種。現在は取り扱い終了となっている。
High-Throughput Sequencing (HtStuf)	High-Throughput Sequencing (HtStuf)	長鎖の導入遺伝子や短鎖のマルチコピー配列の解析に有用な、改変 TA クローニングによるライブラリ調製法とイルミナシーケンスを組み合わせた手法 (Kanizay LB et al. Plant Genome. 2015;8(1))。

用語 (英語)	用語 (和訳)	説明
HiSeq	HiSeq	現在のイルミナ社の高スループットな NGS (Short read 型)。HiSeq2000、2500、3000、4000 など、型番が大きくなるに従い処理量が増加する。
IGV	IGV	Integrative Genomics Viewer の略。大規模な統合ゲノムデータセットの探索用の高性能可視化ツール。マイクロアレイ及び NGS のデータ等に適用できる (http://software.broadinstitute.org/software/igv/)。
index sequence	インデックス配列	シーケンスライブラリ調製の際にライブラリに含まれる DNA フラグメントにライゲーショニングされる配列。バーコード (barcode) またはタグ (tag) とも言われる。
Ion PGM	Ion PGM	ThermoFisher Scientific 社が提供するショートリード型の NGS 機器 (Ion Torrent シリーズ) のひとつ。
Ion Torrent	Ion Torrent	ThermoFisher Scientific 社が提供するショートリード型の一連の NGS 機器。
JELLYFISH	JELLYFISH	DNA 中の k-mers を高速でメモリ効率よくカウントするためのツール (http://www.cbcb.umd.edu/software/jellyfish/)。
Junction Sequence Analysis (JSA)	ジャンクション配列解析 (JSA)	T-DNA とホストゲノムの接合部分の配列を解析することで分子キャラクタリゼーションに関わる情報を得る方法のひとつ (Kovalic et al. Plant Genome. 2012; 5(3):149-163)。
K-mer analysis	K-mer 解析	k-mer とは長さ k の部分配列であり、配列の出現を数えることは、DNA 配列の多くの分析の中心的なステップ。
Lasergene Genomic Suite	Lasergene Genomic Suite	NGS データ解析のソフトウェア (有償) のひとつで、ゲノムアセンブル・マッピング/アライメント、解析結果の可視化等、一連のデータ解析に対応可能 (https://www.dnastar.com/t-products-dnastar-lasergene-genomics.aspx)。

用語 (英語)	用語 (和訳)	説明
library	ライブラリ	NGS 解析のためのサンプル調製により、各断片の末端にアダプター等が結合したゲノム断片の集合体。
long read	ロングリード	通常、1 リードの長さが数 K bp 以上の場合を言う。
MACS2	MACS2	MACS2 (Model-based Analysis of ChIP-Seq) アルゴリズムを使用して、ChIP-及び DNase-seq データの統計的に有意にエンリッチされたゲノム領域を同定するツール (http://chipster.csc.fi/manual/macs2.html)。 ChIP-Seq: クロマチン免疫沈降シーケンス
mapping/alignment	マッピング/アライメント	既知の参照ゲノムに対してシーケンスリードをマップすること。
massively parallel sequencing	超並列シーケンシング	1 回の測定で塩基配列データを大量に取得すること。
mate pair read	メイトペアリード	長い塩基配列の両側を読むため、塩基配列を環状化し、両端を含む接合部分のみを抽出して解読する方法。
MinION	MinION	Oxford Nanopore Technologies 社が提供するロングリード型の NGS 機器。
MiSeq	MiSeq	Illumina 社が提供するショートリード型の NGS 機器のひとつ。
Newbler	Newbler	<i>De novo</i> DNA シーケンスアセンブリ用のソフトウェア。Roche 社の NGS 機器である 454 により作成されたシーケンスデータのアセンブリ用である (http://www.endophyte.uky.edu/ngs/Newbler)。
next generation sequencing (NGS)	次世代シーケンシング (NGS)	超並列シーケンシングとも呼ばれ、既存の数多くのシーケンス技術を包括的に表現する用語。
NextSeq	NextSeq	現在のイルミナ社が販売する中スループットな NGS (Short read 型)。

