

# **Evidence synthesis on the impact of genome editing on plant breeding**

Dissertation

zur Erlangung des Doktorgrades  
der Fakultät der Agrarwissenschaften  
der Georg-August-Universität Göttingen

vorgelegt von

**Dominik Modrzejewski**

geboren in Ludwigsburg, Deutschland

Göttingen, 05/2020

D 7

1. Gutachter: **Herr Prof. Dr. Achim Spiller**  
Georg-August-Universität Göttingen  
Professor für Marketing für Lebensmittel und Agrarprodukte
  
2. Gutachter: **Herr Dr. Ralf Wilhelm**  
Julius Kühn-Institut, Bundesforschungsinstitut für Kulturpflanzen  
Instituts für die Sicherheit biotechnologischer Verfahren bei Pflanzen

Weitere Mitglieder der Prüfungskommission:

**Frau Prof. Dr. Anne-Katrin Mahlein**  
Institut für Zuckerrübenforschung (IfZ)  
An-Institut der Georg-August-Universität Göttingen

**Herr Prof. Dr. Stefan Scholten**  
Georg-August-Universität Göttingen  
Department für Nutzpflanzenwissenschaften & Zentrum für  
Integrierte Züchtungsforschung

Tag der mündlichen Prüfung: 15.07.2020

Meinen Eltern und Elle in Dankbarkeit gewidmet



# Contents

|   |            |
|---|------------|
| <b>ABSTRACT .....</b>   | <b>IV</b>  |
| <b>KURZFASSUNG .....</b>  | <b>V</b>   |
| <b>PUBLICATIONS INCLUDED IN THIS THESIS .....</b>   | <b>VI</b>  |
| <b>LIST OF ABBREVIATIONS .....</b>  | <b>VII</b> |
| <b>1 INTRODUCTION.....</b>  | <b>1</b>   |
| 1.1 GENOME EDITING .....  | 2          |
| 1.1.1 Site-directed nucleases .....   | 2          |
| 1.1.2 Oligonucleotide-Directed Mutagenesis .....  | 6          |
| 1.1.3 Base editing.....   | 6          |
| 1.1.4 Comparison of genome editing and conventional breeding techniques .....                                     | 7          |
| 1.2 OFF-TARGET EFFECTS .....  | 7          |
| 1.2.1 Reasons for the occurrence of off-target effects.....   | 7          |
| 1.2.2 Possibilities to detect off-target effects .....  | 8          |
| 1.2.3 Off-target effects in the context of plant breeding .....   | 9          |
| 1.3 EVIDENCE SYNTHESIS.....   | 9          |
| 1.3.1 Systematic review.....  | 9          |
| 1.3.2 Systematic map .....  | 11         |
| 1.3.3 Disadvantages and challenges of systematic reviews.....   | 12         |
| 1.3.4 CADIMA - an online tool to support the conduct and reporting of systematic reviews and systematic maps..... | 12         |
| <b>2 THE ELSA-GEA PROJECT.....</b>  | <b>14</b>  |
| <b>3 OBJECTIVES OF THIS THESIS .....</b>  | <b>15</b>  |
| <b>4 SYSTEMATIC MAP PROTOCOL.....</b>   | <b>16</b>  |
| 4.1 ABSTRACT .....  | 17         |
| 4.2 BACKGROUND.....   | 17         |
| 4.3 OBJECTIVES OF THE MAP .....   | 21         |
| 4.4 METHODS.....  | 23         |
| 4.4.1 Search strategy .....   | 23         |
| 4.4.2 Study inclusion criteria.....   | 24         |

|          |   |           |
|----------|---|-----------|
| 4.4.3    | Article screening.....  | 24        |
| 4.4.4    | Study quality assessment.....   | 25        |
| 4.4.5    | Data coding strategy.....   | 25        |
| 4.4.6    | Study mapping and presentation.....   | 25        |
| 4.5      | REFERENCES.....   | 26        |
| <b>5</b> | <b>SYSTEMATIC MAP .....</b>   | <b>30</b> |
| 5.1      | ABSTRACT.....   | 31        |
| 5.2      | BACKGROUND.....   | 32        |
| 5.3      | OBJECTIVES OF THE MAP .....   | 35        |
| 5.4      | METHODS.....  | 37        |
| 5.5      | RESULTS.....  | 40        |
| 5.5.1    | General overview of the application of genome editing in model plants and crops.....  | 41        |
| 5.5.2    | Secondary question 1: “What are the traits modified by genome editing in model plants as well as in crops produced for agricultural production?” ...  | 46        |
| 5.5.3    | Secondary question 2: “What is the available evidence for the potential occurrence of associated off-target effects due to the use of genome editing in model plants as well as in crops produced for agricultural production?” ..... | 57        |
| 5.6      | DISCUSSION .....  | 65        |
| 5.6.1    | Market-oriented applications of genome editing.....   | 65        |
| 5.6.2    | Off-target effects of genome editing.....   | 66        |
| 5.6.3    | Knowledge gaps.....   | 66        |
| 5.6.4    | Knowledge clusters .....  | 67        |
| 5.6.5    | Limitations of the systematic map.....  | 67        |
| 5.7      | CONCLUSIONS .....   | 68        |
| 5.7.1    | Implications for policy/management .....  | 68        |
| 5.7.2    | Implication for research .....  | 68        |
| 5.8      | REFERENCES.....   | 68        |
| <b>6</b> | <b>SYSTEMATIC REVIEW.....</b>   | <b>82</b> |
| 6.1      | ABSTRACT.....   | 83        |
| 6.2      | BACKGROUND.....   | 83        |
| 6.3      | OBJECTIVES OF THE REVIEW .....  | 87        |

|          |  |            |
|----------|--|------------|
| 6.4      | METHODS.....   | 87         |
| 6.5      | RESULTS.....   | 94         |
| 6.5.1    | Review descriptive statistics.....   | 94         |
| 6.5.2    | Meta-analysis of <i>high</i> and <i>medium/high validity</i> studies.....                                  | 98         |
| 6.5.3    | Descriptive synthesis of <i>high</i> and <i>medium/high validity</i> studies .....                         | 101        |
| 6.5.3.1  | Hypothesis 1: Number of mismatches .....   | 101        |
| 6.5.3.2  | Hypothesis 2: Position of mismatches .....   | 102        |
| 6.5.3.3  | Hypothesis 3: GC-content.....  | 104        |
| 6.5.3.4  | Hypothesis 4: Nuclease variant .....   | 107        |
| 6.5.3.5  | Hypothesis 5: Delivery method .....  | 109        |
| 6.5.4    | Descriptive synthesis of <i>low</i> and <i>medium/low</i> as well as <i>unclear validity</i> studies ..... | 112        |
| 6.6      | DISCUSSION .....   | 113        |
| 6.6.1    | Factors affecting off-target effects caused by the use of CRISPR/Cas in plants.....                        | 113        |
| 6.6.2    | Reasons for heterogeneity .....  | 114        |
| 6.6.3    | Review limitations.....  | 115        |
| 6.7      | CONCLUSIONS .....  | 116        |
| 6.7.1    | Implication for policy/management .....  | 116        |
| 6.7.2    | Implication for research .....   | 116        |
| 6.8      | REFERENCES.....  | 117        |
| <b>7</b> | <b>SUMMARY AND OUTLOOK.....</b>  | <b>123</b> |
| 7.1      | EVIDENCE SYNTHESIS TOOLS FOR EVIDENCE-BASED DECISION-MAKING.....   | 123        |
| 7.2      | SUMMARY AND OUTLOOK OF THE RESULTS OF THE SYSTEMATIC MAP AND THE SYSTEMATIC REVIEW .....                   | 125        |
| 7.2.1    | Market-oriented applications of genome editing .....   | 126        |
| 7.2.2    | Occurrence of off-target effects .....   | 127        |
| 7.3      | REFERENCES (CHAPTER 1, CHAPTER 7).....   | 128        |
|          | <b>DANKSAGUNG.....</b>   | <b>IX</b>  |
|          | <b>ERKLÄRUNG.....</b>  | <b>X</b>   |

## Abstract

Genome editing techniques such as CRISPR/Cas, TALENs, Zinc-Finger Nucleases, Meganucleases, Oligonucleotide-Directed Mutagenesis and Base editing enable a targeted modification of DNA sequences in a site-directed manner. Despite their simple, time-saving and cost-effective use, there is an ongoing debate about their precise targeting and the extent to which off-target effects occur and matter. Off-target effects are unintended cleavage and mutations at untargeted genomic sites showing similar but not identical sequences compared to the target one. The number of research articles about genome editing in plants is increasing rapidly and evidence synthesis of these articles play an important role in controlling the rapidly expanding evidence base and contributing to evidence-based decision-making. The overall objective of this study was to systematically synthesize and evaluate the evidence on the impact of genome editing in plants. For this purpose, a multi stage conceptual approach was chosen. In a first step, a protocol was developed in order to make the review process more rigorous, transparent and well-defined as well as to minimize reviewer bias. Based on the protocol a systematic map was conducted aiming to provide a comprehensive overview on the applications of genome editing and the potential occurrence of associated off-target effects. The main findings of this map are that although most of the studies are categorized as basic research, nearly 100 market-oriented applications were identified in 28 different crops leading to plants with improved food and feed quality, agronomic value like growth characteristics or increased yield, tolerance to biotic and abiotic stress, herbicide tolerance or industrial benefits. Results further indicate that the large majority of off-target studies were conducted using CRISPR/Cas. Biased detection methods are the dominant approach to detect off-target effects, meaning that potential off-target sequences are first predicted using bioinformatics programs followed by targeted sequencing of these sites to track the occurrence of off-target effects. Off-target effects were detected in several studies. However, studies investigating potential off-target sequences are very heterogeneous in their structure and design. Based on these findings, the objective of the follow up systematic review was to identify and systematically analyze factors that may affect the occurrence of off-target effects. Results of the analysis clearly indicate that an increased number of mismatches between the on-target and potential off-target sequence significantly decreases the occurrence of off-target effects. The observed off-target rate decreases from 59% when there was one mismatch between the on-target and potential off-target sequence towards 0% when at least four mismatches exist. In addition, mismatch/es located within the first eight nucleotides proximal to the PAM significantly decrease the occurrence of off-target effects and there is a tendency that adjoined mismatches increase the likelihood of off-target effects. There was no clear indication that the occurrence of off-target effects shows an obvious trend in relation to the GC-content of the protospacer sequence. Regarding the nuclease variant and the delivery method the given data base is considerably poor, as the large majority of studies applied the standard nuclease SpCas9 and the CRISPR/Cas system was stably delivered in the genome. Only a limited number of studies applied altered nuclease variants and delivery methods. Hence, a general significant impact of the nuclease variants and the delivery method on the occurrence of off-target effects cannot be proved. Based on the results of this study, in order to minimize off-target effects, it is recommended to select a target sequence which differs in at least four disjunct positions from similar genomic sequences.

