Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

31-Oct-2016

English - Or. English

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

HIGH-THROUGHPUT DNA SEQUENCING IN THE SAFETY ASSESSMENT OF GENETICALLY ENGINEERED PLANTS: PROCEEDINGS OF THE OECD WORKSHOP (April 2016)

Series on the Safety of Novel Foods and Feeds
No. 29

JT03404139

ENV/JM/MONO(2016)57
Unclassified
English - Or. English

OECD Environment, Health and Safety Publications

Series on the Safety of Novel Foods and Feeds

No. 29

High-throughput DNA Sequencing in the Safety Assessment of Genetically Engineered Plants: Proceedings of the OECD Workshop (April 2016)

Environment Directorate

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Paris 2016

Also published in the Series on the Safety of Novel Foods and Feeds:

[No. 1, Consensus Document on Key Nutrients and Key Toxicants in Low Erucic Acid Rapeseed (Canola) (2001) – ***REPLACED with revised consensus document No. 24 (2011)***]

[No. 2, Consensus Document on Compositional Considerations for New Varieties of Soybean: Key Food and Feed Nutrients and Anti-nutrients (2001) – ***REPLACED with revised consensus document No. 25 (2012)***]

No. 3, Consensus Document on Compositional Considerations for New Varieties of Sugar Beet: Key Food and Feed Nutrients and Anti-nutrients (2002)

No. 4, Consensus Document on Compositional Considerations for New Varieties of Potatoes: Key Food and Feed Nutrients, Anti-nutrients and Toxicants (2002)

No. 5, Report of the OECD Workshop on the Nutritional Assessment of Novel Foods and Feeds, Ottawa, Canada, February 2001 (2002)

No. 6, Consensus Document on Compositional Considerations for New Varieties of Maize (*Zea mays*): Key Food and Feed Nutrients, Anti-nutrients and Secondary Plant Metabolites (2002)

No. 7, Consensus Document on Compositional Considerations for New Varieties of Bread Wheat (*Triticum aestivum*): Key Food and Feed Nutrients, Anti-nutrients and Toxicants (2003)

No. 8, Report on the Questionnaire on Biomarkers, Research on the Safety of Novel Foods and Feasibility of Post-Market Monitoring (2003)

No. 9, Considerations for the Safety Assessment of Animal Feedstuffs Derived from Genetically Modified Plants (2003)

[No. 10, Consensus Document on Compositional Considerations for New Varieties of Rice (*Oryza sativa*): Key Food and Feed Nutrients and Anti-nutrients (2004) ***REPLACED with revised consensus document No. 28 (2016)***]

No. 11, Consensus Document on Compositional Considerations for New Varieties of Cotton (*Gossypium hirsutum* and *Gossypium barbadense*): Key Food and Feed Nutrients and Anti-nutrients (2004)

No. 12, Consensus Document on Compositional Considerations for New Varieties of Barley (*Hordeum vulgare* L.): Key Food and Feed Nutrients and Anti-nutrients (2004)

No. 13, Consensus Document on Compositional Considerations for New Varieties of Alfalfa and Other Temperate Forage Legumes: Key Feed Nutrients, Anti-nutrients and Secondary Plant Metabolites (2005)

No. 14, An Introduction to the Food/Feed Safety Consensus Documents of the Task Force for the Safety of Novel Foods and Feeds (2006)

No. 15, Consensus Document on Compositional Considerations for New Varieties of the Cultivated Mushroom *Agaricus Bisporus*: Key Food and Feed Nutrients, Anti-nutrients and Toxicants (2007)

No. 16, Consensus Document on Compositional Considerations for New Varieties of Sunflower: Key Food and Feed Nutrients, Anti-nutrients and Toxicants (2007)

No. 17, Consensus Document on Compositional Considerations for New Varieties of Tomato: Key Food and Feed Nutrients, Anti-nutrients, Toxicants and Allergens (2008)

No. 18, Consensus Document on Compositional Considerations for New Varieties of Cassava (*Manihot esculenta* Crantz): Key Food and Feed Nutrients, Anti-nutrients, Toxicants and Allergens (2009)

No. 19, Consensus Document on Compositional Considerations for New Varieties of Grain Sorghum [*Sorghum bicolor* (L.) Moench]: Key Food and Feed Nutrients and Anti-nutrients (2010)

No. 20, Consensus Document on Compositional Considerations for New Varieties of Sweet Potato [*Ipomoea batatas* (L.) Lam.]: Key Food and Feed Nutrients, Anti-nutrients, Toxicants and Allergens (2010)

No. 21, Consensus Document on Compositional Considerations for New Varieties of Papaya (*Carica papaya* L.): Key Food and Feed Nutrients, Anti-nutrients, Toxicants and Allergens (2010)

No. 22, Consensus Document on Molecular Characterisation of Plants Derived from Modern Biotechnology (2010)

No. 23, Consensus Document on Compositional Considerations for New Varieties of Sugarcane (*Saccharum* spp. hybrids.): Key Food and Feed Nutrients, Anti-nutrients and Toxicants (2011)

No. 24, Revised Consensus Document on Compositional Considerations for New Varieties of Low Erucic Acid Rapeseed (Canola): Key Food and Feed Nutrients, Anti-nutrients and Toxicants (2011)

No. 25, Revised Consensus Document on Compositional Considerations for New Varieties of Soybean [*Glycine max* (L.) Merr.]: Key Food and Feed Nutrients, Anti-nutrients, Toxicants and Allergens (2012)

No. 26, Consensus Document on Compositional Considerations for New Varieties of Oyster Mushroom (*Pleurotus ostreatus*)*:* Key Food and Feed Nutrients, Anti-nutrients and Toxicants (2013)

No. 27, Consensus Document on Compositional Considerations for New Varieties of Common Bean (*Phaseolus vulgaris* L.)*:* Key Food and Feed Nutrients, Anti-nutrients and Other Constituents (2015)

No. 28, Revised Consensus Document on Compositional Considerations for New Varieties of Rice (*Oryza sativa*)*:* Key Food and Feed Nutrients, Anti-nutrients and Other Constituents (2016)

ABOUT THE OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 35 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in ten different series: **Testing and Assessment**; **Good Laboratory Practice and Compliance Monitoring**; **Pesticides and Biocides**; **Risk Management**; **Harmonisation of Regulatory Oversight in Biotechnology**; **Safety of Novel Foods and Feeds**; **Chemical Accidents**; **Pollutant Release and Transfer Registers**; **Emission Scenario Documents;** and **Safety of Manufactured Nanomaterials**. More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (http://www.oecd.org/chemicalsafety/).

This publication is available electronically, at no charge.

