Biotechnology Advances 35 (2017) 302–309

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

Molecular characterization of genetically-modified crops: Challenges and strategies

Rong Li [a,b,1], Sheng Quan [a,1], Xiaofang Yan [c], Sukumar Biswas [a], Dabing Zhang [a,b], Jianxin Shi [a,b,*]

[a] Joint International Research Laboratory of Metabolic & Developmental Sciences, Shanghai Jiao Tong University-University of Adelaide Joint Centre for Agriculture and Health, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China
[b] Key Laboratory of Crop Marker Assisted Breeding of Huaian Municipality, Jiangsu Collaborative Innovation Center of Regional Modern Agriculture and Environmental Protection, Huaiyin Normal University, Jiangsu 223300, China
[c] FuturaGene Biotechnology (Shanghai) Co. Ltd., Shanghai 200233, China

ARTICLE INFO

Article history:
Received 22 September 2016
Received in revised form 19 January 2017
Accepted 23 January 2017
Available online 26 January 2017

Keywords:
Cisgenesis
Genome editing
Next generation sequencing
T-DNA
Transgenesis

ABSTRACT

Molecular characterization lays a foundation for safety assessment and subsequent monitoring of genetically modified (GM) crops. Due to the target-specific nature, conventional polymerase chain reaction (PCR)-based methods cannot comprehensively detect unintended gene insertions, let alone unknown GM events. As more and more new developed GM crops including new plant breeding technology (NPBT) generated crops are in the pipeline for commercialization, alternative -omics approaches, particularly next generation sequencing, have been developed for molecular characterization of authorized or unauthorized GM (UGM) crops. This review summarizes first those methods, addresses their challenges, and discusses possible strategies for molecular characterization of engineered crops generated by NPBT, highlighting needs for a global information-sharing database and cost-effective, accurate and comprehensive molecular characterization approaches.

Contents

1. Introduction

Molecular characterization of genetically modified (GM) crops provides structural and expressional information of the insert(s) and stability information of the intended trait(s) (EFSA, 2011a, EFSA, 2012a). Molecular characteristics of GM crops generally include genomic features (such as insertion site, flanking sequence and copy number),

* Corresponding author at: School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China.
E-mail address: jianxin.shi@sjtu.edu.cn (J. Shi).
[1] Authors contribute equally to this work.

http://dx.doi.org/10.1016/j.biotechadv.2017.01.005
0734-9750/

transcriptomic features, proteomic features, and metabolomic features (Fig. 1), and information of these features is fundamental for the research & development, safety assessment, detection, and monitoring of GM crops. Based on molecular characteristics, both intended and unintended effects of transgene(s) can be readily identified (EFSA, 2012a, Schnell et al., 2015), facilitating significantly subsequent efforts for risk assessment of potential effects of GM crops and their derived products on food/feed quality and safety (Schnell et al., 2015).

The principle of substantial equivalent is one of the key principles of risk assessment for GM crops. To see if a GM crop is substantially equivalent to its non-GM recipient, various omics-based systems biology approaches are applied to compare their molecular characteristics (Heinemann et al., 2011). Most of current molecular characterization methods for GM crops are based on PCR based approaches (Fraiture et al., 2015b; Arulandhu et al.; 2016). Although transcriptomic (Coll et al., 2010; Li et al., 2016), proteomic (Gong and Wang, 2013), and metabolomics (Clarke et al., 2013; Simo et al., 2014) characterization have been applied to some GM crops, comprehensive system biology analyses at all levels (DNA, RNA, protein and metabolite) on a GM crop event are still very rare (Ricroch et al., 2011). Since more and more GM crops as well as novel engineered crops derived from new plant breeding technology (NPBT) are in the pipeline to be commercialized and released to environment and/or market in the near future (James, 2016; Parisi et al., 2016), it creates great challenges for molecular characterization. To provide detailed molecular characteristics of both authorized and unauthorized GM (UGM) crops to regulators, retailers, and consumers, development of accurate, comprehensive, and cost-effective molecular characterization methods are urgently needed. In the following we first address the technical aspects of molecular characterization of GM crops, focusing mainly on DNA based technologies; then we will highlight the main challenges and also discuss potential coping strategies.

2. Current methods for molecular characterization of GM crops

2.1. PCR-based methods for characterization of insertion sites and unknown flanking regions

Identifying the event-specific insertion sites upstream or downstream of an exogenous insertion in a GM crop provides direct evidence of the unknown flanking regions and conclusive evidence of the identity of the given GM crop. For this purpose, many PCR-based genome walking methods, such as thermal asymmetric interlaced PCR (TAIL-PCR) (Liu and Whittier, 1995), ligation mediated PCR (LM-PCR) (Mueller and Wold, 1991), and inverse PCR (IPCR) (Ochman et al., 1988), have been originally adopted. Later on, more and more modified or improved PCR-based methods, high throughput or high efficiency Tail PCR (Liu and Chen, 2007; Singer and Burke, 2003), and adaptor mediated PCR (Huang et al., 2007) were developed for routine molecular characterization with high throughput and/or high efficiency. These PCR-based methods all rely on known sequence information of the exogenous insertion, and no single method can be applied universally. Each method possesses its own characteristics (Table 1), and combinations of different methods are often adopted to increase efficiency (Yang et al., 2007). Because PCR-based methods cannot detect undocumented molecular characteristics of GM crops, in many cases, T-DNA insertion information is often underestimated (Yang et al., 2013a). Taking the commercialized soybean event GTS40-3-2 as an example, initially PCR-based methods identified only one inserted copy of the expression cassette of EPSPS gene (5-enolpyruvylshikimate 3-phosphate synthase) in the host's genome (Padgette et al., 1995). However, further studies revealed two additional unintended partial insertions of CP4 EPSPS (72- and 250-base pair, respectively) (Product Safety Center, 2000) and unintended DNA rearrangements at the 3′-NOS junction, causing the molecular characterization of GTS40-3-2 to be amended several times (Windels et al., 2001). Other examples involve GM insect resistant rice TT51-1 and T1c-19 that were developed in China. Initial PCR-based gene walking approaches identified one full insertion cassette each on the Chromosome 10 of TT51-1 (Cao et al., 2011) and the chromosome 11 of T1C-19 (Tang et al., 2006), respectively. Further whole genome next generation sequencing (NGS) approach revealed an additional full insertion cassette each on the chromosome 4 of TT51-1 and on the chromosome 4 of T1c-19, respectively (Yang et al., 2013a). Therefore, more attention should be paid to efforts making PCR-based methods also effective in characterizing unknown GM crops fully taking the advantage of bioinformatics tools and other related approaches.