# Kurzfassung

Genome Editing ist ein Sammelbegriff für mehrere molekularbiologische Techniken, mit denen gezielt Veränderungen in ganz bestimmten Abschnitten der DNA herbeigeführt werden können. Der große Vorteil von Genome Editing liegt in der einfachen, zeitsparenden und kostengünstigen Anwendung verglichen zu konventionellen Pflanzenzüchtungstechniken. Trotzdem wird in der Landwirtschaftsbranche sehr kontrovers über dessen Nutzen diskutiert und auch darüber, inwieweit das Auftreten von unbeabsichtigten off-target Effekten relevant ist. Off-target Effekte können auftreten, wenn der DNA Strang an unbeabsichtigten Stellen im Genom geschnitten wird, die ähnlich, aber nicht identisch zur Zielsequenz sind. Das Ziel dieser Arbeit bestand in einer systematischen Literaturrecherche über mögliche Auswirkungen des Genome Editing in Pflanzen. Dazu wurde ein mehrstufiger konzeptioneller Ansatz gewählt. In einem ersten Schritt wurde ein Protokoll entwickelt, um einen stringenten, transparenten und klar definierten Review Prozess zu gewährleisten. Auf dessen Grundlage wurde eine Systematic Map erstellt, um einen umfassenden Überblick der zur Verfügung stehenden Literatur über mögliche Anwendungen des Genome Editing und das Auftreten von off-target-Effekten zu bekommen. Die Ergebnisse der Map zeigen, dass, obwohl die meisten Studien als Grundlagenforschung einzuordnen sind, bereits annähernd 100 marktorientierte Anwendungen in 28 verschiedenen Kulturen publiziert wurden. Diese Pflanzen zeigten verbesserte Nahrungs- und Futtermittelqualitäten, verbesserte Wachstumseigenschaften oder Ertragssteigerungen, erhöhte Toleranz gegen biotischen und abiotischen Stress, Herbizidtoleranzen oder einen erhöhten industriellen Nutzen. Des Weiteren zeigen die Ergebnisse, dass die meisten off-target Studien mit der CRISPR/Cas Technik durchgeführt wurden. Der vorherrschende Ansatz zur Identifizierung von off-target Effekten war die „biased detection method“, bei der lediglich Stellen im Genom, die ähnlich zur Zielsequenz sind, gezielt auf ungewollte Mutationen hin untersucht werden. Mehrere Studien wurden identifiziert, in denen off-target Effekte nachgewiesen wurden. Darüber hinaus wurde eine große Heterogenität der off-target Studien bezüglich der methodischen Durchführung und der Art der Analyse der off-target Sequenzen festgestellt. Auf Grundlage dieser Ergebnisse wurde ein Systematic Review durchgeführt, mit dem Ziel Faktoren zu identifizieren und zu analysieren, die das Auftreten von off-target-Effekten beeinflussen. Die Ergebnisse der Analyse zeigen, dass eine höhere Anzahl an Fehlpaarungen („Mismatches“) zwischen der Zielsequenz und ähnlichen potentiellen off-target Sequenzen das Auftreten von off-target Effekten signifikant verringern. Die beobachtete off-target-Rate sank von 59% bei der Tolerierung eines Mismatches auf nahezu 0% bei vier oder mehr Mismatches. Darüber hinaus verringern Mismatches, die sich innerhalb der ersten acht Nukleotide zur PAM befinden, das Auftreten von off-target-Effekten signifikant. Auch ist eine Tendenz erkennbar, dass nebeneinanderliegende Mismatches die Wahrscheinlichkeit von off-target Effekten erhöhen. Dagegen lässt sich basierend auf der Datenlage kein Zusammenhang zwischen dem GC-Gehalt des Protospacers und dem Auftreten von off-target Effekten feststellen. Bezüglich der verwendeten Nuklease und der Art und Weise, wie das CRISPR/Cas Konstrukt in die Zelle integriert wird, ist die Datenlage nicht ausreichend, um belastbare Aussagen über deren Einfluss auf das Auftreten von off-target Effekten zu treffen. Um das Risiko von off-target Effekten auf ein Minimum zu reduzieren, sollten Zielsequenzen ausgewählt werden, die sich in mindestens vier nicht nebeneinanderliegenden Positionen von ähnlichen Sequenzen im Genom unterscheiden.

## Publications included in this thesis

### Peer-review publications

#### Chapter 4 has been published

**Modrzejewski, Dominik**; Hartung, Frank; Sprink, Thorben; Krause, Dörthe; Kohl, Christian; Schiemann, Joachim; Wilhelm, Ralf (2018): What is the available evidence for the application of genome editing as a new tool for plant trait modification and the potential occurrence of associated off-target effects: a systematic map protocol. In: *Environ Evid* 7 (1), S. 11. DOI: 10.1186/s13750-018-0130-6.

#### Chapter 5 has been published

**Modrzejewski, Dominik**; Hartung, Frank; Sprink, Thorben; Krause, Dörthe; Kohl, Christian; Wilhelm, Ralf (2019): What is the available evidence for the range of applications of genome-editing as a new tool for plant trait modification and the potential occurrence of associated off-target effects: a systematic map. In: *Environ Evid* 8 (1), S. 11. DOI: 10.1186/s13750-019-0171-5.

#### Chapter 6 has been published

**Modrzejewski Dominik**, Hartung, Frank; Lehnert, Heike; Sprink, Thorben; Kohl, Christian; Keilwagen, Jens, Wilhelm, Ralf (2020): Which Factors Affect the Occurrence of Off-Target Effects Caused by the Use of CRISPR/Cas: A Systematic Review in Plants. *Front. Plant Sci.* 11:574959. DOI: 10.3389/fpls.2020.574959.

## List of Abbreviations

|            |   |
|------------|---|
| ALS        | Acetolactate synthase   |
| AUC        | Area under the curve  |
| BE         | Base editing  |
| BLAST      | Basic Local Alignment Search Tool   |
| BMEL       | Federal Ministry of Food and Agriculture                                  |
| bp         | Base pair   |
| CABI       | CAB International   |
| CAPS       | Cleaved amplified polymorphic sequence                                    |
| Cas        | CRISPR associated proteins  |
| CEE        | Collaboration for Environmental Evidence                                  |
| CJEU       | Court of Justice of the European Union                                    |
| Cpf1       | CRISPR from <i>Prevotella</i> and <i>Francisella</i> 1                    |
| CRISPR     | Clustered Regularly Interspaced Short Palindromic Repeats                 |
| dCas9      | DeadCas9  |
| DNA        | Deoxyribonucleic acid   |
| DOI        | Digital Object Identifier   |
| DSB        | Double-strand break   |
| e.g.       | For example   |
| ELSA-GEA   | Ethical, Legal and Socioeconomic Aspects of Genome Editing in Agriculture |
| EMS        | Ethyl methane sulfonate   |
| eSpCas9    | Enhanced specificity CRISPR associated proteins                           |
| EU         | European Union  |
| FSTA       | Food Science and Technology Abstracts                                     |
| GAP        | Good Agricultural Practice  |
| GC-content | Guanine-cytosine content  |
| GMO        | Genetically modified organism   |
| HDR        | Homology-directed repair  |
| i.e.       | Id est  |
| Indel      | Insertion-deletion  |

|           |  |
|-----------|--|
| IPCC      | Intergovernmental Panel on Climate Change                                |
| ISBN      | International Standard Book Number                                       |
| JMP       | John's Macintosh Project   |
| kbp       | Kilobase pair  |
| KCI       | Korea Citation Index   |
| MN        | Meganucleases  |
| NGO       | Non-governmental organization  |
| NHEJ      | Non-homologous end joining   |
| ODM       | Oligonucleotide-Directed Mutagenesis                                     |
| PAM       | Protospacer-adjacent motif   |
| PCR       | Polymerase-Chain-Reaction  |
| PICO      | Population/ Intervention/ Comparator/ Outcome                            |
| RNA       | Ribonucleic acid   |
| RNP       | Ribonucleoproteins   |
| ROC       | Receiver operating characteristics                                       |
| ROSES     | Reporting standards for Systematic Evidence Syntheses                    |
| SaCas9    | <i>Staphylococcus aureus</i> CRISPR associated proteins                  |
| SAM       | Scientific Advice Mechanism  |
| SciELO    | Scientific Electronic Library Online                                     |
| SDN       | Site-directed nucleases  |
| sgRNA     | Single guide RNA   |
| SpCas9    | <i>Streptococcus pyogenes</i> CRISPR associated proteins                 |
| SpCas9-HF | <i>Streptococcus pyogenes</i> CRISPR associated proteins - High fidelity |
| TALE      | Transcription Activator-Like Effector                                    |
| TALENs    | Transcription Activator-Like Effector Nucleases                          |
| USA       | United States of America   |
| USDA      | United States Department of Agriculture                                  |
| WGS       | Whole genome sequencing  |
| WoS       | Web of Science   |
| ZFN       | Zinc-Finger Nucleases  |

# 1 Introduction

Over the last decades, technological progress in agriculture and plant breeding has formed the basis for high yields and the production of high-quality agricultural products [1]. This has contributed significantly to a stable food supply in most parts of the world. However, in future new challenges need to be encountered. It is estimated that the world agricultural production needs to increase by 50% by 2050 in order to feed an increasing world population and to cope with an increased demand meet products [2]. In addition, climate change is associated with increased extreme weather events like droughts or floods as well as changing dynamics of pests and diseases [3]. Besides, the available area for agricultural production is limited and further agriculture expansion has a considerable impact on nature, as naturally diverse landscapes are replaced by arable land [1, 3]. The current Special Report of the Intergovernmental Panel on Climate Change (IPCC) "Climate Change and Land Systems" has drawn up numerous recommendations to address these challenges, including changes in consumer behavior and eating habits, reduced food waste and "technical improvements" in agriculture like advances in plant breeding [3].