For this and many other Biosafety and Novel Food & Feed Safety publications, consult the OECD's World Wide Web site (www.oecd.org/biotrack/)

or contact:

OECD Environment Directorate,
Environment, Health and Safety Division
2, rue André-Pascal
75775 Paris Cedex 16
France

E-mail: ehscont@oecd.org

FOREWORD

The OECD's Task Force for the Safety of Novel Foods and Feeds decided at its first session, in 1999, to focus its work on the development of science-based *consensus documents*, which are mutually acceptable among member countries. These consensus documents contain information for use during the regulatory assessment of a particular food/feed product. In the area of food and feed safety, consensus documents are being published on the nutrients, anti-nutrients or toxicants, information of its use as a food/feed and other relevant information.

The scope of this document is different from that of the 'consensus documents'. It constitutes the ***proceedings of the OECD Workshop on High-throughput DNA Sequencing in the Safety Assessment of Genetically Engineered Plants***, held on 18 April 2016.

The document collates the summaries of the presentations delivered during the workshop, which were kindly prepared by the lecturers. It was edited by the OECD Secretariat in collaboration with the workshop organisers from Belgium, Canada, the Netherlands and the United Sates, and the Bureau of the Task Force for the Safety of Novel Foods and Feeds.

This document is published under the responsibility of the Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology of the OECD.

TABLE OF CONTENTS

INTRODUCTION

On April 18, 2016, the Organisation for Economic Co-Operation and Development (OECD) hosted a workshop on High-throughput DNA Sequencing in the Safety Assessment of Genetically Engineered Plants, held at the OECD Conference Centre in Paris, France. The workshop reviewed available information and provided a forum for countries to share experiences on new techniques used for molecular characterisation of genetically-engineered (GE) plant varieties during the risk/safety assessment process. The workshop presented the principles of high-throughput DNA sequencing and included topics such as the basics of the technique as applied to molecular characterisation and the bioinformatics necessary for compiling and interpreting the resulting data. Representatives from Canada, Belgium, the Netherlands, the United States and the Business and Industry Advisory Council (BIAC) shared their experiences and insights on applying this technology during the GE crops safety assessment. The final agenda of the workshop is reproduced in the Annex to this document.

Organised at the initiative of the OECD Task Force for the Safety of Novel Foods and Feeds (Task Force), the workshop was also open to delegates of the Working Group on Harmonisation of Regulatory Oversight in Biotechnology (Working Group), and other interested experts. A total of 85 individuals (biosafety regulators and experts) from 27 countries, the European Union and three international bodies (Food and Agriculture Organization of the United Nations, NEPAD-African Biosafety Network of Expertise, ILSI Research Foundation) participated in the workshop. This proceedings document is intended to provide interested individuals with a summary of the workshop discussion.

The workshop benefited from a specific grant provided by the USDA-Foreign Agriculture Service which supported its organisation and provided travel support allowing for participation from eight non-OECD member countries. This contributed to developing the country capacities in environmental safety and the safety of novel foods and feeds, and to strengthening cooperation between national Authorities responsible for the safety assessment of GE products worldwide.

This workshop was co-organised by Philippe Herman (Belgium), Luc Bourbonnière (Canada), Gijs Kleter (Netherlands) and Jason Dietz (United States) in collaboration with the Secretariat of the OECD's Environment Directorate, Health and Safety Division (Bertrand Dagallier and Peter Kearns). The agenda for the workshop was set by the organisers during the course of 2015, and the relevant high-level scientists and regulators were identified to deliver presentations and exchange views and experiences with the participants.

PROCEEDINGS OF THE WORKSHOP

PART 1. PRINCIPLES OF HIGH-THROUGHPUT DNA SEQUENCING

1. BASICS OF THE HIGH-THROUGHPUT DNA SEQUENCING USED FOR MOLECULAR CHARACTERISATION OF GENETICALLY ENGINEERED CROPS

Mauro Petrillo, Joint Research Centre, European Commission

The Human Genome Project took about 13 years and cost millions of dollars and considerable effort to be completed in 2003. The advent of Next Generation Sequencing (NGS) technologies in the period 2004-2006 provided a new and incredible impulse to research. In fact, compared to standard classical DNA sequencing (Sanger), NGS allows the reading of DNA on a massive scale and for little cost. After more than 10 years, we are now collecting trillions of bases as billions of sequences (reads), generating huge Big Data that needs to be analysed.

Then the NGS technologies currently available on the market were presented:

- 454 Roche technology, based on pyrosequencing;
- ION Torrent, based on ion sensors to detect proton release;
- SOLiD, which uses DNA ligase and a two-base-encoded probes system;
- Ilumina technology, based on sequencing by synthesis.

For each of them, a summary of the pros and cons was described, in order to highlight the fact that none of them, like any technology, is "perfect".

In addition, a global outline of the new NGS technologies which are arriving soon was provided:

- PACBIO, which allows Single Molecule Real Time Sequencing;
- Nanopore, a portable NGS device which can be easily transported in the pocket.

Differently from the previous ones, they are able to read very long stretches of DNA, without fragmentation and the associated assembly through complex and tedious protocols.

Moreover, an overview of the possible applications was explained. Different approaches exist, such as de-novo genome sequencing, transcriptome sequencing, targeted sequencing, ultra-deep sequencing. According to the aim of the experiment, the right approach should be used and, as a consequence, this affects the selection of the most appropriate NGS technology.

Only in the last five years has NGS been applied for characterisation of genetically engineered (GE) crops, maybe because NGS has only recently become more precise and cheap. Through NGS it is possible to acquire new information about GE crops and to develop ad hoc NGS-based strategies for their characterisation. However, challenging questions remain to be addressed, especially when referring to its use in routine control laboratories:

- Is NGS still too expensive for routine laboratory use?
- How many samples have to be analysed?
- What's the limit of detection of this technology?
- Is NGS also suitable for quantification of GE content as is required by legislation?
- How to analyse the large amount of produced data in a standardised, agreed and harmonised manner?

The end of the presentation focused on the role of Bioinformatics in the process of data analysis of NGS output, thus providing an introduction to the next presentation.

2. BIOINFORMATIC ANALYSIS OF HIGH-THROUGHPUT OR MASSIVELY PARALLEL DNA SEQUENCING DATA FROM GENETICALLY ENGINEERED CROPS

Michael L. Kotewicz, Ph.D., United States Food and Drug Administration, Centre for Food Safety and Applied Nutrition

On November 15, 2013, the U.S. Food and Drug Administration received its first regulatory submission in support of the safety assessment of food from a genetically engineered (GE) crop in which the inserted DNA was characterised using data derived from high-throughput DNA sequencing, or more accurately massively-parallel DNA sequencing (MPS). MPS data were used to characterise the insert structure, insert-genome junctions and generational stability for a GE maize variety described in Biotechnology Notification File 145.