2.2. Next generation sequencing based methods for comprehensive characterization of insertion sites and unknown flanking regions

As evidenced in several recent reviews, the combinations of abovementioned PCR based approaches with NGS appear to be more accurate and more comprehensive for molecular characterization of GM (or UGM) crops (Bodi et al., 2013; Mertes et al., 2011; Arulandhu et al., 2016). The process involves an initial enrichment of unknown adjacent sequences of known GM elements before NGS, using target-specific primers (not semi-random or random primers). Therefore, not all PCR-based methods can be effectively coupled with NGS to characterize GM and particularly UGM crops universally (Arulandhu et al., 2016). So far, none of the available enrichment methods can enrich long enough DNA fragments down- or up-stream of a known insert in a very sensitive manner, and none of them has been demonstrated to be effective in

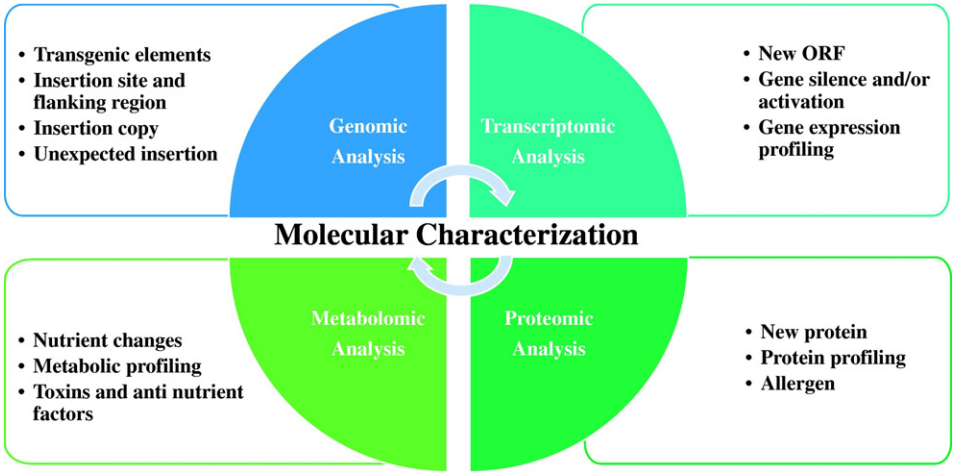

Fig. 1. Molecular characterization of genetically modified crops at genomic, transcriptomic, proteomic, and metabolomic levels.

Table 1
Different PCR based genome walking methods used for molecular characterization of GM crops.

Methods [ref.]	Principle or procedure	Applicability	Application	DNA treatment	Advantages	Disadvantages
Tail-PCR [Liu and Whittier, 1995]	TAIL-PCR uses nested long specific primers from known sequence with a high melting temperature combined with short arbitrary primers with a low melting temperature in three consecutive reactions under interspersing stringency PCR cycles	Most widely used but hardly work when endogenous promoter is used.	Rapeseed: RF1, RF2, MS1, T45, Topas 19/2, GT73 (Wu et al., 2014), Oxy-235 (Yang et al., 2008); maize: MON863 (Yang et al., 2005b), BVLA430101 (Rao et al., 2016b); rice: KeFeng-6 (Wang, 2011); wheat: B73–6-1 (Xu et al., 2013); cotton: BG2–7 (Zhang et al., 2016)	No	Quite efficient	Insufficient sensitivity and specificity; a low rate of positivity; cloned sequence size <1.0 kb; time consuming (4 days) and complex operation
LM-PCR [Mueller and Wold, 1991]	LM-PCR first digests DNA with blunt end restriction enzymes, and adds a linker cassette of known DNA sequence to the end of the unknown DNA fragment, finally amplifies selective DNA fragment using one primer binding to the link and another primer binding to the known portion of the DNA sequence	When only one exogenous end is known.	Maize: MON810 (Holck et al., 2002), MON863 (Zhu et al., 2008).	DNA digestion DNA ligation	Versatility for genomic typing and do not require prior knowledge of the sequence, long amplification product	Relatively expensive and time-consuming
SiteFinding-PCR [Tan et al., 2005]	SiteFinding-PCR starts with SiteFinder primer at a low temperature, followed by exponential amplification of target molecules with gene-specific and SiteFinder primers, and screening with another gene-specific and a vector primer	Especially for long flanking region amplification.	Rice: KMD1 (Babekova, 2009), MIR162 (Liang et al., 2014)	PCR product cloning and vector screening and sequencing	Simple, specific, sensitive, reliable, inexpensive, long specific product	Long procedure, time consuming
I-PCR [Ochman et al., 1988]	Inverse PCR (I-PCR) digests the DNA by restriction enzymes and cyclizes the resulting sticky end products to form a ring, and amplify this ring DNA using two primers that complementary to the known portion of the sequence and only differ in orientation	When only one exogenous end is known	Maize: LY038 (Zhang et al., 2011), Bt11 (Zimmermann et al., 2000); soybean: DP-356043-5 (356043) (Xu et al., 2011)	DNA digestion DNA ligation	Relative fast, reliable, amplified fragment can be as long as 4 kb	Complexed procedure, depends largely on chosen enzymes