Plant breeding helps to decrease environmental impacts on cultivation systems by providing varieties resistant to plant diseases or pests, and tolerant to abiotic stress. This may reduce pesticide use and result in less intense management efforts (e.g. irrigation). Further yield improvement can reduce the area required for food production and may balance with areas e.g. for nature conservation [4]. Plant breeding essentially relies on the utilization of genetic variation within the breeding material. For thousands of years, species have been domesticated by selecting the best progeny with desirable characteristics. Since the discovery of Mendel's laws scientific and technological progress has led to the continuous development of new plant breeding tools [5]. As shown in Figure 1.1, within the last decades the nature of technical intervention has shifted from the plant level via cell and tissue level to DNA level.

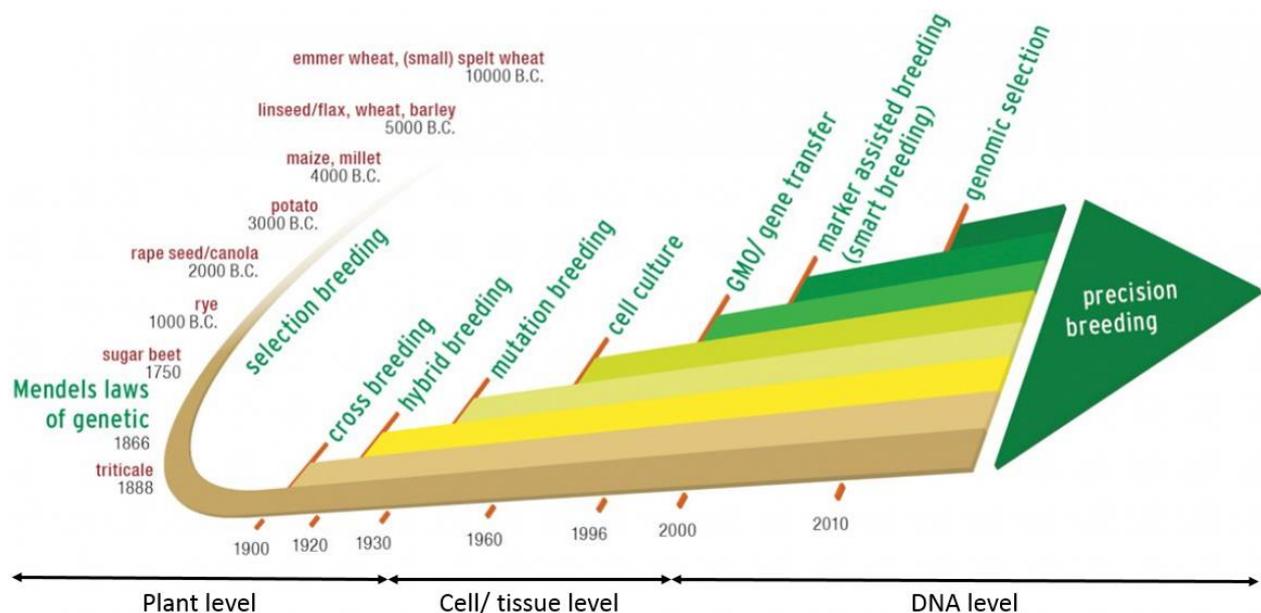


Figure 1.1 Milestones in plant breeding [6].

Nevertheless, key characteristics of plant breeding remain crossing and selecting desired plants, while all the newly developed breeding techniques aim to reach breeding goals in a more efficient and specific way. One milestone in plant breeding was the artificial induction of mutations aiming to increase the natural occurring mutation rate [7]. The induction of these undirected mutagenesis methods lead to several unwanted mutations that have to be eliminated by laborious backcrossing and selection steps in order to select a target trait in a desired genetic background [7]. In 1983, a further milestone was achieved, as for the first time recombinant DNA was delivered to plant cells using *Agrobacterium tumefaciens* [8, 9]. From this time on, it has been possible to work at a single gene level with genetic material from any organism, generating plants that cannot be bred conventionally. Nevertheless, induced mutations using chemical and physical mutagens as well as the “classical” transgenic approach show limited efficiencies and unintended side effects due to the random targeting [10]. In 1996, Zinc-Finger Nucleases (ZFN) was reported as the first genome editing technique enabling a targeted modification of a DNA sequence in a site-directed manner [11]. Since then, further techniques like Oligonucleotide-directed mutagenesis (ODM), Meganucleases (MN), Transcription Activator-Like Effector Nucleases (TALENs), Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated proteins (CRISPR/Cas) and Base editing (BE) have been developed. All these techniques are summarized under the umbrella term genome editing and are promising tools to enhance basic research like analyzing gene functions and to revolutionize plant breeding. However, there is an ongoing controversial debate about the benefits of genome editing for agriculture and the extent and significance of off-target effects.

This thesis aimed to systematically map and evaluate the evidence available on the impact of genome editing in plants by focusing on two aspects: the applications of genome editing and the occurrence of associated off-target effects. To achieve this, a multi stage conceptual approach was adopted. At first, a methodological systematic map protocol was developed that serves as a guidance for the evidence synthesis. Based on the protocol, a systematic map was conducted aiming to provide a comprehensive overview of the applications of genome editing and the occurrence of associated off-target effects. Results of this map determined that most of the off-target studies were conducted by applying CRISPR/Cas and that the analyzed potential off-target sequences within these studies were very heterogeneous in structure and design. Based on these findings, a systematic review was performed in order to identify and systematically analyze factors that may affect the occurrence of off-target effects by applying CRISPR/Cas.

## 1.1 Genome editing

Genome editing is a collective term for several new plant breeding techniques that are able to induce a targeted modification of DNA sequences in a site-directed manner in humans, animals, plants and bacteria. Genome editing comprises three molecular approaches that efficiently induce targeted alterations in the DNA: (i) Site-directed nucleases (SDN) (ii) Oligonucleotide-directed mutagenesis (ODM) and (iii) Base editing (BE).

### 1.1.1 Site-directed nucleases

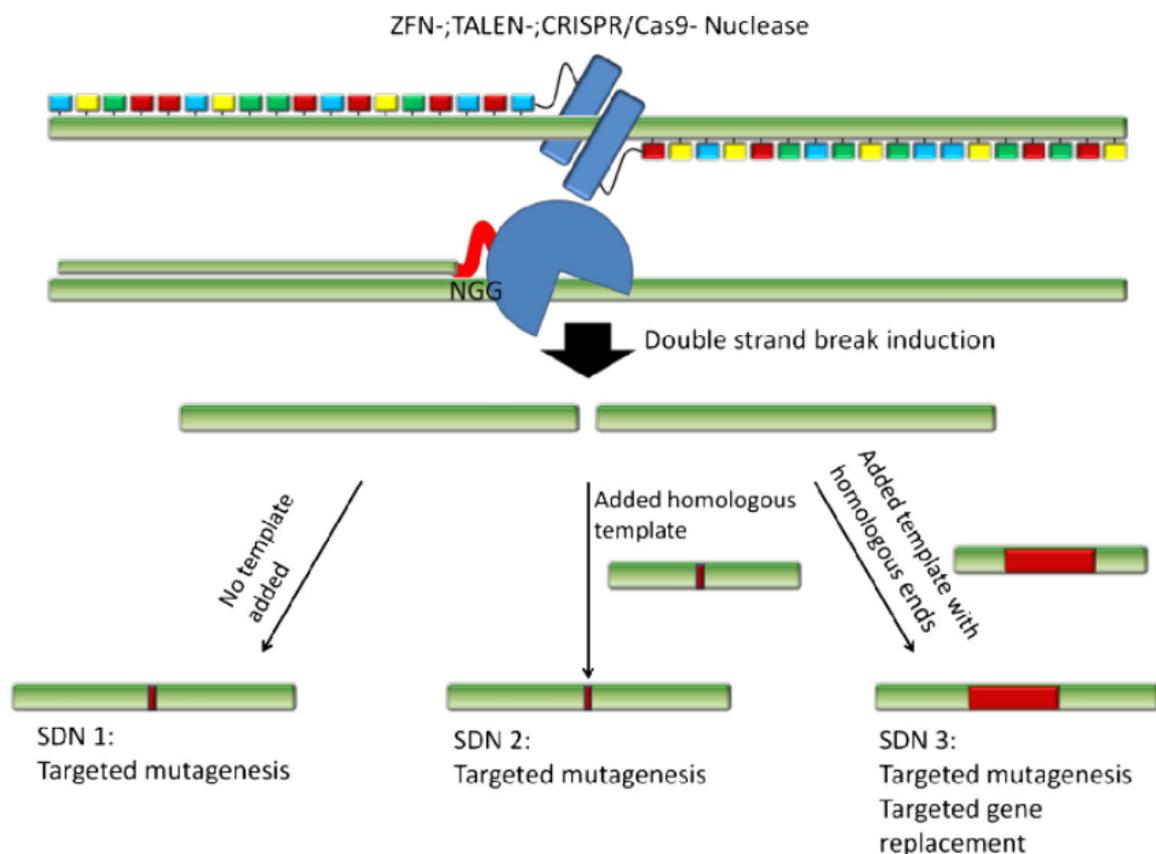
There are four similar systems of SDN: Meganucleases, Zinc-Finger Nucleases, Transcription Activator-Like Effector Nucleases (TALENs) and Clustered Regularly Interspaced Short

Palindromic Repeats/CRISPR associated proteins (CRISPR/Cas). SDN induce double-strand breaks (DSBs) in the DNA, which are subsequently repaired by the host cellular repair mechanisms. The type of repair can be categorized in three main types as shown in Figure 1.2 [5, 12–15]:

SDN1: No additional template is added and the DSB is repaired by non-homologous end joining (NHEJ) resulting in indel mutations.

SDN2: A repair template is added which, except for a few nucleotides, is identical to the sequences in which the DSB is introduced. Then, the DSB is repaired via homology-directed repair (HDR), causing nucleotide substitution or, depending on the template used, targeted indels of a specific size.