This is a short discussion of the bioinformatic processes and software used to generate, quality control, and analyse MPS data for characterisation of inserted DNA in GE crops, and it focuses on analysis of data from the Illumina sequencing platform. There are consistent logic paths and methods for handling and analysing GE event data despite ongoing developments in software and sequencing platforms. From 2013 to 2016, MPS technologies employing genome fragmentation with in *vitro* transposons and final sequencing by synthesis have gained dominance. These technologies allow: 1.) sequencing of multiple samples within one run or multiplex sequencing (index primers); 2.) hybridisation and Bridge Amplification, and 3.) the creation of primer sites for paired end sequencing (Illumina Sequ. Tech., 2013).

Plant genomes vary greatly in size from 24 chromosomes in rice, totalling 389 million bases (Mb) to 20 chromosomes in corn containing a total of 2,300 Mb. In these sequencing by synthesis reactions, the data, massive sets of 100-200 base pair reads equalling approximately 100 million reads for a rice genome, require different types of bioinformatic analysis. Three types of bioinformatic analysis will be discussed.

2.1 Raw read quality control software

In a typical analysis, using a GE rice plant containing a 10 kb[1] plasmid insertion intended to express a new gene, 100 million sequence reads are generated (Yang et al., 2013). This is 100 million pieces of data, each about 100 bases long. In preparing the data for analysis, non-relevant sequence from multiplexing adaptors at the ends of sequence reads as well as priming adaptors have to be removed, and critically, generally at the ends of these molecules, there will be sequencing errors that can be identified and removed. There may also be errors from other, different sources such as: optic and fluid handling errors; reaction and reagent flow errors; polynucleotide track run errors, and; sequence dependent location errors. Sequence reads resulting from these errors have to be removed using bioinformatic software programs. Typically, the ends of these sequence reads are trimmed and excluded from the data analysis in order to focus on the highest fidelity portion of the sequence reads.

Massively-parallel DNA sequencing by synthesis relies on multiple overlapping reads, or coverage, typically 20, 50 or even several hundred reads. As a result, there will be differences in sequence reads at some positions with full agreement of a base call at most other positions. Reads with statistically determined errors will need to be trimmed to remove the errant portion of the sequence read or completely

[1] kb unit stands for 'kilobase', equal to 1000 base pairs of DNA or RNA.

removed from further analysis because they do not provide high fidelity sequence data. This is one of the most important quality control steps. FastQC is a popular user-driven program that can be used to remove such reads from further analysis. Vendor-driven software packages such as CLCBio, PacBio, or DnaStar, give a menu-driven path with several settings, options, and steps to first remove adaptors and then remove reads with errors. These options and steps, again act as a quality control filter. Quality (N) of MPS data is measured using probability-based calculations, presented on a logarithmic scale. For example, an N20 filter or cut off removes reads with error rates above 1 in 100, N30, reads with a probability of errors above 1 in 1,000, N40, above 1 in 10,000. These software packages generate graphic and tabular results of these operations.

2.2 Assembly software

The assembly process itself is a full course of study, and will not be discussed in any detail here. Assembly software uses data from the sequence reads and overlaps them to produce longer sequences called 'contigs'. The subsequent assembly of contigs can ultimately lead to the process of generating reference genomes. Assembly software is often part of vendor packages from Celera, CLCbio, PacBio, and also includes stand-alone programs such as Velvet and Spades. There are some programs used in assemblies, BLAST, Clustal, Muscle, Mummer, and Mauve relating to alignment, that are important in junction sequence analysis, this is discussed in more detail in the next section.

2.3 Junction sequence analysis and alignment software

The most critical procedures of sequence data processing for the review of GE crops are junction analysis. The complete set of sequence reads from a GE plant is aligned to the vector DNA sequence that was used to create the GE plant. This process identifies a distinctive set of sequence reads that contain both vector sequence as well as plant sequences. These hybrids define the site where the vector inserted into the plant genome. In the simplest of cases, there will be two junction sequence classes, one at the left and one to the right of the site where the vector inserted into the plant genome. In other cases there will be more than two sets of junction sequences. These indicate multiple insertions, and require more detailed examinations to determine if any further rearrangements have occurred.

Two examples of searching GE reads for junctions in rice were presented. Yang et al. (2013) presented bioinformatic plans based on three different scenarios of information known about a GE plant. The second example used CLCbio software to find sequence reads in a rice Illumina HiSeq2000 data set at NCBI in the Sequence Read Archive (SRA) files, matching a transformation vector sequence (NCBI SRA, 2015).

References:

Illumina Sequencing Technology (2013), *Video providing an overview of the DNA sequencing workflow on an Illumina sequencer*, posted on Oct. 23, 2013 at https://www.youtube.com/watch?v=womKfikWlxM.

Yang L., C. Wang, A. Holst-Jensen, D. Morisset, Y. Lin and D. Zhang (2013), "Characterization of GM events by insert knowledge adapted re-sequencing approaches", *Natureresearch Journal, Scientific Reports No.3:2839,* http://www.nature.com/articles/srep02839

NCBI Sequence Read Archive (2015), *SRX187564: Whole genome re-sequencing can improve basis for transgene safety assessments, Sequence SRR567751; Organism:* Oryza sativa, *TT51-1 (rice), submitted by Shanghai Jiao Tong University,* National Center for Biotechnology website, National Library of Medicine, United States of America, available at: http://www.ncbi.nlm.nih.gov/sra/?term=SRR567751.

Acknowledgements:

M. Kotewicz thanks Mark K. Mammel and Jayanthi Gangiredla for their assistance in preparing this talk.

PART 2. APPLICATION OF HIGH-THROUGHPUT DNA SEQUENCING TO MOLECULAR CHARACTERISATION OF GENETICALLY ENGINEERED CROPS

3. *CANADIAN EXPERIENCES AND INSIGHTS*: WHOLE GENOME SEQUENCING INTERNAL GUIDANCE FOR MOLECULAR EVALUATORS

Jennifer Holtzman, Ph.D., Novel Foods Section, Health Canada

3.1 Summary

Health Canada has led the drafting of an internal guidance document on whole genome sequencing (WGS), in collaboration with the Canadian Food Inspection Agency (CFIA) and Agriculture and Agri-food Canada (AAFC), for evaluators who assess the molecular characterisation data provided by companies as part of their pre-market submissions for novel foods, novel feeds, and plants with novel traits. The purpose of the guidance is to provide evaluators at Health Canada and the CFIA with the technical and background knowledge required to continue to deliver high quality and timely safety assessments of genetically modified organisms (GMOs) in light of the industry's movement towards the use of this rapidly developing technology in their molecular characterisations.