dealing with complex processed food and feed samples. In addition, for reliable sequencing results, a high and even coverage of enriched sequences is required, which actually is quite difficult for some regions in the genome, particularly those rich in GC contents (Bodi et al., 2013). To solve these issues, whole genome sequencing approach was later introduced (Wahler et al., 2013; Yang et al., 2013a; Guttikonda et al., 2016; Guo et al., 2016; Holst-Jensen et al., 2016). The first reported approach using NGS to characterize and detect UGM crops is TranSeq (Yang et al., 2013a). TranSeq successfully characterizes two insect resistant GM rice events (TT51-1 and T1c-19) comprehensively combining paired-end re-sequencing and three bioinformatics modules. When using module 3, its identification of insertions and flanking regions does not depend on prior knowledge of the insert and vector sequences, demonstrating its high potential for the identification of UGM crops. Beside, TranSeq can be applied for the determination of insertion copies (Yang et al., 2013a). Another pilot study for rapid molecular characterization of GM crops based on NGS whole genome sequencing method was done in the same year using an EU un-authorized GM rice event LLRice62 (Wahler et al., 2013). Both studies prove that NGS-based approach could figure out single, multiple or complex insertions present in a DNA sample in one experiment with or without the prior information of insertion. In contrast, other whole genome sequencing approaches including genome shot gun sequencing rely on prior knowledge on the sequence information of insertions and host reference genome (Guttikonda et al., 2016, Guo et al., 2016; Holst-Jensen et al., 2016). Notably, the successful application of NGS for molecular characterization of GM crops depends largely on establishing successful bioinformatics pipelines and fit-for-purpose libraries, particularly in the case of UGM detection.

Other key issues for the feasibility of NGS in molecular characterization of GM crops include efficient cost, simplified data analysis pipeline, and high applicability to complex samples with GM or UGM crop ingredients at varying low concentrations. There has been no report regarding the detection limit of NGS for molecular characterization of GM crops so far, let alone any report about the standardization of NGS methods in the process. Although tremendous progresses have been made in the development of NGS technologies, different NGS platforms have different sequencing mechanisms and outputs (read lengths and read numbers) (Holst-Jensen et al., 2016), understanding of each NGS platform's advantages and shortcomings before applying it in the molecular characterization of GM crops is very important. Before NGS can fully come into the play, PCR-based methods will still be the dominant approaches for molecular characterization of GM crops.

2.3. DNA based methods for the determination of insertion copy number

Southern blot is traditionally used to identify the copy number of a gene or insertion in plant genome. It is also routinely used in molecular characterization of GM crops, and a successful Southern blot analysis involves proper designing of probe and careful selection of restriction digestion enzymes; both depend on prior sequence information of the gene or insertion. Southern blot is time consuming and laborious, and the result may not accurately reflect the presence of rearranged GM copies that lack the selected restriction enzyme sites (Pérez Urquiza and Acatzi Silva 2014; Yang et al., 2005a).

Real time quantitative PCR (qPCR) assay can accurately quantify the level of a GM crop by comparing with an endogenous reference gene, which provides a simplified, accurate alternative to Southern blot in determining gene copy number (Ingham et al., 2001; Yang et al., 2005a). Because of its simplicity, accuracy, robustness, and low cost, qPCR should be established as a standard assay for high-throughput GM crop number determination. Unfortunately, no validation has been

reported for its routine application in copy number determination although it's been widely tested in many GM events during the past decade. Comparing to its widely application in GM crop identification and quantification, it is still a long way before qPCR can be routinely used in GM copy number determination (Ingham et al., 2001; Yang et al., 2005a). Notably, qPCR is often targeted to only a small region in the gene sequence and sometimes fails to detect the presence of truncated or mutated copies of a GM insertion (Ingham et al., 2001). Currently, Southern blot and qPCR approaches are often used to complement or validate each other's results in GM crops' molecular characterization.

Digital PCR (dPCR) technology, an end-point DNA measurement, shows great potential as an alternative to qPCR to determine insert copy number in GM crops. It overcomes the dependency of qPCR on a DNA calibrant and identifies the absolute gene copy number in a sample. Comparing with qPCR, dPCR has enhanced productivity and increased reliability in copy number determination (Pérez Urquiza and Acatzi Silva 2014). Similarly to the case for qPCR, no dPCR platform has been validated or standardized for copy number determination despite that it has been applied in several GM crops (Dong et al., 2015; Felix-Urquidez et al., 2016). A recent comparative study of four dPCR platforms using a certified plasmid DNA reference material for accurate quantification of DNA copy number has highlighted the importance of full understanding of the measurement bias and uncertainty of each digital PCR platform before their application. Given its nature of absolute quantification, copy number values inferred from dPCR assays should be more reliable and accurate than those from qPCR which is still a relative assay of copy number. In the future, more efforts should be put into the standardization of dPCR method, as well as reducing the cost. A recent report of multiplex quantification of 12 EU authorized GM maize events using droplet dPCR has demonstrated its throughput and cost-effectiveness compared with qPCR (Dobnik et al., 2015), indicating a great potential of the application of droplet dPCR in determining GM copy number.

2.4. Other methods for insertion copy number determination

While it has been previously discussed that NGS can be applied to comprehensively characterize GM and UGM events, a recent publication has also reported that it could be effectively applied in estimating the insertion copy number of GM events (Yang et al., 2013a). However, this application is still in its early explorative stage and needs further investigations and validations.

2.5. Database of DNA based molecular characterization methods

Currently, there are many public available databases for GM detection methods, such as GMDD[1], GMO Crop Database[2], GMO COMPASS[3], Biodiv LMO[4], GMOfinder (Gerdes et al., 2012), GMOseek (Morisset et al., 2014), GMOMETHODS (Gerdes et al., 2012), JRC GMO-Amplicons (Petrillo et al., 2015), JRC GMO-Matrix (Angers-Loustau et al., 2014) and ISAAA GM Approval Database[5]. Not all of them include comprehensive molecular characterization information for approved GM crops, let alone information regarding UGM crops and patented GM crops in development pipeline. It is thus necessary to put a global joint effort in developing a database specifically dedicated to molecular characterization of GM crops.