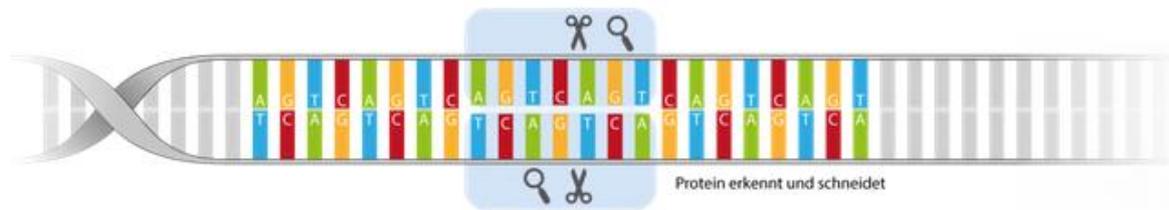
SDN3: The repair template harbors a recombinant DNA sequence additional to the homologous sequences in which the DSB is made and the break is repaired via HDR, resulting in more complex alterations like the insertion of foreign genes.



**Figure 1.2** Overview of the repair mechanism of site-directed nucleases using either Meganucleases, Zinc-finger nucleases, Transcription Activator-Like Effector Nucleases (TALENs) or Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated protein (CRISPR/Cas) [5].

### Meganucleases

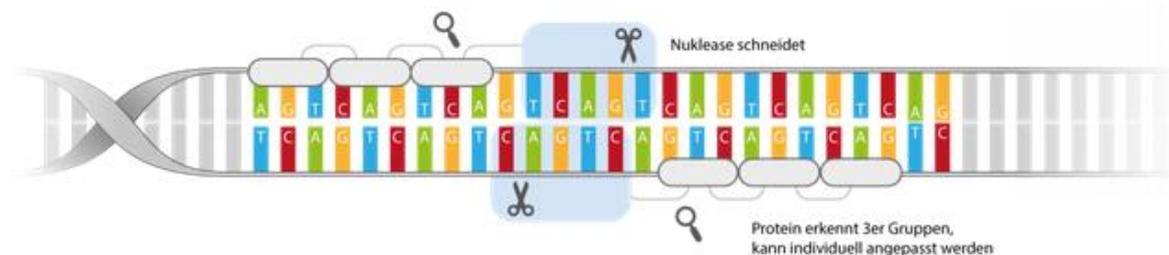
MN are naturally occurring endonucleases that can be modified to bind to a specific DNA sequence and cleave it [16]. The advantage of MN is their small size, making them suitable for use in the majority of delivery techniques [17]. However, the DNA-binding domain cannot be separated from the catalytic domain challenging the construction of novel MN (Figure 1.3) [18].



**Figure 1.3** Inducing a double-strand break by using Meganucleases [19].

### Zinc-finger nucleases

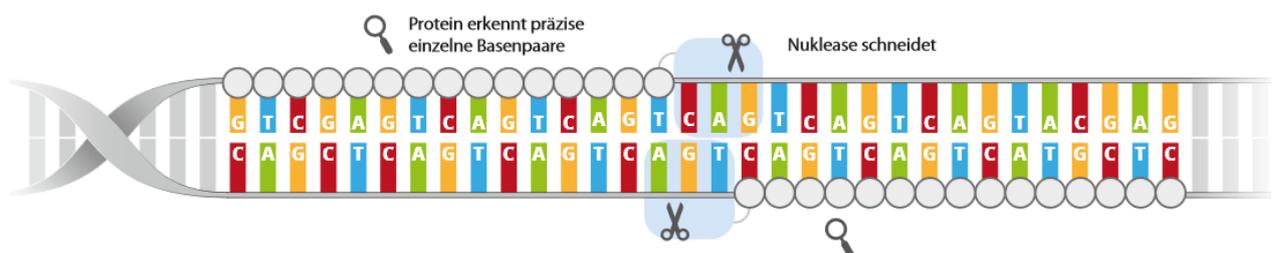
ZFN are generated by fusing two independent protein domains. A zinc-finger protein, which comprises up to six zinc-finger domains, each able to identify a nucleotide triplet of a specific DNA sequence, is fused with a synthetic endonuclease domain (most frequently FokI) (Figure 1.4) [5, 18]. Since the nuclease is active as a dimer only, two ZFN are necessary in close proximity to target and cut a sequence in the genome.



**Figure 1.4** Inducing a double-strand break by using Zinc-finger nucleases [19].

### TALENs

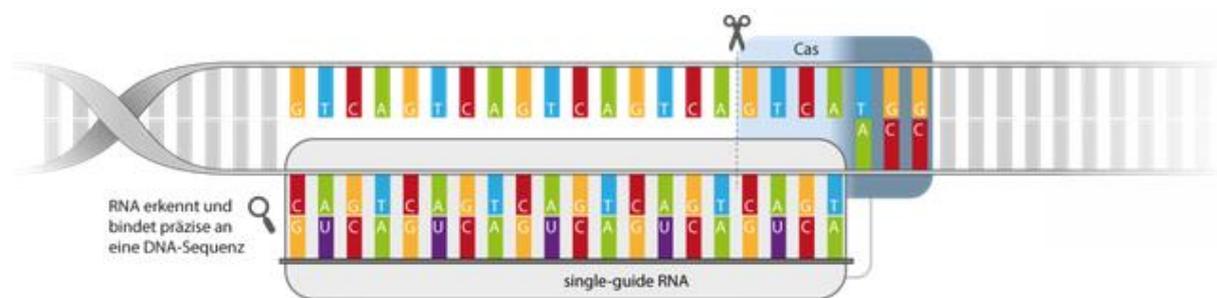
TALENs are composed of two functional parts. The first part consists of the TALE which is originally derived from the *Xanthomonas* species and is crucial for binding to a specific DNA sequence. The TALE is composed of a 34-amino acid repeat, each binding specifically to a single nucleotide in the target DNA [20]. In order to mediate the introduction of a targeted DSB, the TALE is fused to a FokI endonuclease domain (Figure 1.5).



**Figure 1.5** Inducing a double-strand break by using TALENs [19].

### CRISPR/Cas

The most widely used CRISPR/Cas9 system is derived from *Streptococcus pyogenes* consisting of two elements. An artificial single guide RNA (sgRNA) directs the nuclease to a specific DNA sequence (Figure 1.6). Afterwards, the Cas9 endonuclease induces a DSB at this targeted DNA sequence [21]. To induce a DSB, a protospacer-adjacent motif (PAM), such as 5'-NGG-3', has to be present in the site-specific target [22]. Recently, other CRISPR-based nucleases such as Cas12a (also known as Cpf1) from *Prevotella* and *Francisella* have been reported [23].



**Figure 1.6** Inducing a Double-strand break by using CRISPR/Cas [19].

### Comparison of site-directed nucleases

Table 1.1 summarizes and compares the individual characteristics of different SDN. Taking the ease of engineering, cost for development, simplicity, success rate and ease of multiplexing into account the CRISPR/Cas approach looks more attractive compared to MN, ZFN and TALENs (Reviewed by [10, 22, 24–27]).

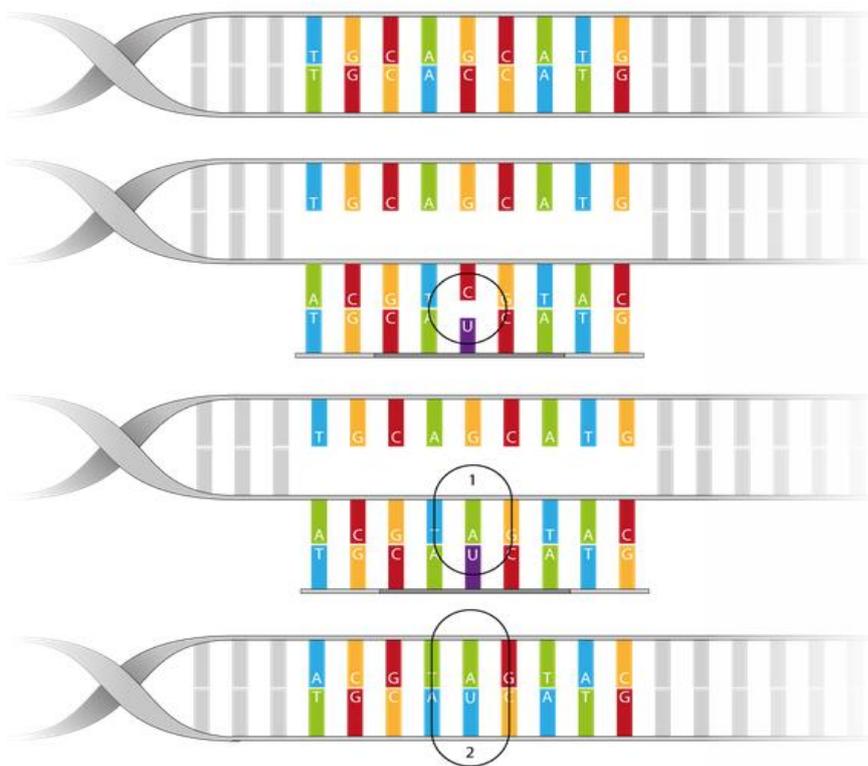
**Table 1.1** Comparison of site-directed nucleases.

|                             | <b>Zinc-finger nucleases</b>  | <b>Meganucleases</b>  | <b>TALENs</b>   | <b>CRISPR/Cas</b>  |
|-----------------------------|---|---|---|--|
| <b>Module</b>               | Protein based   | Protein based   | Protein based   | RNA based  |
| <b>Ease of engineering</b>  | Technically challenging as DNA-binding and catalytic domains are not separated  | Technically challenging in generation as requires protein engineering | Technically challenging in generation as requires protein engineering                             | Easy to engineer compared to other SDN   |
| <b>Recognition site</b>     | 18-36 nucleotides per ZFN pair  | 12 – 40 nucleotides   | 30-40 nucleotides per TALEN pair  | 20-24 nucleotides + 3-5 bp PAM   |
| <b>Success rate</b>         | Low   | Low   | High  | High   |
| <b>Ease of multiplexing</b> | Low   | Low   | Low   | High   |
| <b>Ploidy level</b>         | Not studied   | Not studied   | Useful in case of polyploidy  | Useful in case of polyploidy   |
| <b>Cost for development</b> | High  | High  | Medium  | Low  |
| <b>Advantages</b>           | It is possible to edit various types of genome modification (gene knockout, reporter, specific alleles)                             | Almost any given DNA sequences can be designed                        | Almost any given DNA sequence can be designed to target. Targeting of biallelic genes is possible | High efficiency<br>Easy and cheap<br>Targeting of biallelic and multiplex genes is possible              |
| <b>Disadvantages</b>        | Inefficient for genome editing because of the inadequate knowledge on designing a construct that recognizes a specific DNA sequence | It is more difficult to design ZFN compared to CRISPR/Cas             | Difficult to design, since the iterative cloning of TALE repeats is mistake prone.                | Target sequences are limited by the requirement for the PAM motif<br>Higher chance of off-target effects |

| Zinc-finger nucleases                                      | Meganucleases | TALENs  | CRISPR/Cas |
|--|---------------|---|------------|
| Time consuming compared to other genome editing techniques |               | Selection of targets is limited by the availability of a thymine at the 5' end of the target sequence |            |

### 1.1.2 Oligonucleotide-Directed Mutagenesis

In contrast to SDN, the ODM technique requires neither a nuclease nor a DSB [28]. The mechanism of ODMs is based on the use of chemically synthesized oligonucleotides for the induction of site-specific mutations in the genome. The alteration generally affects between one and four adjacent nucleotides resulting in point mutations [29]. The oligonucleotide is a modified DNA or DNA/RNA molecule of 20-100 nucleotides. It is homologous to a genomic sequence except for the nucleotide(s) that is/are supposed to be modified [29, 30]. The introduced oligonucleotide binds to the targeted DNA sequences and this sequence is then modified by the host cell's mechanism of mismatch repair (Figure 1.7) [12].