3.2 Background

Products derived from GMOs are required under Canadian law to undergo a safety assessment prior to their sale in Canada. The molecular characterisation portion of the assessment requires that companies provide data on changes to the genome of the modified organism that were introduced, including the number of DNA insertion sites in the genome, the presence of complete or partial copies of any inserted DNA, whether the inserted DNA is intact compared to what was expected, and the absence of any antibiotic resistance markers or other unintended genetic elements used during development.

This characterisation has traditionally been accomplished using classical molecular biology techniques with which evaluators are quite familiar. In 2013, some companies began providing WGS data in support of the molecular characterisation of their products. The technologies for rapid sequencing of whole genomes have become increasingly affordable and accessible to companies, including through outsourcing to specialised laboratory service providers. While the technology is well established and widely used across many disciplines of biomedical research, the assessment of WGS analysis requires some knowledge of bioinformatics that was new to some of the molecular evaluators at Health Canada and the CFIA.

As a first step toward bridging the gap in expertise, the CFIA led a workshop on March 30-31, 2015 in Ottawa, which brought together experts from industry and academia, as well as regulators from across the Canadian government, the United States, and Mexico. Based on the suggestions gathered at the meeting, it was concluded that an internal guidance document could be formulated by drawing on a combination of peer-reviewed literature and consultation with Government of Canada experts in DNA sequencing and bioinformatics. This approach presented the advantage of a substantially shortened timeline without the need for additional expenditures or specialised expertise compared to other approaches that were considered (e.g. producing new sequencing data).

An advanced draft was completed within six months of the initial meeting and was presented in a workshop aimed at molecular evaluators from Health Canada and the CFIA held in November 2015 in Ottawa. The objectives of the guidance are to provide background information at an appropriate level of technical detail, familiarise evaluators with terminology and fundamental concepts in genome sequence analysis, and explain the advantages and limitations of WGS. Simulated data was used to prepare case studies in order to demonstrate the application of the guidance in a realistic context.

3.3 Impact and next steps

This capacity building exercise has had an immediate impact by facilitating the review of current submissions containing WGS data analysis components. The internal guidance was recently finalised by the working group and is in the process of being distributed to evaluators. As a next step, the working group will extract key principles from the internal guidance to generate a concise external key points document for industry.

4. *BELGIAN EXPERIENCE AND INSIGHTS:* APPLICATION OF HIGH THROUGHPUT DNA SEQUENCING TO MOLECULAR CHARACTERISATION OF GENETICALLY ENGINEERED CROPS

Katia Pauwels, Ph.D., Scientific Institute of Public Health (WIV-ISP), Belgium

Molecular characterisation data of genetically engineered (GE) plants provide valuable information for GE plant developers (e.g. for selection purposes), public risk/safety assessors (to assess the risk /safety of GE plant with regard human health and environment) and enforcement laboratories for the development and validation of detection methods. For many years Southern blot and Sanger sequencing have been a 'gold method' for obtaining molecular characterisation data in the context of their risk/safety assessment. Now the increasing DNA sequencing throughput possibilities at continuously decreasing costs combined with their digital nature and tuneable resolution have paved the way for their application in the domain of molecular characterisation of GE plants.

To anticipate the use of data generated by high-throughput sequencing technologies (HTS) in upcoming risk/safety regulatory dossiers and to identify crucial points for their evaluation, the Biosafety and Biotechnology Unit has co-organised with the Platform Molecular Biology and Biotechnology of the WIV-ISP (Scientific Institute of Public Health) an international workshop (held in Brussels, Belgium, November 25, 2013) to examine how HTS could offer an added value for the molecular characterisation of GE plants compared to traditional molecular approaches including conventional Southern blot/Sanger sequencing. Data requirements as described in the Commission Implementing Regulation (EU) No 503/2013 were used as a starting point. Participants from various countries (from and outside the European Union) and different sectors (risk assessors, product developers, enforcement laboratories, competent authorities) shared their scientific views and expertise. In addition, particular attention was given to the potential added value of HTS for molecular characterisation with respect to current and emerging developments in plant breeding techniques such as stacked events, GE plant developed by the use of RNA interference, cisgenesis or targeted mutagenesis (use of oligonucleotides or site-directed nucleases (SDN)).

Depending on the library and the HTS platform used, reads usually do not exceed ≤ 300 nucleotides. Therefore, for GE plants obtained by the non-targeted insertion of a large cassette of foreign DNA, sequence verification and validation is still performed by means of Sanger sequencing. However, the (EU) No 503/2013 requirement to re-sequence inserts and flanking regions may favour HTS compared to conventional Southern blot/Sanger sequencing in case of stacked events because of the more streamlined

and standardised procedure along with an increased cost-effectiveness. We also note that upcoming HTS platforms will facilitate transcriptome profiling, which will be of particular relevance for GE plants developed by the use of RNA interference technology. With regard cisgenic plants, for which the inserted cassettes may show high sequence identity or similarity to endogenous sequences of the recipient genome, it is possible that further developments in HTS technology will offer better perspectives compared to the classical Southern blot/Sanger sequencing. For another class of plants developed by novel breeding techniques, the plants obtained by targeted mutagenesis, whole genome sequencing has already been applied for detecting the number of indels[2] and single –nucleotide polymorphisms in genome edited rice.

However, irrespective of the technical possibility to perform such analysis, it is important to keep in mind that it will not be possible to distinguish off-target mutations induced by erroneous targeting of the engineered nucleases (e.g. CRISPR-Cas9 system) from spontaneous mutations that occur during cell culture. Moreover, given that plant genomes are very dynamic and plastic, it should be questioned to what extent the unintended presence of small inserts or small nucleotide changes in the plant genome needs to be identified and characterised taking into account that potential associated unintended effects are also assessed on the basis of agronomic, phenotypic and compositional properties. Therefore, from a risk assessor's point of view and irrespective of the technical possibilities offered by HTS, focus should remain on the gathering of data that actually inform risk assessment.

Other questions raised along with the use of HTS are associated to the substantial amount of raw data generated and the choice of algorithms. Different analysis programmes or parameters in handling these data may lead to different conclusions on the molecular characterisation. Hence, the set-up of criteria to retrieve and present relevant and interpretable data of good quality will be both important and challenging. Therefore we advocate the set-up of a common workflow for the generation of relevant and interpretable data by HTS will facilitate a scientifically sound assessment of GM plants.