2.6. Other methods based on various -omics

Compared with routine application of PCR based approaches in molecular characterization of GM crops, much less progress has been made in the past decades to characterize GM crops using other –omics tools. The major reason is not the lack of established platforms (Hall and de Maagd 2014) but rather a lack of understanding of the importance of applying such analyses in comply with substantial equivalent principle, as well as concerns for technical difficulties and high cost generally associated with other –omics tools (Gong and Wang 2013; Ricroch et al., 2011; Stewart et al., 2013; Yang et al., 2013b). Different from genomes, the changes of transcriptomes, proteomes, and metabolomes are affected by not only genetic factors (including genetic modification), but also various internal (such as developmental stages, tissue and cell types) and external (such as environment) factors. In many cases, the changes caused by factors rather than genetic modification could be larger than those caused by genetic modification alone (Zhou et al., 2009; Ricroch et al., 2011). Therefore, other –omics approaches except genomic need to be carefully designed to evaluate exact changes caused by GM insertions when applying them in molecular characterization.

Another important issue in interpreting transcriptomic, proteomic, and metabolomic studies of a GM crop is to bring in the concept of natural variation (Coll et al., 2010; Clarke et al., 2013; Yang et al., 2013b; Rao et al., 2016a). By the principle of substantial equivalent, if changes of one dimension factor in a GM crop are comparable to those of its non-GM counterpart (EFSA, 2011b), this GM crop can be regarded at least as safe as its non-GM counterpart. However, if changes in a GM crop are not comparable to those of its non-GM counterpart, do we rule this GM crop unsafe? The answer is clearly not a simple yes or no, because in any cases, natural variation of the considered factor(s) cannot be neglected. Previous -omics studies have indicated that natural variation can explain most of the profile changes between GM and non GM crops (Clarke et al., 2013; Coll et al., 2010; Harrigan et al., 2010; Rao et al., 2016a). For such large scale comparative –omics studies, in addition to choose correct comparators and make sure all crops are grown under the same conditions with the same developmental stages, natural variation has to be taken into consideration. By doing so, we can provide conclusive and convincing comparative results to properly understand the nature of the substantial equivalent principle.

Changes observed in -omics studies may also be brought upon from different platform settings of the same omics study. Various platforms used in transcriptomic, proteomic and metabolomic characterization of GM crops have been well reviewed elsewhere (Heinemann et al., 2011; Gong and Wang 2013; Ricroch et al., 2011; Holst-Jensen et al., 2016; Stewart et al., 2013; Garcia-Canas et al., 2011), albeit none of them has been validated for routine use. Due to the limitation of each single platform, combinations of platforms are often advised to apply for molecular characterization of GM crops, aiming to have better coverage of global alterations (both intended and unintended) of the expression patterns of RNA, protein and metabolites. While studies utilizing individual –omics approaches have been reported on different GM crops, simultaneous studies of a given GM crop at genomic, transcriptomic, proteomic, and metabolomic levels are still rare. To reduce the high cost for individual research institutions to conduct extensive and repetitive –omics studies, public research hubs or the third part service centers need to step in and collaborate with research institutions in the future to establish crop natural variation databases using data generated with well-established and validated -omics platforms.

3. Current challenges and coping strategies

3.1. Stacked GM crops

By 2015, GM crops approved by at least one country have increased to 363 GM events in 26 crops globally. Among them, stacked GM crops accounted for about 32.5% of all planted GM crops (James, 2016), and in US alone, stacked GM maize accounted for 77% of total planted GM maize (Chegao et al., 2015). The number of stacked genes in a given stacked GM crop has increased as well. An extreme example is SMARTSTAX CORN, a product of Monsanto and Dow AgroSciences, containing six insect resistance and two herbicide tolerance genes[6].

Stacked GM crops are produced from various processes including conventional breeding, multiple transformation, re-transformation, multigene introduction, and co-transformation (Taverniers et al.,

2008). Depending on what techniques have been utilized, the regulatory requirements for stacked GM crops differ significantly. For F1 generation or subsequent offspring from hybridization, molecular characterization focus on the integrity of the transgene site and potential interactions among stacked events, and they are treated the same as their parent events. For stacked crops produced through multiple transformation, re-transformation, multigene introduction, or co-transformation, multiple gene insertion sites may occur in the genome and they are treated as primary transformants for risk assessment purposes (EFSA, 2007). For stacked GM crops, PCR-based methods for single trait GM crops are not able to comprehensively reveal the complete molecular characteristics. Since generation of stacked GM crops are becoming an important trend in future GM crop development (James, 2016), there are currently increased needs for new molecular characterization approaches to perform unequivocal identification of stacked events in non-GM or GM seed lots. The complexity of stacked genes/traits has made it more challenging to develop suitable molecular characterization methods for stacked GM crops. It is urgently necessary to explore methods to characterize stacked GM crops without knowing the nature of stacked genes/traits.

3.2. UGM crops

UGM crops include GM crops that are approved but misused, approved in one jurisdiction but not in another (asynchronous authorizations), unapproved but released (intermingling), or unapproved but escaped from field trials (particularly in the stage of seed production). There have been increased reports of UGM incidents that tremendously affected both international trade and local economies (Holst-Jensen et al., 2012). The presence of UGM crops in the market highlights the need for international collaborations on synchronous authorization of a given GM crop, the need for development of communication channels to share information on GM crops under field trials, and the need for development of high throughout and cost effective screening methods to characterize and detect UGM crops. Considerable effort has been put into the development of suitable molecular characterization and detection methods for UGM crops, which have been reviewed in this and several other literatures (Fraiture et al., 2015a; Arulandhu et al., 2016; Holst-Jensen et al., 2016). When should we switch from the traceability strategy for authorized GM crops that relies heavily on claimed molecular features and target-specific characterization methods to a comprehensive screening strategy that depends largely on more general and broader characterization matrix assisted methods? Such a strategic shift will necessitate integrated well-developed basic methods with bioinformatics instead of developing and validating new characterization methods, and will focus more on synchronous authorization and development of data sharing databases which will not be easily accomplished.

3.3. GM crops generated by new technologies

Currently, many new plant breeding techniques (NPBT) including cisgenesis, intragenesis, near intragenesis, and genome editing have been applied as potential alternatives to transgenesis to generate engineered crops with various new traits. The applications of these new techniques have raised considerable concern and debate on their regulation (Lusser and Davies 2013; Broeders et al., 2012). In transgenesis the transferred T-DNA cassette including the coding sequence and regulatory elements is from other cross-incompatible species; in cisgenesis the T-DNA cassette is from crossable relative species or native species; in intragenesis the T-DNA cassette is *in vitro* rearranged with genetic elements from both host and cross-compatible donors; and in near intragenesis the T-DNA cassette consists mainly of host derived genetic elements with minimal recombinant elements (Holme et al., 2013; Singh et al., 2015).