**Figure 1.7** Inducing a Double-strand break by using Oligonucleotide-directed mutagenesis [19].

### 1.1.3 Base editing

BE is a recently developed approach which enables a targeted conversion of one DNA base into another in a programmable manner [31]. It is based on the CRISPR/Cas9 system, in which a deadCas9 enzyme (dCas9) is coupled with the enzyme cytidine-deaminase. This fusion protein is still able to bind to the sgRNA and is directed to the target sequence. However, the dCas9 enzyme

is not able to induce a DSB anymore. Instead, it mediates the direct conversion of cytosine to uridine within a window of approximately five nucleotides by the cytidine-deaminase [31]. The technique is suitable for the introduction of point mutations but so far, only two conversions are possible: The C-G base pair can be converted to T-A and A-T can be converted to G-C [32].

### 1.1.4 Comparison of genome editing and conventional breeding techniques

Genome editing offers substantial advantages compared to conventional breeding techniques, opening up new dimensions for scientists and plant breeders:

- (i) Genome editing provides the opportunity to induce targeted mutations in one or a few genes. Previous mutation-initiating methods using chemicals or radiation induce undirected mutations in many sequences randomly in the genome. It is not known in which gene sequences the changes occur. In a subsequent selection process, those containing the desired alterations need to be selected from a large number of plants [33].
- (ii) Genome editing provides the opportunity to selectively replace entire genes from both closely (cisgene) as well as distantly related organisms (transgene) without inserting foreign DNA sequences at all [12, 34]. In contrast, using conventional genetic engineering, traces of recombinant DNA, from the used gene shuttle (e.g. bacteria, virus or plasmid), persist in the modified organism leading to clearly characterized genetically modified organisms [34].
- (iii) Polyploid plant species like hexaploid wheat were to date largely inaccessible for simultaneously targeted alterations. Genome editing provides the opportunity to induce simultaneous mutations in all six alleles [35]. This was demonstrated in several polyploidy plants including maize, cotton, canola, potato, apple, peanuts and citrus [36].
- (iv) The precision and efficiency of the technique facilitates the elucidation of the function of poorly understood genes or gene variants. So far, work in higher plants has usually required a very large number of experimental plants to switch-off or modify individual genes. In addition, it took months or even years. Due to the efficiency of CRISPR/Cas, several genes can now be changed simultaneously in numerous organisms within a few weeks [34].

## 1.2 Off-target effects

### 1.2.1 Reasons for the occurrence of off-target effects

Off-target effects may occur when a genome editing technique induces unintended cleavage and mutations in a DNA sequence. It is estimated that off-target effects occur mainly in sequences that are similar to the target one but located at another site in the genome. For example, the standard CRISPR/Cas technique has a sgRNA of 20 nucleotides. With a certain probability different sequences of the DNA are very similar, as illustrated in Figure 1.8. Nucleotide deviations between the on-target and potential off-target sequences are called mismatches. Since the sgRNA is able to tolerate some mismatches, off-target effects may occur in these similar DNA-sequences [17, 37, 38]. Therefore, it is recommended to carefully select and modify a target sequence that has no further similar sequences in the genome [39–41].

|              | guideRNA                                | PAM   | # mismatches |
|--------------|---|-------|--------------|
| On-target    | G G T T C T T C C T G T A C G A G A G C | C G G |              |
| Off-target 1 | G G T T C T T C C T G T A A G A G A G C | N G G | 1            |
| Off-target 2 | G G T T C G G C C T G T A C G A G A G C | N G G | 2            |
| Off-target 3 | G C G T C T T C C T G T A C C A G A G C | N G G | 3            |
| Off-target 4 | G G T G G T T C C T G T A C G C G A C C | N G G | 4            |
| Off-target 5 | T C T T C T T C G T G T C A G A G A G C | N G G | 5            |

**Figure 1.8** Illustration about potential off-target effects in plants using CRISPR/Cas9.

With regard to CRISPR/Cas, further factors are described in literature as potentially affecting the occurrence of off-target effects are:

- (i) The position of mismatches proximal to the PAM [41–43].
- (ii) The applied nuclease variants of the CRISPR-system (e.g. SpCas9, Cpf1,...) and engineered variants of the standard nuclease SpCas9 (e.g. eSpCas9, SpCas9-HF or SpCas9 (1.1) [44–47].
- (iii) The guanine-cytosine content (further referred to as GC-content) of the target sequence [37, 48].
- (iv) The kind of delivery of the CRISPR/Cas system into the plant cell (e.g. using DNA, RNA or Ribonucleoproteins (RNPs)) [49].

Detailed explanations of the individual factors potentially affecting off-target effects are provided in chapter 6.

### 1.2.2 Possibilities to detect off-target effects

Off-target effects can be detected by using either biased or unbiased detection methods. In the biased approach, the search for off-target effects is restricted to sequences that are similar to the target sequence. In contrast, the unbiased approach examines the whole genome for the occurrence of off-target effects.

The biased approach to detect off-target effects consists of two steps:

1. Sequence alignment programs like BLAST [50], Cas-OFFinder [51] or CRISPR-P [52] are used to identify sequences in the genome with high similarity to the target one.
2. Detection methods like Sanger sequencing or targeted deep sequencing are used to analyze the potential off-target sequences for the occurrence of off-target effects [46].

A drawback of this approach is the risk of overlooking off-target mutations in further sequences in the genome that are not assessed for off-target activity [40].

The unbiased approach is able to identify off-target effects in a completely unrestricted way without a priori identification of similar sequences. It requires genome-wide sequencing to detect off-target effects anywhere in the genome [40, 53]. Drawbacks of this approach are high costs, which means that usually only a few clones are sequenced for each target, and hence low-frequency off-target effects might be missed because of low coverage [54]. In addition, mutations do not necessarily

originate from a genome editing technique, but may occur spontaneously or by using conventional breeding techniques like somaclonal variation [12].

### 1.2.3 Off-target effects in the context of plant breeding

Since plant breeding is the contextual background, it is worth comparing the occurrence of off-target effects in the context of naturally occurring mutations and routinely used breeding techniques like regular crossing, undirected mutagenesis using chemical mutagens or irradiation or the regeneration of plants from cell culture (somaclonal variation) [12]. The natural mutation rate of *Arabidopsis thaliana* is approximately one mutation per 150,000 kilo base pairs (kbp) which means that considering the genome size of *Arabidopsis* around one mutation occurs per generation [32, 55]. Using chemicals like EMS or irradiation to induce mutations in plants the mutation rate increases dramatically, e.g. Jander et al. (2003) identified at least 700 mutations in EMS-mutagenized *Arabidopsis* lines [56]. In addition, experiments show that somaclonal variation increases the mutation rate by a factor up to 250 compared to spontaneous mutations [57]. Compared to these techniques, genome editing off-target effects occur less frequently and to a lower extent. Nevertheless, there is an ongoing debate about the potential occurrence of off-target effects and the extent to which it matters in plants breeding [58, 59].

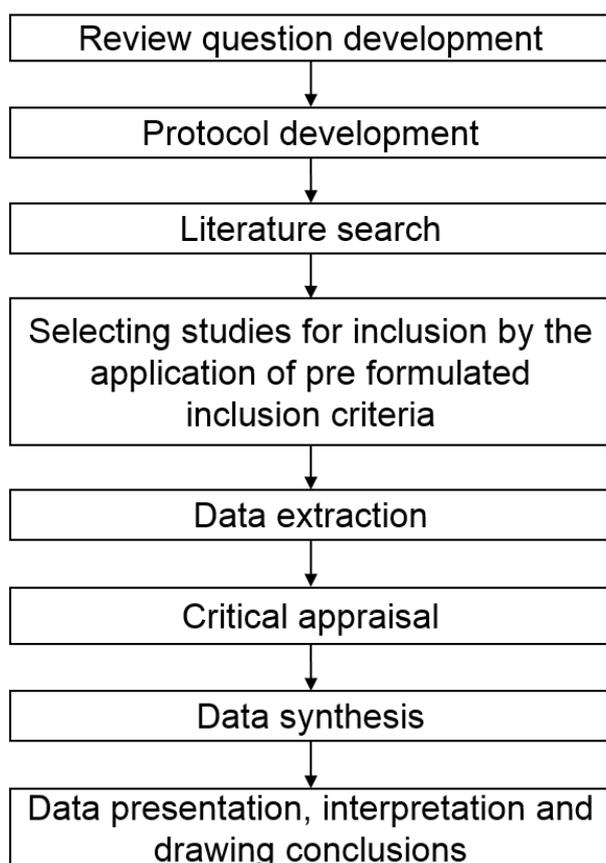
## 1.3 Evidence synthesis

The number of research articles is rising rapidly. It is estimated that the number of articles published is increasing by about 3% and 3.5% per year [60]. In total, the number of published research articles passed 50 million in 2009 [61] and in 2018, around 42,000 peer-reviewed journals collectively published over three million articles [60]. Evidence synthesis plays an important role in order to control the rapidly expanding evidence base and to inform evidence-based decision-making. Evidence synthesis tools were first established in the field of healthcare and then expanded to several other disciplines, including international development, social welfare, policy, crime and justice, education, environmental management and food/ feed safety assessment [62, 63]. Today, evidence synthesis, including evidence review and synthesis methodology, is widely used in disciplines where science can inform decision-making. Powerful tools including systematic maps and systematic reviews were developed to systematically identify, collect, evaluate and summarize research results to assess a scientific question. These tools provide a transparent and reproducible process aiming at maximizing objectivity and minimizing bias [64].