Reference:

Pauwels K., S.C.J. De Keersmaecker, A. De Schrijver, P. du Jardin, N.H.C. Roosens and P. Herman (2015), "Next-Generation sequencing as a tool for the molecular characterisation and risk assessment of genetically modified plants: added value or not?", *Trends in Food Science & Technology Vol. 45(2)*, pp. 319-326.

5. *DUTCH EXPERIENCES AND INSIGHTS*: USE OF RNA SEQUENCING FOR THE CHARACTERISATION OF GENETICALLY ENGINEERED PLANTS

Esther Kok, Ph.D., RIKILT Wageningen UR, the Netherlands

There are different approaches to confirm the successful integration or indeed to identify potentially present unintended effects of a genetic modification in new genetically engineered (GE) varieties. In most countries this is, in accordance with FAO/WHO and OECD guidelines, performed by: i) a detailed molecular biological analysis, and ii) a phenotypic, agronomic and compositional comparison of the GE plant and its conventional counterpart. Also, in those cases where the molecular biological data are not conclusive, the compositional data are usually considered as most informative to assess for any potentially present unintended side effect of a genetic modification. In addition to this, in some countries a 90-day feeding trial with the whole GE plant variety and its conventional counterpart is mandatory as well. In other countries this type of animal feeding trials with the whole food is considered not sensitive enough as an indicator of potential unintended changes to be performed on a routine basis, besides ethical reasons

[2] Indels: abbreviation used for the **in**sertions and **del**etions of bases in the DNA of an organism

for not performing animal trials unless strictly necessary. All approaches have in common that they aim to identify potentially present unintended effects of a breeding strategy that may include genetic modification, while it is good to keep in mind there are only very few cases where unintended side effects have been observed and there are no examples where unintended effects were shown to be of toxicological relevance.

As the compositional analysis is thus a central part of the risk assessment of new (GE) plant varieties, it has been advocated in different national and international scientific reports to further develop advanced omics technologies as a more informative strategy to perform a compositional analysis. Plant breeders nowadays often use detailed compositional analyses to screen for unintended characteristics in the new plant variety. Basically the same data can also be used to screen new plant varieties for potentially present unintended effects in a way that is similar to the current targeted approaches, but resulting in more detailed compositional comparisons.

At this moment in time transcriptomics (RNA sequencing (RNASeq)) is the most informative omics approach, as RNASeq data generally cover over 90% of the transcriptome. RNA sequencing can be used in two ways to screen for unintended side effects of the genetic modification: 1) by comparing the RNASeq profile with profiles from conventional varieties that we consider as safe, and 2) by identifying novel transcripts in the whole pool of RNA transcripts. This latter approach can also be used to confirm the intended effect.

Both approaches have been assessed in the European GRACE project. RNASeq data have been obtained, for instance, from a GE maize variety that contained the MON810-construct. The RNASeq data from the MON810 variety and from the conventional variety were aligned to the maize reference genome, and the reads that could not be mapped to the reference genome were analysed in a *de novo* assembly. Based on this *de novo* assembly the gene coding for the Cry1A(b) protein could easily be identified, and no other new transcripts were found that could indicate a potential unintended effect of toxicological concern.

The second approach in the GRACE project was to compare the RNASeq profiles of the GE varieties with similar RNASeq profiles of conventional varieties that we consider as safe. As mentioned, it has been advocated in the past already to use omics technologies to perform a more informative and more cost-effective compositional analysis. It proved, however, that the large RNASeq datasets made it often difficult to perform an informative data analysis, as many individual transcripts will be shown to differ between repeated analyses, even if the repeats concern the same plant material. And also, the analysis of RNASeq data should take into account that there is much natural variation between plant samples due to the genotype, also in conventional varieties, but also due to even small differences in growth conditions, whether related to the soil type, the fertilisation and spraying regime, or to local climatological conditions. This should all be taken into account when assessing new RNASeq profiles.

To do this effectively, a one-class SIMCA (Soft Independent Modelling of Class Analogies) classification tool has been developed that allows classifying a new profile in terms of similarity to the profiles that form the one class of profiles from conventional varieties. The one class model has been developed in such a way that profiles of varieties that cannot be considered as similar to the profiles of the conventional varieties included in the one class, should fall outside of the one class. Similarly, safe new profiles may be classified as outside of the model ('false positives'), in this case it will be necessary to look into the details of the altered physiology whether there are any indications of differences with conventional varieties that may be of toxicological concern. It is clear that the 'out' classification as such does not imply that a particular new plant variety is not safe; it will only trigger a further assessment of available data. In this way the RNASeq one class approach is similar, but much more informative when compared to the current compositional analyses that are targeting micro-nutrients and anti-nutrients in plant samples. The application in the GRACE project has shown that the different GE varieties (maize and potato)

that were used in the project were all classified as inside the one class model. Using the same approach it was shown that genetically more distant varieties, were, although safe for human consumption, classified as outside of the one class model. Similarly, based on metabolomics data, samples that were clearly of inferior quality were also classified as outside of the one class model.

Based on these results it can be further confirmed that RNASeq and other omics approaches may be more informative as well as more cost-effective with relation to identifying potentially present unintended side effects of a plant breeding strategy compared to current targeted approaches, especially when the latter are routinely combined with animal feeding trials with whole foods.

Acknowledgements:

RIKILT Wageningen UR, the Netherlands*: Dr. Jeroen van Dijk, Dr. Martijn Staats, Marleen Voorhuijzen, Martijn Slot M.Sc., Dr. Roberta Mariot, Dr. Joseph Evaristo, Rico Hagelaar, Dr. Gijs Kleter.* CRAG, Spain*: Dr. Maria Pla, Dr. Maria Corujo.* CSIR, South Africa*: Dr. Eugenia Barros.* WUR Biometris, the Netherlands*: Dr. Hilko van der Voet.* WUR Plant Breeding, the Netherlands*: Dr. Ronald Hutten, Prof. Dr. Richard Visser.* University of Nijmegen, the Netherlands*: Dr. Jeroen Jansen.*

6. *UNITED STATES EXPERIENCES AND INSIGHTS*: HIGH THROUGHPUT DNA SEQUENCING IN THE SAFETY ASSESSMENT OF GENETICALLY ENGINEERED CROPS

Michael L. Kotewicz, Ph.D., United States Food and Drug Administration, Centre for Food Safety and Applied Nutrition

By April 1995, under the "Statement of Policy: Foods Derived from New Plant Varieties" from the U.S. Food and Drug Administration (U.S. FDA 1992), a half dozen voluntary premarket food safety consultations on genetically engineered (GE) varieties of soy, tomato, potato, cotton, squash and canola had been completed. Almost 20 years later, in November 2013, the first consultation using high-throughput DNA sequence data, in this case, massively-parallel DNA sequencing (MPS) in support of the safety assessment of food from a GE plant was submitted to the FDA for GE maize variety, Biotechnology Notification File (BNF) 145. A historical perspective that describes the formation of GE plant safety concerns can be found at the FDA website (U.S. Public Health Service, 2006).