Intragenesis and cisgenesis use the same gene pools available for traditional breeding, and no foreign genetic elements including selection markers and vector sequences are present in the end products (Fig. 2A–C). However, because of random integration, potentially interrupting and silencing the resident genes or other host sequences occur in intragenesis and cisgenesis. Additionally, it has been reported that microbial regulatory sequences and/or selectable marker genes are still found in some cisgenesis and intragenesis products (Holme et al., 2013, Jo et al., 2014). Therefore, both are currently regulated as transgenesis in EU (EFSA, 2012b).

Since it is supposed to be no exogenous inserted elements in cisgenic and intragenic crops, molecular characterization methods for transgenes cannot be used alone to characterize the genetic modification. At the genomic level, characteristics of the inserted elements (such as the orientation) and insertion sites could still be useful for DNA based characterization, the characteristics of typical event specific motifs that are associated with *Agrobacterium* mediated insertion sites are also helpful (Holst-Jensen et al., 2012). At transcriptomic, proteomic and metabolomic levels, if an engineered gene is involved in a particular pathway, the expression pattern of the targeted gene, its encoded protein, and final metabolic product in the pathway, can be characterized with various -omics technologies. The interpretation of these -omics data could still be difficult because concomitant changes of corresponding RNAs, proteins and metabolites do not always happen. The situation will get worse in the case where an engineered gene is not involved in a particular pathway. Therefore, combination of NGS with bioinformatics analysis may be the best solution to characterize cisgenic and intragenic crops as they have been demonstrated in the characterization of UGM crops (Yang et al., 2013a). In near intragenic crops, the minimal vector backbone characteristics could be important for PCR based characterization as well as for the initial sequence enrichment in NGS approach. Nevertheless, in the case of cisgenic and intragenic crops, we still lack effective DNA or NGS based methods to characterize them.

Genome editing is a highly specific and efficient tool with a great potential to generate new improved crops. It is based on site-directed mutagenesis mediated by engineered nucleases. The most commonly used engineered nucleases include meganuclease, zinc finger nucleases (ZFN), transcription activator-like effector nuclease (TALEN), and clustered regular interspaced short palindromic repeats/CRISPR-associated nuclease (CRISPR/Cas) (Araki et al., 2014, Cardi & Varshney, 2016, Jones, 2015). In principle, an engineered nuclease initially creates a double strand break (DSB) in the target sequence, and the DSB can be repaired inaccurately *in vitro* through two mechanisms: an error prone process named non-homologous end joining (NHEJ) that often generates indels (insertions or deletions), and a homologous recombination process that results in replacement/insertion or deletion depends on the presence or absence of a donor DNA, respectively (Pauwels et al., 2014) (Fig. 2D). While an engineered nuclease gene is stably integrated into the host genome initially, it will be lost eventually in the next generation because of random assortment and chromosome segregation (the intended edit remains) (Jones, 2015). Therefore, crops generated by genome editing are virtually indistinguishable from those generated by conventional mutagenesis or natural mutation. Genome editing occurred within a few nucleotides is difficult to be distinguished from natural mutation. Compared with the random insertion of foreign genetic materials in transgenesis, targeted NPBT minimizes the probability of position effects and unintended gene disruption. Therefore, there are debates about whether or not such crops should be considered as transgenic crops or if less stringent safety assessment requirements could be applied to them. When genome editing is applied to insert DNA fragments or whole genes *via* homologous recombination into the host genome at pre-determined locations, off-target mutagenesis and induced monoallelic modification could still occur (Araki et al., 2014; Jones, 2015; Pauwels et al., 2014; Wolt et al., 2016). Therefore, risks induced by NPBT cannot be completely ruled out, and crops generated by genome editing still need screening and characterization,