### 1.3.1 Systematic review

A systematic review is defined as “*an evidence synthesis method that aims to answer a specific question as precisely as possible in an unbiased way. The method collates, critically appraises, and synthesizes all available evidence relevant to the question*” [64].

Conducting a systematic review consists of eight steps as indicated in Figure 1.9. The single steps are summarized based on the guideline of Collaboration for Environmental Evidence (CEE) [64].



**Figure 1.9** Basic steps to follow when conducting a systematic review.

### **1. Step: Developing a review question**

Well-framed and focused questions are mandatory for conducting a systematic review. To assure this, key elements have to be specified. The most common question type is the PICO-question which was developed for medical practice. P describes the **P**opulation of interest, I the **I**ntervention the population is exposed to, C the **C**omparator and O all relevant **O**utcomes from the proposed intervention. As the methodology of a systematic review has broadened to other disciplines, several further question types have been developed.

### **2. Step: Protocol development**

Before conducting a systematic map or a systematic review, a protocol has to be published (including a peer-review process). The protocol can then be used to guide the review team through the review process. Additionally, the protocol helps to make the review process more rigorous, transparent and well-defined as well as minimizing reviewer bias, for example due to ad-hoc decisions made when conducting the review.

### **3. Step: Literature search**

The search for literature should be well structured to identify as many relevant articles as possible. Literature should capture scientific literature databases and catalogues covering relevant subject areas as well as grey literature. Grey literature comprises, for example, research that has been published in non-commercial form like degree theses or government reports but also searching websites of companies and organizations. To test the comprehensiveness of the search strategy a scoping search should be carried out to test the identified records against an a priori defined test library with articles of known relevance.

#### **4. Step: Article screening**

Eligibility criteria need to be precisely defined considering all key elements of the review question in order to determine which research studies gained from the literature search are indeed relevant for the review. Two reviewers screen all articles for relevance independently. The selection process should be precisely reported. To minimize the risk of inconsistent screening a kappa-test is conducted to increase inter-reviewer agreement. To ensure transparency, articles included in the review, but also articles excluded at full text level should be listed.

#### **5. Step: Data extraction**

Relevant data should be extracted and collated as described in the protocol. One article can contain several studies when different populations, interventions or outcomes are described within this article. All data extracted should be listed in a table or spreadsheet and cross-checked by a second reviewer to improve transparency, repeatability and objectivity. In the case of relevant data being missed, the authors of the respective articles should be contacted and asked to provide the information missing.

#### **6. Step: Critical appraisal**

A critical appraisal of all studies included has to be conducted on study-by-study basis. The critical appraisal aims to identify all relevant sources of bias including both internal validity (risk of bias within a specific study) and external validity (generalizability). Based on the results of the critical appraisal, studies can be categorized as having high, medium, low or unclear validity. A high level of transparency is expected by providing detailed information about the critical appraisal.

#### **7. Step: Evidence synthesis**

A narrative synthesis has to be provided in order to overview the available evidence and to present the body of evidence. Depending on the studies included in the review and the data gained therein, suitable synthesis methods (e.g. meta-analysis, narrative reporting) have to be chosen for data analysis and synthesis. Additionally, effect modifiers that may affect the review results should be investigated.

#### **8. Step: Data presentation, interpretation and conclusion making**

Conducting a systematic review aims to provide reliable evidence relating to a specific review question. Therefore, a careful reporting and interpretation of the results is mandatory. Reviewers have to avoid reaching conclusions beyond the evidence base or presenting only parts of the results. Additionally, limitations of the evidence synthesis as well as implications for policy and research should be described.

### **1.3.2 Systematic map**

Sometimes the evidence base is unclear and it is not expedient to conduct a systematic review by formulating a specific research question directly. In order to get a comprehensive overview by collating, describing and mapping findings for a given topic, it can be useful to conduct a systematic map a priori. The systematic map is based on a broader and open-framed review question and can support the decision as to whether a more specific synthesis of outcomes is suitable to be addressed in a systematic review [64]. The systematic mapping process is conducted with the same comprehensive method as for a systematic review, except that a critical appraisal is not mandatory.

Table 1.2 summarizes the major methodological differences of a systematic map and systematic review compared to a traditional review [64, 65]:

**Table 1.2** Major methodological differences between a traditional review, systematic map and systematic review.

|   | <b>Traditional Review</b>                                 | <b>Systematic Map</b>   | <b>Systematic Review</b>  |
|---|---|---|---|
| <b>Study question</b>   | Often broad in scope                                      | Broad in scope or focused and explicit  | Focused and explicit  |
| <b>Protocol</b>   | Not mandatory   | Mandatory   | Mandatory   |
| <b>Eligibility criteria for inclusion or exclusion of studies</b> | Not always explicitly stated                              | Pre-defined and documented; objectively applied                                       | Pre-defined and documented; objectively applied                                       |
| <b>Description of the review method</b>                           | Seldom reported   | Reported and predefined in the protocol   | Reported and predefined in the protocol   |
| <b>Literature search</b>  | Not always extensive                                      | Structured to identify as many relevant studies as possible including grey literature | Structured to identify as many relevant studies as possible including grey literature |
| <b>Critical appraisal of study validity</b>                       | Variable  | Variable  | Included  |
| <b>Reporting of studies outcomes</b>                              | Selective reporting; often of study author interpretation | Full reporting of relevant outcomes   | Full reporting of relevant outcomes (numerical results)                               |
| <b>Synthesis</b>  | Usually narrative, sometimes selective                    | Usually narrative   | Quantitative synthesis (meta-analysis) when possible                                  |

### 1.3.3 Disadvantages and challenges of systematic reviews

The key characteristics of systematic reviews including comprehensive methods applied the high reporting standard to ensure transparency and repeatability helps to maximize objectivity by minimizing bias. Therefore, systematic reviews are seen as the gold standard when synthesizing primary research [63]. However, despite these advantages, systematic reviews also have challenges which have to be overcome. One weak point of systematic reviews is their resource intensity, including time, money and man power due to the peer-review protocol, the rigorous methodologies and documentation effort [63]. Dependent on the experience in this field a systematic review takes between 12 and 24 months. However, in some situations decision maker have a short timeframe for their decision and can't wait until a systematic review has been conducted [66]. A further challenge is that access is needed to several databases and peer-review journals which can be very expensive [67].

### 1.3.4 CADIMA - an online tool to support the conduct and reporting of systematic reviews and systematic maps

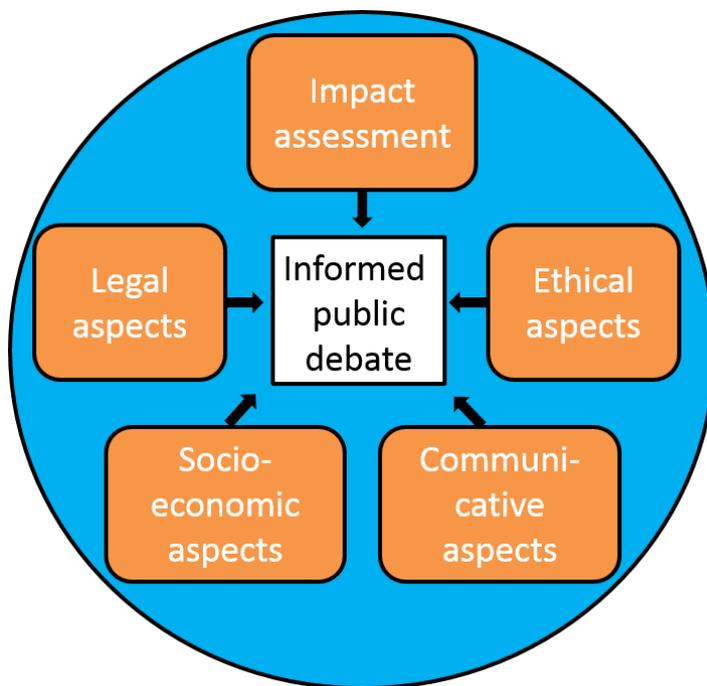
CADIMA is an open access tool aiming to increase the efficiency of the review process and to facilitate reporting of all steps when conducting a systematic map or a systematic review [63]. It was developed by a collaboration between the Julius Kühn-Institute and the CEE.

**CADIMA:**

- (i) Is suited to be used for further disciplines beside medical science.
- (ii) Was developed to support throughout all steps of the review/ mapping process.
- (iii) Facilitates the coordination of the review team.
- (iv) Ensures documentation of the entire review synthesis process [63].

## 2 The ELSA-GEA project

The ELSA-GEA project was funded by the German Federal Ministry of Education and Research (01GP1613B). Within this project, a comprehensive and transdisciplinary approach was applied to address the broad aspects of genome editing in the agricultural context. Ethical, legal, socio-economic and communicative aspects as well as risk assessment approaches were addressed in a systematic process for crop plants. Based on a broad spectrum of results from the different disciplines, structured information were provided to enable a rational societal dialogue on genome editing in agriculture. Publications and the project webpage [www.dialog-gea.de](http://www.dialog-gea.de) aimed to inform policy makers and professional stakeholders on the one hand and a broader public on the other hand. The results of this project aimed to support an informed public debate and provided fact-based input for decision-makers in politics, economy, science and society (Figure 2.1).



**Figure 2.1** Overview on the transdisciplinary approach of the ELSA-GEA project.

### 3 Objectives of this thesis

The number of research articles about genome editing in plants is increasing sharply and evidence synthesis of these articles play an important role in order to control the rapidly expanding evidence base and to inform evidence-based decision-making. The overall objective of this study was to systematically identify, collect, evaluate and summarize the available evidence about the application of genome editing in plants. To realize this a conceptual multi-stage approach was applied.