Inserted DNA in foods (including foods for Humans and other animals) from GE plants is routinely characterised as part of food safety assessment. Characterisation addresses: 1) organisation of inserted DNA, including the site(s) of insertion as well as possible plant genome or vector rearrangements including truncations; 2) Identification of expressed substances, especially from plant-vector junctions; and 3) generational stability. Traditionally, Southern blotting data, with relatively low resolution of insertion and deletion events (500 to 1,000 base pair), were used for characterisation of insert copy number, structure, and chromosome junctions. In contrast, for a GE rice plant genome with 24 chromosomes and a total of 389 million bases (Mb) containing a novel 10-40,000 base pair insertion, massively-parallel sequencing can produce 10,000,000 to 100,000,000 sequence reads for complete coverage of the genome.

Analysis of GE plants focuses on a small part of data from MPS; specifically, those sequence reads where vector sequences are joined to plant genome sequences. In its simplest form, junction analysis takes the millions of 100-200 base pair (bp) sequence reads from the genome of the GE plant and aligns them to the sequence of the vector used to transform the plant for the creation of a new trait. When sequencing a complex plant genome containing a relatively tiny insertion of vector sequence, bioinformatic software

is used to separate the bulk of DNA sequence reads, 99.999% containing only plant DNA, from the reads that contain vector sequences and reads containing vector-plant genome junctions.

Distinct hybrid sequences, containing sequences from both the vector and plant genome, define the junction site(s) of the insertion into the plant genome. In simple cases there will be two junction sequences, one to the left and one to the right of the insertion site. In other cases, there will be more than two sets of junction sequences. These indicate multiple insertions which may require more detailed examinations to determine the different sites of insertion as well as if any rearrangements or truncations of vector and plant DNA have occurred.

Another major consideration of MPS is coverage; i.e. how many times a base in a sequence has been found in the collected set of overlapping reads of a sequence. This is partly a quality of data consideration, but more importantly, coverage determines the degree of certainty that junction sequences will be recovered in a MPS data collection consisting largely of plant genome sequence reads. Calculations to measure the probability of detection of an insertion and minimise the possibility of missing an insert, show that 75-fold coverage is sufficient (Clarke and Carbon, 1976). There are other sequence considerations, in addition to MPS data analysis, where polymerase chain reaction (PCR) –directed sequence data for the insert as well as PCR sequence data across junctions supports MPS data junction analysis.

To date, FDA has completed consultations on food from three GE plant varieties that have used MPS data to characterise the inserted DNA. Although the submission of MPS junction data was new in these submissions, with basic requirements met for quality and coverage considerations, the fundamental similarity of the biochemistry of MPS produces data that are equivalent to those obtained using traditional sequencing; both techniques allow for the detailed evaluation of insert structure and junction sequences. The first consultation using MPS to characterise the inserted DNA was for a maize variety with insect resistance and herbicide tolerance. The MPS data submitted were from Illumina HiSeq short sequencing reads averaging 100 bp, and coverage was greater than 75-fold for the genome. The insert junction and the number of junctions based on recovery of unique junction site sequences were determined. In addition to MPS data, directed sequencing was performed on PCR products amplified from the 11 kb insertion event and from the insertion junctions to confirm the sequence of the entire insert and the insertion site. MPS data and junction sequence analyses on a total of five breeding generations were used to demonstrate generational stability as well as Mendelian inheritance of the traits. Spiking experiments using 1 and 1/10 genome equivalents of vector plasmids were also used to demonstrate required sensitivity of detection. Finally, analysis of the read coverage of a representative single-copy endogenous plant gene provides an additional quality determination for the MPS data.

Other consultations using MPS to characterise the inserted DNA included an insect resistant soybean and maize engineered to have increased ear mass. In these two submissions, MPS data for junction analysis, spiked control sensitivity, single copy endogenous gene coverage, and finally directed sequencing to confirm junctions and insert structure completed MPS scrutiny of the events. For the maize variety, data regarding sensitivity of MPS were also reported. A small number of reads (4) that mapped to the transformation plasmid backbone were noted, relative to thousands of reads that mapped to the T-DNA insertion. This brought to the fore another concern relating to signal-to-noise or contamination issues, which are currently being examined in numerous laboratories for MPS data. In this case, the reads mapping to the plasmid backbone were interpreted to be the result of a small amount of plasmid vector inadvertently present in the sequencing reaction based on the fact that the data contained thousands of reads mapping to the T-DNA insertion and only four reads mapping to plasmid backbone.

Looking towards future submissions containing MPS data, sufficiently complex insertion events may still warrant more extensive directed sequencing as they have in the past, i.e. head-to-tail multiple

insertions with truncations. Complex insertions may warrant sequence reads longer than 100 bp due to multiple copies of inserts and tandem inserts, or prompt at the very least, a more extended series of directed sequencing and PCR verification of complex arrangements. As has been noted, the organism, the complexity and nature of the insertion event(s) will still require a case-by-case evaluation of data, including the MPS data presented in terms of coverage and supplemental directed sequence data. Extended "host" range GE events, i.e. new crop species being modified and the difficulty of generational stability determinations for some plants will continue to add complexity to safety assessment of food from new GE plants.

In conclusion, FDA has received a number of GE plant consultations in which MPS data were used to characterise the inserted DNA. These data were successfully used to address the food safety considerations traditionally addressed through Southern blotting. High-throughput DNA sequencing, including massively parallel DNA sequencing, is a tool that can be used to address the same considerations historically addressed by Southern blotting.

References:

U.S. Food and Drug Administration (1992), "Statement of Policy-Foods Derived from New Plant Varieties", *FDA Federal Register, Volume 57, May 29, 1992,* available at: http://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/Biotechnology/ucm096095.htm.

U.S. Public Health Service (2006), "History of the U.S. Food and Drug Administration", *Interview of James Maryanski by Suzanne Junod, August 18, 2006*, Department of Health & Human Services, National Library of Medicine, available at: http://www.fda.gov/downloads/AboutFDA/WhatWeDo/History/OralHistories/SelectedOralHistoryTranscripts/UCM315241.pdf.

Clarke, L. and J. Carbon (1976), "A colony bank containing synthetic Col El hybrid plasmids representative of the entire E. coli genome", *Cell Vol. 9(1)*, pp. 91-99.