用語 (英語)	用語 (和訳)	説明
PacBio RS II	PacBio RS II	Pacific Biosciences 社の主力 NGS (Long read 型)。
paired-end read	ペアエンドリード	断片化した塩基配列の両端を解読する方法。
Peak call	ピークコール	測定装置(シーケンサー)によって得られた測定値を塩基配列に変換すること。Base call と同義。
PERL	PERL	UNIX や Linux を操作する際に使用するスクリプト言語。
Phred score	Phred スコア	DNA 塩基配列のクロマトグラムファイルを読み取り、塩基を分析するベースコールプログラム Phred が割り当てる品質スコア (http://www.phrap.com/phred/)。
PRINSEQ	PRINSEQ	ゲノム及びメタゲノムデータのフィルタリング、再フォーマット、またはトリミングに使用できるツール(http://prinseq.sourceforge.net/)。
Q score	Q スコア	シーケンスクオリティスコアのこと。シーケンスクオリティスコアとはベースコールが正しく行われない確率の測定値を指す (例: Q30 ではベースコールが正しく行われない確率が 1000 分の 1)。
raw read	raw read	NGS 解析で得られた生データ (リード)。
Ray	Ray	NGS データの de novo 解析をするソフトウェア (http://denovoassembler.sourceforge.net/)。
read(s)	リード	NGS 解析で得られる A、T、C 及び G 塩基配列が連なった一本の配列のこと。
Reference genome	リファレンスゲノム	既知のゲノム配列。
re-sequencing	re-sequencing	既にリファレンスゲノムがあるサンプルについてシーケンス解析を行うこと。
RNA sequencing (RNA-seq)	RNA-seq	RNA サンプルを NGS 解析し、RNA 配列と発現量の両方を解析すること。
RPKM (Reads Per Kilobase of exon Model per million mapped reads)	RPKM	RNA-seq 解析に用いるアルゴリズムのひとつで、得られたシーケンスデータから各遺伝子の発現量を算出する際に用いる。

用語 (英語)	用語 (和訳)	説明
SAM	SAM ファイル	Sequence Alignment/Map Format の略でファイル形式の名称のひとつ。リファレンスゲノム配列にシーケンス配列を Mapping した結果が保存されたファイル。
SAMtools	SAMtools	NGS データの相互解析に用いられるプログラム(http://www.htslib.org/)。
Scaffolding sequence	スキャホールド配列	コンティグ配列同士をつなぎ合わせた配列。
SeqMan Ngen	SeqMan Ngen	NGS シーケンスデータのアセンブリ用ソフトウェア(https://www.dnastar.com/t-nextgen-seqman-ngen.aspx)。
sequence depth	シーケンス深度	同じ箇所を何回シーケンスしたかを表す尺度。冗長度とも言われる。「20×」であれば 20 回同じ箇所をシーケンスしたことを意味する。
short read	ショートリード	通常、1 リードの長さが 1000 bp 未満の場合を言う。
single read	シングルリード	断片化した塩基配列の一端からのみ解読する方法。
Snpeff	Snpeff	変異アノテーションと影響の予測を行うツール (http://snpeff.sourceforge.net/SnpEff.html)。
SOAPdenovo	SOAPdenovo	<i>de novo</i> アセンブリ解析に使用するソフトウェア。
SOLiD	SOLiD	ThermoFisher Scientific 社が提供するショートリード型の NGS 機器のひとつ。
Southern by- Sequencing (SAS)	Southern by- Sequencing (SAS)	シーケンスキャプチャー及び NGS を組み合わせにより、プラスミド骨格の検出、挿入コピー数確認、ジャンクション配列の確認が行える方法 (Zastrow-Hayes GM et al. Plant Genome. 2015; 8(1).)。
SSAKE	SSAKE	ショートリード用の <i>de novo</i> アセンブリツール (http://www.bcgsc.ca/platform/bioinfo/software/ssake)。

用語 (英語)	用語 (和訳)	説明
Target (Capture) sequence/sequencing (TCS)	ターゲットキャプチャシーケンス/シーケンシング	解析対象とするターゲット領域に相補的なDNA配列 (プローブもしくはベイト) を用いて濃縮し、その部分のみをシーケンス解析する方法。
tophat	tophat	RNA-seq の解析でよく用いられるソフトウェア。Mapping 後の塩基配列の連続性及び Gap から、遺伝子の構造 (エクソン領域及びイントロン領域) を推定する。
trimming	トリミング	シーケンスした生データの DNA 配列から解析に不要な配列を除くこと。
Trinity	Trinity	RNA-Seq データからのトランスクリプトーム解析で使用する de novo アセンブル用のソフトウェア (https://github.com/trinityrnaseq/trinityrnaseq/wiki)。
Velvet	Velvet	ショートリード用の de novo アセンブリツール (https://www.ebi.ac.uk/~zerbino/velvet/)。
whole exome sequence or whole exome sequencing (WES)	全エクソームシーケンスまたは全エクソームシーケンシング (WES)	全ゲノムのエクソン領域のみシーケンスする方法。
whole genome sequence or whole genome sequencing (WGS)	全ゲノムシーケンスまたは全ゲノムシーケンシング (WGS)	全ゲノムの塩基配列をシーケンスする方法。