The main objectives were to:

- (i) Develop a methodological systematic map protocol as an a priori guide and reference to the conduct of the synthesis (Chapter 4).
- (ii) Systematically map the available evidence in order to provide a general overview of the application of genome editing in plants (Chapter 5).
- (iii) Provide an overview of the traits modified by genome editing in plants (Chapter 5).
- (iv) Provide an overview of the available evidence about the occurrence of off-target effects due to the use of genome editing in plants (Chapter 5).
- (v) Assess whether a specific section of the available evidence base is suitable for an in-depth analysis by a systematic review (Chapter 5).
- (vi) Conduct a systematic review in order to collect and synthesize the available evidence about factors that may affect the occurrence and detection of off-target effects caused by the application of CRISPR/Cas in plants (Chapter 6).

## 4 Systematic map protocol

---

### **What is the available evidence for the application of genome editing as a new tool for plant trait modification and the potential occurrence of associated off target effects: A systematic map protocol**

---

Chapter 4 has been published:

**Modrzejewski, Dominik**<sup>1</sup>; Hartung, Frank<sup>1</sup>; Sprink, Thorben<sup>1</sup>; Krause, Dörthe<sup>1</sup>; Kohl, Christian<sup>1</sup>; Schiemann, Joachim<sup>1</sup>; Wilhelm, Ralf<sup>1</sup> (2018): What is the available evidence for the application of genome editing as a new tool for plant trait modification and the potential occurrence of associated off-target effects: a systematic map protocol. In: *Environ Evid* 7 (1), S. 11. DOI: 10.1186/s13750-018-0130-6.

<sup>1</sup>Institute for Biosafety in Plant Biotechnology, Federal Research Centre for Cultivated Plants, Julius Kühn-Institut, Erwin-Baur-Straße 27, 06484 Quedlinburg, Germany

#### **Authors' contributions**

**DM**, FH, TS, CK, DK, JS and RW conceived the review question. **DM** undertook pilot research. **DM** drafted the protocol text with support from FH, TS, CK, DK and RW. **DM** will coordinate the mapping process, analysis and presentation of the results. **DM**, FH, TS and DK will screen the articles. All authors read and approved the final manuscript

## 4.1 Abstract

### Background

Plant breeding is a developing process and breeding techniques have continuously evolved over time. In recent years, genome editing techniques such as Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated proteins (CRISPR/Cas), Transcription Activator-Like Effector Nucleases (TALENs), Zinc-Finger Nucleases (ZFN), Meganucleases (MN) and Oligonucleotide-Directed Mutagenesis (ODM) enabled a precise modification of DNA sequences in plants. Genome editing has already been applied in a wide range of plant species due to its simplicity, time saving and cost-effective application compared to earlier breeding techniques including classical mutagenesis. Although genome editing techniques induce much less unintended modifications in the genome (off-target effects) compared to classical mutagenesis techniques, off-target effects are a prominent point of criticism as they might cause genomic instability, cytotoxicity and cell death.

### Methods

The aim of this systematic map is to address the following primary question: “*What is the available evidence for the application of genome editing as a new tool for plant trait modification and the potential occurrence of associated off-target effects*”? The primary question will be considered by two secondary questions: One is aimed at the traits being modified by genome editing in plants and the other explores the occurrence of off-target effects. The systematic map will focus on model plants as well as on plants produced for agricultural production that were subjected to genome editing techniques. Academic and grey literature will be searched in English and German language. Inclusion/ exclusion criteria were developed for the two secondary questions and will be applied on title/ abstract and full text stage. Included studies will be catalogued in a searchable and open access database and study results will be summarized using descriptive statistics. Furthermore, the extracted data will serve as a preparatory step for further in-depth analysis, e.g. by a systematic review.

## 4.2 Background

Technological progress in agriculture and plant breeding has contributed significantly to a stable food supply and has formed the basis for high yields and the agricultural production of high quality products. However, in an ever changing world new challenges are encountered within the next decades. Aside the demands of the growing global population and limited fossil resources, climate change is a driver of breeding efforts as it is associated with increased extreme weather events like droughts or floods as well as changing dynamics of pests and diseases. Agriculture needs to ensure and increase the world agricultural production to serve extended demands with limited environmental resources like soil and water [1]. In contrast, intensification of agriculture causes considerable impact on nature, as naturally diverse landscapes are replaced by arable land for the cultivation of few plant species. Biodiversity is threatened through habitat loss and pesticide use which in turn is considered to increase disease and pest pressure [2]. Additionally, environmental impacts on agriculture are becoming increasingly important in societal debates.

To solve all these challenges a combination of Good Agricultural Practice (GAP) and innovation in plant breeding is needed. GAP addresses environmental, economic and social sustainability leading to safe and healthy food [3]. Examples for GAP are the preservation of natural soil fertility through suitable crop rotations, fertilization and plant protection according to the principles of integrated cultivation or a balanced and species-appropriate animal husbandry [3]. Besides, plant breeding is of crucial importance to manage environmental impacts on cultivation systems by providing varieties resistant to plant diseases or pests, and tolerant to abiotic stress. This may reduce pesticide use and result in less intense management efforts (e.g. irrigation). Further yield improvement will reduce the area required for food production and may balance with areas e.g. for nature conservation [2].

Plant breeding essentially relies on the utilization of genetic variation within the breeding material that can be used for crossing and selection steps to develop improved varieties. New genetic variation can occur naturally by spontaneous mutations that enable populations to adapt to changing environmental conditions. However, as the mutation rate is fairly low and at random, plant breeders and scientists artificially induce mutations for already several decades. The first generation of mutation breeding used chemical and physical mutagens to generate a plurality of nonspecific mutations. The increased mutation rate results in plants with a few positive, a lot of neutral and several negative characteristics. Thus, laborious backcrossing and selection steps are necessary in order to select for a desired trait.

Nevertheless, today more than 3200 mutant varieties from 214 different plant species have been generated through undirected mutagenesis and have been officially registered [4].

**Glossary: [2, 5]****Backcrossing**

Backcrossing is a crossing of a hybrid with one of its parents in order to achieve offspring's that are genetically closer to the selected parent. This way desired heterologous traits from the hybrid can be transferred into the genetic background of a parental line. Since crossing recombines all genes many backcrosses are necessary to achieve considerable dilution of unwanted genes from the hybrid.

**Mutagen**

A mutagen is an agent that increases the mutation rate within an organism or cell, e.g. X-rays, gamma-rays or chemicals like ethyl methane sulfonate (ESM).

**Mutation breeding**

A plant breeding approach using mutagens to enhance genetic variation. The resulting random mutations can generate new gene variations with positive traits that can be selected for further breeding. However, several of these mutations are negative and diminish the viability of the plant.

**Selection**

A process in breeding by which the breeder chooses only those individuals that show desired trait(s).

In 1983, the first recombinant DNA was delivered to plant cells using *Agrobacterium tumefaciens* [6, 7]. From this time on, it is possible to work at a single gene level with genetic material from any organism generating plants that cannot be bred by conventional breeding techniques [8, 9]. Nevertheless, the induced mutations using chemical and physical mutagens as well as the “classical” transgenic approach show limited efficiencies due to the random targeting of the modified site [10].

In recent years, genome editing techniques have been developed enabling a precise modification of DNA sequences in a specific and site-directed manner [11].

To date, genome editing comprises two molecular approaches that efficiently induce targeted alterations in genomes: (1) site-directed nucleases (SDN) and (2) oligonucleotide-directed mutagenesis (ODM).

Site-directed nucleases induce double-strand breaks (DSBs) in the DNA which are subsequently repaired by the hosts own cellular mechanisms. The type of repair can be categorized in three main types [11–14]:

- (1) No additional template is added and the DSB is repaired by non-homologous end joining (NHEJ) resulting in small insertion-deletion (indel) mutations. This approach is defined as SDN1.
- (2) A repair template is added which, except for a few nucleotides, is identical to the sequences in which the DSB is introduced. Then, the DSB is repaired via homology-directed repair (HDR), causing nucleotide substitution or targeted indels. This approach is defined as SDN2.
- (3) The repair template harbors a recombinant DNA sequence additional to the homologous sequences in which the DSB is made and the break is repaired via HDR, resulting in more complex alterations i.e. the insertion of foreign genes. This approach is defined as SDN3.

All three types of DSB repair can be induced using one of the following engineered nucleases:

(i) Meganucleases (MN), (ii) Zinc-Finger Nucleases (ZFN), (iii) Transcription Activator-Like Effector Nucleases (TALENs) and (iv) Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated proteins (CRISPR/Cas).

### **Meganucleases**

Meganucleases are naturally occurring endonucleases that can be modified to bind to a specific DNA sequence and cleave it [15]. The advantage of Meganucleases is their small size, making them appropriate to a majority of delivery methods [16]. However, the DNA-binding domain cannot be separated from the catalytic domain challenging the construction of MN [4]. MN have been applied successfully for genome editing in plants such as *Arabidopsis* [17], maize [18] and cotton [19].

### **Zinc-finger nucleases**

In 1996, zinc-finger nucleases were reported as the first programmable site-specific nucleases [20]. ZFN are generated by fusing two independent protein domains. A zinc-finger protein, which comprises up to six zinc-finger domains each able to identify a nucleotide triplet of a specific DNA

sequence, is fused with a synthetic endonuclease domain (most frequently FokI) [4, 9, 29]. Since the nuclease is active as a dimer, two zinc-finger nucleases are necessary in close proximity to target and cut a sequence in the genome.

### **TALENs**

Similarly to ZFN, TALENs are also composed of two functional parts. The first part consists of the TALE which is originally derived from the *Xanthomonas* species and is crucial for the binding to a specific DNA sequence. The TALE is composed of a 34-amino acid repeat, each binding specifically to a single nucleotide in the target DNA [21]. In order to mediate the introduction of a targeted DSB, the TALE is fused to a FokI endonuclease domain. Compared to ZFN where each repeat recognizes a cluster of three nucleotides and interferes with neighboring repeats, the design of specific TALE DNA-binding domains is easier and amenable to programming [22]. Similar to ZFN, TALENs are most frequently used as pairs to introduce a DSB at a specific target site of the DNA [9].