7. *INDUSTRY'S EXPERIENCE AND INSIGHTS*: THE USE OF NEXT GENERATION SEQUENCING DATA FOR REGULATORY SUBMISSIONS

Jafar Mammadov[1], Kristina Burgin[2], Huihua Fu[3]; Veerle Habex[4], Mary Locke[5], Andre Silvanovich[6], and Taiwo Koyejo[7]

[1]Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268; [2]Syngenta Crop Protection LLC, 3054 East Cornwallis Road, Research Triangle Park, NC 27709; [3]BASF Plant Science LLP, 26 Davis Drive, Research Triangle Park, NC, 27709; [4]Bayer CropScience, Technologiepark 38, 9052 Gent, Belgium; [5]DuPont Pioneer, Global Regulatory Science, 8325 NW 62nd Ave. P.O. Box 7062, Johnston, Iowa 50131-7062; [6]Monsanto Company, 800 N. Lindbergh Blvd., St. Louis, MO 63167; [7]CropLife International, A.I.S.B.L.326 Avenue Louise, Box 35, 1050 Brussels, Belgium.

7.1 Introduction

Molecular characterisation is an important part of safety assessment of transgenic crops. The main endpoints of molecular characterisations are: 1) insert/copy number; 2) absence/presence of backbone; 3) across generational stability; 4) T-DNA structural integrity; 5) T-DNA nucleotide integrity, and 6) flanking border sequences. Traditionally, the first four endpoints have been achieved by Southern blot analysis, while the fifth and sixth endpoints rely on Sanger sequencing. The emergence and rapid evolution of next-generation sequencing (NGS) technologies over the past few years have offered novel, rapid, and

cost-effective options for molecular characterisation of transgenic crops. Since 2012 the plant biotechnology industry has published several reports which share its experience in the application of NGS technologies for molecular characterisation of transgenic crops. In 2012, Monsanto Company (Kovalic et al., 2012) successfully demonstrated the application of whole genome sequencing (WGS) approach to determine T-DNA insert and copy number in soybean by sequencing the junction regions between the T-DNA and the flanking border genomic sequences. Moreover, it was proposed that through the evaluation of insert and copy number, the integrity and stability of insertions across multiple generations can be determined. In 2015, DuPont Pioneer reported the successful application of targeted sequence capture (TSC) coupled with NGS for high throughput event sorting (Zastrow-Hayes et al., 2015). Finally, in 2016 Dow AgroSciences published data related to the comparative analysis of WGS and TSC approaches for the characterisation of single and stacked transgenic events and compared the results and inferences with traditional method with respect to key criteria required for regulatory submissions (Guttikonda et al., 2016). In this report, we summarised industry experience in NGS-based molecular characterisation of transgenic crops for regulatory submissions focusing primarily on the WGS approach.

7.2 NGS read mapping reveals insert/copy number and presence/absence of backbone

To determine the number of copies of the T-DNA inserted in the genome and their locations as well as presence/absence of backbone, genomic DNA of the transgenic plant is randomly sheared and sequenced. Random shearing produces a mixture of three types of fragments – those derived solely from the plant genome (> 99.999%), those derived solely from the T-DNA and those derived from regions spanning the T-DNA integration site and thus consisting of both the T-DNA and the host genomic DNA. Paired-end reads generated from this third type of fragment will have one read of a pair mapped to the T-DNA and its mate mapped to the plant genome. In a subset of these junction pairs, one read of the pair will span the insertion-genomic DNA junction with a portion of read derived from the transgene and the other derived from the genome. These will be referred to as 'junction reads' or 'junction sequences'. With sufficient NGS coverage, from 10X (Guttikonda et al., 2016) up to 75X (Kovalic et al., 2012), the identification of a two junction read classes indicates a single copy insertion, whereas heterogeneous population of greater than two junction reads would point to multiple insertions in the genome. As WGS generates sequences from the entire genome, any vector backbone segments that are present in the transgenic plant will be sequenced and these reads will map back to the plasmid reference sequence. If no reads map back to vector backbone of the construct, this suggests a clean integration of the T-DNA and absence of any vector backbone in the transgenic plant.

7.3 T-DNA integrity and flanking borders

T-DNA integrity and flanking borders of transgenic crops can be determined by mapping reads back to a plasmid and genomic reference sequences. However, the application of NGS to determine these endpoints could be limited to the events with no endogenous regulatory elements and long repetitive sequences within T-DNA.

7.4 Stability is evaluated by mapping NGS reads from multiple generations

To determine generational stability of a T-DNA, several generations of an event can be subject to WGS. Consistency in number and sequence of junctions will indicate no differences in copy number and integrity of T-DNA suggesting that T-DNA is stably inherited across generations.

7.5 Conclusion

NGS has been shown as another technological option to carry out molecular characterisation of transgenic crops along with traditional Southern Blot and Sanger sequencing. This technology provides

direct sequence readout, is free of artefacts resulted from incomplete digestion, cross hybridisation, mobility issues, and radio-active hazards. In contrast to Southern Blot technique, NGS is amenable to automation, which makes it less labour-intensive. Due to its paired-end chemistry, NGS could also be used to resolve complex aberrations within T-DNA and junction regions. NGS- and Southern blot-based molecular characterisation yields identical conclusions in determining insert/T-DNA copy number, presence/absence of backbone, and generational stability. The use of NGS in determining T-DNA integrity and flanking border sequences depends on the nature of T-DNA. Currently, due to the short length of reads, the application of NGS could be limited to the events with no endogenous regulatory elements and long repetitive sequences within T-DNA.

References:

Guttikonda, S.K., P. Marri, J. Mammadov, L. Ye, K. Soe, K. Richey, J. Cruse, M. Zhuang, Z. Gao, C. Evans, S. Rounsley and S.P. Kumpatla (2016), "Molecular Characterization of Transgenic Events Using Next Generation Sequencing Approach", *PLoS ONE Vol. 11(2):e0149515*, doi:10.1371/journal.pone.0149515.

Kovalic D., C. Garnaat, L. Guo, Y. Yan, J. Groat, A. Silvanovich, L. Ralston, M. Huang, Q. Tian, A. Christian, N. Cheikh, J. Hjelle, S. Padgette and G. Bannon (2012), "The Use of Next Generation Sequencing and Junction Sequence Analysis Bioinformatics to Achieve Molecular Characterization of Crops Improved Through Modern Biotechnology", *The Plant Genome Vol. 5(3)*, pp. 149-163, doi: 10.3835/plantgenome2012.10.0026.

Zastrow-Hayes G.M., H. Lin, A.L. Sigmund, J.L. Hoffman, C.M. Alarcon, K.R. Hayes, T.A. Richmond, J.A. Jeddeloh, G.D. May and M.K. Beatty (2015), "Southern-by-Sequencing: A Robust Screening Approach for Molecular Characterization of Genetically Modified Crops", *The Plant Genome Vol. 8(1)*, pp. 1-15, doi:10.3835/plantgenome2014.08.0037.