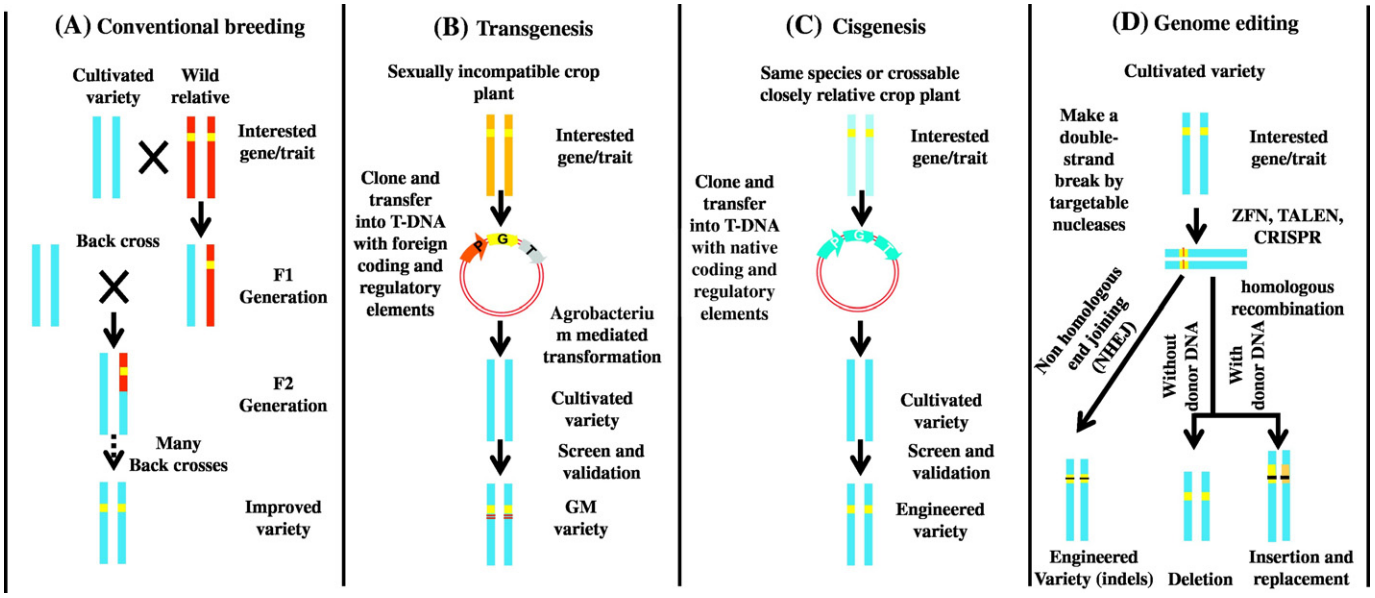

Fig. 2. Illustration of different principles and procedures for engineering crops. Blue bar: cultivated variety chromosome; red bar: wild relative crop chromosome; orange bars: sexual incompatible crop chromosomal; light blue bar: native or cross compatible organism chromosome; yellow small cube: interested gene; double dark red lines: vector or marker gene sequence. P: promoter; G: engineered gene; T: terminator.

focusing on both the confirmation of intended effects and the discovery of potential off-target modifications (Araki et al., 2014). Due to the economic, social, and academic importance of the application of NPBT in crop breeding, worldwide consensus and actions on regulatory issues on NPBT-generated crops are urgently needed. While EU treats gene-edited crops as transgenic, at least two gene-edited crops have been deregulated in US (EFSA, 2012b, Jones, 2015, EFSA, 2012c).

Genome editing produces crops with indels in a targeted gene, involving changes of deletion or insertion of a few nucleotide bases only, and making it difficult to distinguish them from natural variation. Currently, we do not know the exact frequency of genome editing to produce insertion or replacement in the targeted gene *via* homologous recombination. Therefore, in the case of characterizing crops with unknown genome editing, we are not sure if single nucleotide polymorphism information could be helpful. Notably, genome editing acts on the whole gene sequence, including promoter sequence. If off-target genome editing occurs, it would affect the host both genetically and epigenetically, which merits further investigation.

4. Concluding remarks

Current regulations on crops safety put great emphasis on the regulation of transgenic crops because of concerns such as co-introduction of selectable markers (particularly antibiotic resistance genes) and microorganism originated DNA sequences, or unintended effects caused by random insertion. NPBT offers new opportunities to transfer genetic materials from closely related species or native genetic pools into a host by precise site-directed mutagenesis, and thus eliminate the introduction of foreign genetic elements. Because of this, NPBT may accelerate crop breeding processes since it relieves to some extent the worries from general public and regulators about undesirable effects caused by random gene insertions and disruptions associated with conventional transgenic crops. Despite of the apparent advantages of NPBT over conventional transgenesis, we still cannot rule out the likelihood of random occurrences of unintended effects associated with NBPT, and sufficient risk assessment of crops generated by NBPT is still necessary at the current stage.

Molecular characterizations of GM crops should include different analyses on DNA, RNA, protein and metabolite levels. Previously, most efforts have been put into the development of DNA based methods, among which PCR based methods have played indispensable roles. Nowadays, even when NGS based technologies have shown great promises to be a potential alternative, the conventional PCR based DNA analysis are still the workhorse for routine molecular characterization of GM crops. NGS can offer accurate and comprehensive view of the structural information of GM crops, particularly in the cases of UGM and stacked GM crops (Yang et al., 2013a, Arulandhu et al., 2016). However, NGS associated higher cost and technical difficulties make it unlikely that NGS would be applicable in large scale to replace PCR based methods anytime soon. Currently it is not certain if NGS can be applied to mixed samples with stacked traits of varying concentrations and what is its limit of detection (LOD), let alone its quantitative use. Future molecular characterizations of GM crops must combine genomic analysis with other analyses targeting RNAs, proteins, and metabolites. Additional efforts should be directed into establishment of information-sharing databases containing molecular characterization data for GM crops that have been approved, under review, or in field trials. Efforts should also be spent to develop suitable characterization and detection methods for UGM crops, stacked GM crops and NPBT generated crops.

Acknowledgements

This work was supported by grants from the China National Transgenic Plant Special Fund (2016ZX08012-002), the EU Seventh Framework Programme project DECATHLON (613908) and the Programme of Introducing Talents of Discipline to Universities (111 Project, B14016).

References

Angers-Loustau, A., et al., 2014. JRC GMO-matrix: a web application to support genetically modified organisms detection strategies. BMC Bioinf. 15, 417.
Araki, M., et al., 2014. Caution required for handling genome editing technology. Trends Biotechnol. 32, 234–237.
Arulandhu, A.J., et al., 2016. DNA enrichment approaches to identify unauthorized genetically modified organisms (GMOs). Anal. Bioanal. Chem. 408, 4575–4593.
Babekova, R., 2009. Development of an event-specific real-time PCR detection method for the transgenic Bt rice line KMD1. Eur. Food Res. Technol. 228, 707–716.
Bodi, K., et al., 2013. Comparison of commercially available target enrichment methods for next-generation sequencing. J. Biomol. Tech. 24, 73–86.
Broeders, S.R., et al., 2012. How to deal with the upcoming challenges in GMO detection in food and feed. J. Biomed. Biotechnol. 2012, 402418.