CONCLUDING REMARKS

The moderator provided a summary of the day's presentations, underlining the interest of the high-throughput DNA sequencing techniques as a powerful tool for breeders as well as for regulators during the risk assessment process. Given the recent but speedy development of these techniques, it is crucial for national authorities to be kept fully aware of their limits and possibilities, and be informed of their increasing use in the development of new products in many regions of the world. The workshop was timely for fulfilling the information need and allowing exchange of views and experience between scientists, industry and regulators on the new technology and its potential use in risk assessment. It will also be valuable background reference as the OECD Task Force for the Safety of Novel Foods and Feeds plans its work for 2017 and beyond.

It was confirmed that the proceedings of the workshop were to be prepared by the lecturers and the organisers for publication as part of the OECD Food and Feed Safety Series in the course of the year.

The moderator thanked the presenters for sharing their expertise through their informative presentations. The moderator also thanked the attendees for their questions and discussion. Finally, the moderator thanked the organising committee and the OECD Secretariat for putting this workshop together.

ANNEX: AGENDA OF THE WORKSHOP
held at OECD Headquarters, Paris, France, on 18 April 2016

MORNING SESSION: INTRODUCTION AND BASIC PRINCIPLES OF HIGH-THROUGHPUT DNA SEQUENCING

10:00	**Registration and welcome** Welcome and final registration of participants. Workshop documentation is provided.
10:30	**INTRODUCTION TO THE WORKSHOP** Presentation of workshop aim and background information. i) central role of molecular characterisation in the safety assessment of genetically-engineered (GE) plants, including that of inserted DNA and products of transcription and translation (RNA, proteins); ii) wide range of potential applications for the high-throughput DNA sequencing technology; and iii) focus on GE plants safety assessment, for which the technology provides a different method to characterise the inserted DNA compared to currently-used techniques. The workshop organisers were from Belgium, Canada, the Netherlands and the United States. Dr. Kathleen Jones, U.S. FDA, Chair of the OECD Task Force for the Safety of Novel Foods and Feeds, is designated to chair the workshop. Speakers: *Organising committee and workshop Chair*
11:00	**PRINCIPLES OF HIGH-THROUGHPUT DNA SEQUENCING** Overview of genome sequencing and new methods that may be used in the molecular characterisation of GE plants: i) Basics of the high-throughput DNA sequencing methodology; ii) existing platforms used for this purpose; iii) bioinformatics and approaches needed to process the data, and iv) endpoints potentially addressed by this technology during the GE plant safety assessment, similar to those currently covered by Southern blots, Sanger sequencing, and Northern blotting for expression analysis.
	Basics of the high-throughput DNA sequencing used for molecular characterisation of GE crops Presentation discussing how DNA sequencing eventually evolved into massively parallel high-throughput DNA sequencing technology, basic chemistry and analytical considerations behind its use for characterising rDNA inserts and their RNA transcripts in GE plants. It will also explain why technology providers may choose to use this methodology over other methods. Speaker: *Mauro Petrillo, European Commission, Joint Research Centre (JRC)*

11:30	**Bioinformatics** Overview of the bioinformatics necessary for compiling and interpreting the large quantities of data emanating from high-throughput DNA sequencing. Presentation of some key bioinformatics considerations when associated with the characterisation of rDNA inserts in GE plants. Speaker: *Michael Kotewicz, Ph.D., U.S. Food and Drug Administration (FDA)*
11:50	**General discussion**
12:30	Lunch break

AFTERNOON SESSION: EXPERIENCES/INSIGHTS ON HIGH-THROUGHPUT DNA SEQUENCING, AND GENERAL DISCUSSION

14:00	**APPLICATION OF HIGH-THROUGHPUT DNA SEQUENCING TO MOLECULAR CHARACTERISATION OF GE CROPS** Country experiences in considering genome sequencing data in the safety assessment of foods and feeds derived from GE plants.
	Canadian experiences and insights Highlights on the outcomes of a workshop held to engage various stakeholders, incl. experts in Next Generation Sequencing (NGS) technology and bioinformatics. Progress on drafting guidance and training for scientific evaluators on NGS data interpretation and critical assessment in the context of GM food submission. Speaker: *Jennifer Holtzman, Ph.D., Health Canada*
14:30	**Belgian experience and insights** At the end of 2013, a workshop on the utility of NGS for characterisation of GE Organisms was jointly hosted by the Biosafety & Biotechnology Unit and the Biotechnology and Molecular Biology Platform, both units of the Scientific Institute of Public Health (WIV-ISP). Feedback, obtained from the workshop and the WIV-ISP analysis on the added value of NGS for the purpose of GE plants' molecular characterisation, is provided. Speaker: *Katia Pauwels, Ph.D., WIV-ISP, Belgium*
15:00	**Dutch experiences and insights** Experiences with use of RNA sequencing for the characterisation of GE plants. Speaker: *Esther J. Kok, Ph.D., RIKILT Wageningen UR*
15:30	Coffee break

15:45	**United States experiences and insights** Focus on how the U.S. Food and Drug Administration (US FDA)'s experience considering genome sequence data as part of consultations on the safety assessment of foods derived from GE plants. The US FDA has completed several consultations where genome sequence data derived from newer DNA sequencing methods has been considered. Management of these data in a regulatory dossier. Speaker: *Michael Kotewicz, Ph.D., U.S. Food and Drug Administration (FDA)*
16:15	**The use of next-generation sequencing data for regulatory submissions** Focus on Industry's experience with NGS tools for the characterisation of GE crops both in the product development and in the regulatory safety assessment stages. Speaker: *Jafar Mammadov, Ph.D., BIAC*
16:45	**GENERAL DISCUSSION AND CONCLUSION** The four previous speakers (item 3) engage in discussions on cross-cutting issues between their presentations, as well as other outstanding questions. Wrap-up session and conclusion by the workshop Chair. Moderator: *Workshop Chair*
18:00	**Closure of the workshop**
	Development of a short proceedings document Feedback from the workshop, including a summary of the presentations and discussions, will be briefly presented at the 23rd meeting of the OECD Task Force for the Safety of Novel Foods and Feeds, to be held on 19-21 April 2016. The workshop organising committee intends to publish a short OECD proceedings document describing the talks presented during the event. The paper will be for information only, and will not have conclusions or recommendations. Writers: *All presenters, and organising committee*

The workshop organisers and the OECD Secretariat extend their gratitude to the workshop speakers for sharing their expertise on this important topic.