Cao, Y., et al., 2011. Characterization of the transgenic rice event TT51-1 and construction of a reference plasmid. J. Agric. Food Chem. 59, 8550–8559.
Cardi, T., Varshney, R., 2016. Cisgenesis and genome editing: combining concepts and efforts for a smarter use of genetic resources in crop breeding. Plant Breed. 135, 139–147.
Chegao, A., et al., 2015. Low level presence vs. coexistence vestigial presence of stacked-transgenic events and consequences of unintentional releases into crop fields. The Seventh International Conference on Coexistence Between Genetically Modified (GM) and Non-GM Based Agricultural Supply Chains (GMCC-15). Amsterdam, The Netherlands, November 17–20, 2015.
Clarke, J.D., et al., 2013. Assessment of genetically modified soybean in relation to natural variation in the soybean seed metabolome. Sci. Rep. 3, 3082.
Coll, A., et al., 2010. Natural variation explains most transcriptomic changes among maize plants of MON810 and comparable non-GM varieties subjected to two N-fertilization farming practices. Plant Mol. Biol. 73, 349–362.
Dobnik, D., et al., 2015. Multiplex quantification of 12 European Union authorized genetically modified maize lines with droplet digital polymerase chain reaction. Anal. Chem. 87, 8218–8226.
Dong, L., et al., 2015. Comparison of four digital PCR platforms for accurate quantification of DNA copy number of a certified plasmid DNA reference material. Sci. Rep. 5, 13174.
EFSA, 2007. Guidance document of the scientific panel on genetically modified organisms for the risk assessment of genetically modified plants containing stacked transformation events. EFSA J. 512, 1–5.
EFSA, 2011a. Guidance for risk assessment of food and feed from genetically modified plants. EFSA J. 9, 2150.
EFSA, 2011b. Guidance on selection of comparators for the risk assessment of genetically modified plants and derived food and feed. EFSA J. 9, 2149.
EFSA, 2012a. Guidance on the risk assessment of food and feed from genetically modified animals and on animal health and welfare aspects. EFSA J. 10, 2501.
EFSA, 2012b. Scientific opinion addressing the safety assessment of plants developed through cisgenesis and intragenesis. EFSA J. 10, 2561.
EFSA, 2012c. Scientific opinion addressing the safety assessment of plants developed using zinc finger nuclease 3 and other site-directed nucleases with similar function. EFSA J. 10, 2943.
Felix-Urquidez, D., et al., 2016. Development, optimization, and evaluation of a duplex droplet digital PCR assay to quantify the T-nos/hmg copy number ratio in genetically modified maize. Anal. Chem. 88, 812–819.
Fraiture, M.A., et al., 2015a. Validation of a sensitive DNA walking strategy to characterise unauthorised GMOs using model food matrices mimicking common rice products. Food Chem. 173, 1259–1265.
Fraiture, M.A., et al., 2015b. Current and new approaches in GMO detection: challenges and solutions. Biomed. Res. Int. 392872.
Garcia-Canas, V., et al., 2011. MS-based analytical methodologies to characterize genetically modified crops. Mass Spectrom. Rev. 30, 396–416.
Gerdes, L., et al., 2012. GMOfinder—A GMO screening database. Food Anal. Methods 5, 1368–1376.
Gong, C.Y., Wang, T., 2013. Proteomic evaluation of genetically modified crops: current status and challenges. Front. Plant Sci. 4, 41.
Guo, B., et al., 2016. Identification of genomic insertion and flanking sequence of G2-EPSPS and GAT transgenes in soybean using whole genome sequencing method. Front. Plant Sci. 7, 1009.
Guttikonda, S.K., et al., 2016. Molecular characterization of transgenic events using next generation sequencing approach. PLoS One 11, e0149515.
Hall, R.D., De Maagd, R.A., 2014. Plant metabolomics is not ripe for environmental risk assessment. Trends Biotechnol. 32, 391–392.
Harrigan, G.G.L.D., et al., 2010. Natural variation in crop composition and the impact of transgenesis. Nat. Biotechnol. 28, 402–404.
Heinemann, J.A., et al., 2011. Molecular profiling–a tool for addressing emerging gaps in the comparative risk assessment of GMOs. Environ. Int. 37, 1285–2193.
Holck, A., et al., 2002. 5′-Nuclease PCR for quantitative event-specific detection of the genetically modified Mon810 MaisGard maize. Eur. Food Res. Technol. 214, 449–454.
Holme, I.B., et al., 2013. Intragenesis and cisgenesis as alternatives to transgenic crop development. Plant Biotechnol. J. 11, 395–407.
Holst-Jensen, A., et al., 2012. Detecting un-authorized genetically modified organisms (GMOs) and derived materials. Biotechnol. Adv. 30, 1318–1335.
Holst-Jensen, A., et al., 2016. Application of whole genome shotgun sequencing for detection and characterization of genetically modified organisms and derived products. Anal. Bioanal. Chem. 408, 4595–4614.
Huang, S.C., et al., 2007. An improved method to identify the T-DNA insertion site in transgenic *Arabidopsis thaliana* genome. Russ. J. Plant Physiol. 54, 822–826.
Ingham, D.J., et al., 2001. Quantitative real-time PCR assay for determining transgene copy number in transformed plants. BioTechniques 31, 132–140.
James, C., 2016. 20th Anniversary of the Global Commercialization of Biotech Crops (1996 to 2015) and Biotech Crop Highlights in 2015. ISAAA Brief 51.
Jo, K.R., et al., 2014. Development of late blight resistant potatoes by cisgene stacking. BMC Biotechnol. 14, 50.
Jones, H.D., 2015. Regulatory uncertainty over genome editing. Nat. Plants 1, 14011.
Li, Y., et al., 2016. Successful detection of foreign inserts in transgenic rice TT51-1 (BT63) by RNA-sequencing combined with PCR. J. Sci. Food Agric. http://dx.doi.org/10.1002/jsfa.7913.
Liang, C., et al., 2014. Detecting authorized and unauthorized genetically modified organisms containing vip3A by real-time PCR and next-generation sequencing. Anal. Bioanal. Chem. 406, 2603–2611.
Liu, Y.G., Chen, Y., 2007. High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences. BioTechniques 43, 649–656.
Liu, Y.G., Whittier, R.F., 1995. Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics 25, 674–681.
Lusser, M., Davies, H.V., 2013. Comparative regulatory approaches for groups of new plant breeding techniques. Nat. Biotechnol. 30, 437–446.
Mertes, F., et al., 2011. Targeted enrichment of genomic DNA regions for next-generation sequencing. Brief. Funct. Genomics 10, 374–386.
Morisset, D., et al., 2014. GMOseek: a user friendly tool for optimized GMO testing. BMC Bioinf. 15.
Mueller, P.R., Wold, B., 1991. Ligation-mediated PCR: applications to genomic footprinting. Methods 2, 20–31.
Ochman, H., et al., 1988. Genetic applications of an inverse polymerase chain reaction. Genetics 120, 621–623.
Pérez Urquiza, M., Acatzi Silva, A.I., 2014. Copy number ratios determined by two digital polymerase chain reaction systems in genetically modified grains. Metrologia 51, 61–66.
Padgette, S.R., et al., 1995. Development, identification, and characterization of a glyphosate-tolerant soybean line. Crop Sci. 35, 1451–1461.
Parisi, C., et al., 2016. The global pipeline of GM crops out to 2020. Nat. Biotechnol. 34, 31–36.
Pauwels, K., et al., 2014. Engineering nucleases for gene targeting: safety and regulatory considerations. Nat. Biotechnol. 31, 18–27.
Petrillo, M., et al., 2015. JRC GMO-Amplicons: a collection of nucleic acid sequences related to genetically modified organisms. Database (Oxford) 2015.
Product Safety Center, 2000. Updated molecular characterization and safety assessment of Roundup Ready® Soybean event 40-3-2. Monsanto Company Confidential Report pp. 1–20.
Rao, J., et al., 2016a. Metabolic changes in transgenic maize mature seeds over-expressing the *Aspergillus niger* phyA2. Plant Cell Rep. 35, 429–437.
Rao, J., et al., 2016b. Development of event-specific qualitative and quantitative PCR detection methods for the transgenic maize BVLA430101. Eur. Food Res. Technol. 242, 1277–1284.
Ricroch, A.E., et al., 2011. Evaluation of genetically engineered crops using transcriptomic, proteomic, and metabolomic profiling techniques. Plant Physiol. 155, 1752–1761.
Schnell, J., et al., 2015. A comparative analysis of insertional effects in genetically engineered plants: considerations for pre-market assessments. Transgenic Res. 24, 1–17.
Simo, C., et al., 2014. Metabolomics of genetically modified crops. Int. J. Mol. Sci. 15, 18941–18966.
Singer, T., Burke, E., 2003. High through TAIL-PCR as a tool to identify DNA flanking insertions. In: Grotewold, E. (Ed.), Plant Functional Genomics II Methods in Molecular Biology. Humana Press Inc., Totowa, NJ.
Singh, A., et al., 2015. Alternative to transgenesis: cisgenesis and intragenesis. Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools. Springer International Publishing, pp. 345–367.
Stewart, D., et al., 2013. Metabolomics for the safety assessment of genetically modified (GM) crops. Metabolomics in Food and Nutrition, pp. 192–216.
Tan, G., et al., 2005. SiteFinding-PCR: a simple and efficient PCR method for chromosome walking. Nucleic Acids Res. 33, 13.
Tang, W., et al., 2006. Development of insect-resistant transgenic indica rice with a synthetic cry1C* gene. Mol. Breed. 18, 1–10.
Taverniers, I., et al., 2008. Gene stacking in transgenic plants: towards compliance between definitions, terminology, and detection within the EU regulatory framework. Environ. Biosaf. Res. 7, 197–218.
Wahler, D., et al., 2013. 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. Food Anal. Methods 6, 1718–1727.
Wang, W., 2011. Event-specific qualitative and quantitative detection of transgenic rice Kefeng-6 by characterization of the transgene flanking sequence. Eur. Food Res. Technol. 232, 297–305.
Windels, P., et al., 2001. Characterisation of the roundup ready soybean insert. Eur. Food Res. Technol. 213, 107–112.
Wolt, J.D., et al., 2016. The regulatory status of genome-edited crops. Plant Biotechnol. J. 14, 510–518.
Wu, Y., et al., 2014. Characterization of genomic integration and transgene organization in six transgenic rapeseed events. J. Integr. Agric. 13, 1865–1876.
Xu, J., et al., 2013. Flanking sequence determination and event-specific detection of genetically modified wheat B73-6-1. Acta Biochim. Biophys. Sin. 45, 416–421.
Xu, W., et al., 2011. Establishment and evaluation of event-specific qualitative and quantitative PCR method for genetically modified soybean DP-356043-5. Eur. Food Res. Technol. 233, 685–695.
Yang, L., et al., 2013a. Characterization of GM events by insert knowledge adapted re-sequencing approaches. Sci. Rep. 3, 2839.
Yang, L., et al., 2008. Qualitative and quantitative event-specific PCR detection methods for oxy-235 canola based on the 3′ integration flanking sequence. J. Agric. Food Chem. 56, 1804–1809.
Yang, L., et al., 2005a. Estimating the copy number of transgenes in transformed rice by real-time quantitative PCR. Plant Cell Rep. 23, 759–763.
Yang, L., et al., 2005b. Event specific qualitative and quantitative polymerase chain reaction detection of genetically modified MON863 maize based on the 5′-transgene integration sequence. J. Agric. Food Chem. 53, 9312–9318.
Yang, R., et al., 2007. Event-specific qualitative and quantitative PCR detection of roundup ready event GT73 based on the 3′-integration junction. Plant Cell Rep. 26, 1821–1831.
Yang, X.S., et al., 2013b. Omics technologies reveal abundant natural variation in metabolites and transcripts among conventional maize hybrids. Food Nutr. 4, 335–341.

Zhang, N., et al., 2011. Event-specific qualitative and quantitative PCR detection of LY038 maize in mixed samples. Food Control 22, 1287–1295.
Zhang, X., et al., 2016. Structure of exogenous gene integration and event-specific detection in the glyphosate-tolerant transgenic cotton line BG2-7. PLoS One 11, e0158384.
Zhou, J., et al., 2009. Metabolic profiling of transgenic rice with cryIAc and sck genes: an evaluation of unintended effects at metabolic level by using GC-FID and GC–MS. J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 877, 725–732.
Zhu, H., et al., 2008. A specific qualitative and real-time PCR detection of MON863 maize based on the 5′-transgene integration sequence. J. Cereal Sci. 48, 592–597.
Zimmermann, A., et al., 2000. Event specific transgene detection in Bt11 corn by quantitative PCR at the integration site. LWT Food Sci. Technol. 33, 210–216.
http://gmdd.shgmo.org/.
http://cera-gmc.org/GMCropDatabase/.
http://www.gmo-compass.org/eng/home/.
http://bch.cbd.int/database/lmo-registry/.
http://www.isaaa.org/gmapprovaldatabase/cropslist/.
http://www.agbios.com/dbase.php.