### **CRISPR/Cas9**

The most widely used CRISPR/Cas system is derived from *Streptococcus pyogenes* consisting of two elements. An artificial single guide RNA (sgRNA) is directing the nuclease to a specific DNA sequence. Afterwards, the Cas endonuclease induces a DSB at this targeted DNA sequence [23]. To induce a DSB, a protospacer-adjacent motif (PAM), such as 5'-NGG-3', has to be present in the site specific target [24]. The difference to TALENs and ZFN is that, instead of a protein, a short sgRNA is used for target recognition within the CRISPR/Cas9 system. This sgRNA can be easily adapted to match the target sequence. Compared to ZFNs and TALENs the CRISPR/Cas9 system is easier, faster and more flexible since only the sgRNA has to be adapted to a new sequence instead of the whole binding proteins as it is the case for MN, ZFN or TALENs [25].

### **Oligonucleotide-directed mutagenesis**

In contrast to MN, ZFN, TALENs and CRISPR/Cas9 the ODM technique does not require a nuclease and a DSB [26]. The mechanism of ODMs is based on the use of chemically synthesized oligonucleotides for the induction of site-specific mutations in the genome. The alteration generally affects one to four adjacent nucleotides resulting in point mutations [27]. The oligonucleotide is a modified DNA or DNA/RNA molecule of 20 - 100 nucleotides. It is homologous to a genomic sequence except for the nucleotide(s) that is/are supposed to be modified [8, 27]. The introduced oligonucleotide binds to the targeted DNA sequences and this sequence is then modified by the host cell's mechanism of mismatch repair [11].

Genome editing offers substantial advantages compared to previous mutation breeding techniques and conventional genetic engineering in terms of speed and precision. Genome editing provides the opportunity to selectively mutate or modify one or a few genes (SDN1, SDN2). In addition, it is now possible to precisely modify or selectively replace (SDN3) entire genes from both closely as well as distantly related organisms [11]. By the use of conventional genetic engineering traces of recombinant DNA, from the viruses or bacteria that were used as gene shuttle persist in the modified organism leading to clearly characterized genetically modified organisms. In contrast, by applying genome editing it is possible to modify crops without inserting foreign DNA sequences at all [28]. This may reduce the regulatory burden for plant breeders and increase the acceptance of genome editing within society. Based on the simplicity, time saving and cost effective

application of genome editing, it has already been applied in a wide range of cultivars. Genome editing has been used for:

- (i) Analyzing gene functions (e.g. effect of the RAV2 gene for salt stress in rice [29]).
- (ii) Improvement of product quality (e.g. decreased linolic acid in soybean [30]).
- (iii) Development of disease resistant varieties (e.g. virus resistant cucumber [31]).
- (iv) Developing of herbicide tolerant varieties (e.g. resistance to the herbicide chlorsulfuron in oil seed rape [32]).
- (v) Improved adaption to abiotic stress, (e.g. drought tolerance in maize [33]).

Even in plants like hexaploid wheat that were so far largely inaccessible for targeted genetic alterations the simultaneous mutation of all six alleles was successfully performed [34]. All these open new dimensions for the scientific, plant breeding and agricultural community.

Compared to randomly induced mutations by chemicals or irradiation, the number of unintended mutations (off-target effects) is greatly reduced by genome editing techniques [11]. Nevertheless, their application does not completely or *per se* exclude the occurrence of off-targets. Off-targets are changes in a certain DNA sequence being similar to the targeted one but located at another site in the genome. Mainly, they occur due to the lack of exclusiveness and/ or length of the recognition site [11, 35, 36]. Several methods have been developed to predict and identify off-target sites linked to the use of genome editing. One can differentiate between the prediction of off-target effects using *in silico* methods and the detection of off-targets using either biased detection methods to analyze individual DNA sequences or unbiased detection methods where genome-wide off-target analysis are conducted [37]. Depending on the detection method being used the results of identifying off-target mutations vary widely. Although genome editing techniques induce much less off-target effects compared to classical mutagenesis techniques, off-targets are an important point of criticism as they may possibly cause genomic instability, cytotoxicity and cell death [38, 39, 40].

Risk assessors and decision makers are depending on the provision of a reliable body of evidence to support conclusions about potential risks being associated with the application of genome editing. Thus, the provided overview on the available evidence on the occurrence of off-target effects will be of crucial importance. Furthermore, this systematic map facilitates an objective debate by informing interested stakeholder communities in a transparent and retraceable manner about the status of research, the progress in genome editing in plants and the available evidence about the potential occurrence of associated off-target effects. The results of the systematic map will be discussed on a stakeholder conference as well as within an expert group established as part of the ELSA-GEA project. These meetings will be an important part to identify relevant key aspects that should further be analyzed within a systematic review.

### 4.3 Objectives of the map

Due to its strong implications of plant breeding, genome editing is of particular relevance to scientists, regulators and policy-makers in the EU and worldwide. Therefore, we want to survey the available evidence about applications of genome editing in plants. The main objectives are:

- (i) Overview of the traits modified by genome editing in model plants as well as in crops produced for agricultural production.
- (ii) Overview of the available evidence about the occurrence of off-target effects due to the use of genome editing techniques in model plants as well as in crops produced for agricultural production.
- (iii) Identification of the volume of the available literature, evidence clusters and key characteristics of the evidence base to inform interested stakeholder communities.
- (iv) Identification of knowledge gaps concerning the occurrence of off-target effects in order to inform decision makers which future research might be needed for a risk assessment.
- (v) Assessment whether the available evidence base is suitable for in-depth analysis such as by a systematic review.

The primary question of the systematic map is: *“What is the available evidence for the application of genome editing as a new tool for plant trait modification and the potential occurrence of associated off-target effects”?*

To answer this primary question, it is reconsidered by two secondary questions related to (1) the traits modified by genome editing and (2) the occurrence of off-target effects due to the use of genome editing.

### **Secondary question one**

*“What are the traits modified by genome editing in model plants as well as in crops produced for agricultural production?”*

Population: Any model plant or crop produced for agricultural production.

Intervention: One of the following genome editing techniques was used to induce an alteration in the plant genome: Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated proteins (CRISPR/Cas), Transcription Activator-Like Effector Nucleases (TALENs), Meganucleases (MN), Zinc-Finger Nucleases (ZFN), Oligonucleotide-Directed Mutagenesis (ODM).

Outcome: The alteration of the genome (i.e. insertion, deletion or replacement of nucleotides) induced by the use of one of the genome editing techniques.

### **Secondary question two**

*“What is the available evidence for the potential occurrence of associated off-target effects due to the use of genome editing in model plants as well as in crops produced for agricultural production?”*

Population: Any model plant or crop produced for agricultural production.

Intervention: One of the following genome editing techniques was used to induce an alteration in the plant genome: Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated proteins (CRISPR/Cas), Transcription Activator-Like

Effector Nucleases (TALENs), Meganucleases (MN), Zinc-Finger Nucleases (ZFN), Oligonucleotide-Directed Mutagenesis (ODM).

Outcome: The occurrence of potential off-target events was assessed.

## 4.4 Methods

### 4.4.1 Search strategy

To test the comprehensiveness of the search strategy a scoping search was carried out to validate the search string and to test it against a priori selected articles of relevance (Additional file 1). To revise the adequacy of the search string an iterative process was applied by testing search strings in Web of Science (WoS), recording numbers of hits and testing them against the test library. The development of the search string is shown in additional file 2.

The search string will be composed of two parts: The first part defines the population of interest and comprises less specific terms like crop, plant or seed and in addition model plants and crops produced for agricultural production including their English and Latin names to ensure broad coverage. The second part defines the intervention, i.e. the genome editing technique applied to induce an alteration in the plant genome (CRISPR, TALENs, ZFN, MN or ODM).

The search terms describing each key element will be combined by the Boolean operator “OR” and the different key elements will be combined with the “AND” operator. Wildcards (‘\*’) will be used to search for variant word endings. The final search terms shown in additional file 3 will be adapted to the specific needs of each database to which it will be applied to. Database searches will be conducted in English and German language. Articles published after 1996, when the first study about a genome editing technique was published, will be considered.

The following online publication databases will be searched to identify academic literature. Access to these databases is ensured by institutional subscriptions:

- (i) Scopus
- (ii) PubMed
- (iii) Science direct
- (iv) Agris
- (v) Web of Science (WoS)
- (vi) Biological Abstracts
- (vii) BIOSIS Previews
- (viii) CAB Abstracts
- (ix) FSTA
- (x) SciELO Citation Index

Furthermore, Google Scholar (<https://scholar.google.com>) will be searched using 30 different combinations of the most relevant (model) plants and genome editing terms. The first 20 search

results, organized by relevance, of each combined search term will be assessed at the title/ abstract stage. Additionally, the search engine Google will be used to identify companies working with genome editing and to search on websites of government agencies for clues on the application of genome editing for market approval.

Furthermore, the references of each review article will be scanned for further relevant papers. All hits from each database will be imported into an EndNote X8.0.1 library file. Duplicates will be removed using the appropriate function within the EndNote software. After removing duplicates the remaining records will be imported into the open-access and non-profit database CADIMA [41] to increase transparency and traceability during the review process.

#### **4.4.2 Study inclusion criteria**

In order to be included in the systematic map each article has to meet all the following inclusion criteria:

Relevant population: Any model plant or crop produced for agricultural production as well as higher fungi was used. Ornamental and medicinal plants as well as yeast will be excluded.

Relevant intervention: At least one of the following genome editing techniques was used to induce an alteration in the plant genome: Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated proteins (CRISPR/Cas), Transcription Activator-like Effector Nuclease (TALENs), Zinc-Finger Nuclease (ZFN), Meganuclease (MN), Oligonucleotide-Directed Mutagenesis (ODM)

Relevant outcome: Due to the use of a genome editing technique an alteration in the plant genome was reported (insertion, deletion or replacement). Other techniques which do not induce a DSB and therewith do not employ the DNA repair mechanism of the cell will be excluded; among which are TALE (without a nuclease) and variations of a non-functional dCas9 (deadCas9) fused to a methylase, demethylase or transcription factor.

Primary data: Only those references will be included which comprise primary data referring to the use of a genome editing technique to induce a sequence alteration in the plant genome. If there is any doubt about the availability of primary data of an article on title/abstract stage, it will be kept for full text assessment.

#### **4.4.3 Article screening**

When applying the selection criteria at title/abstract stage, a consistency check will be conducted by all participating reviewers aiming to determine the inter-reviewer agreement. A minimum of 50 references or 10% of the total number up to a maximum of 200 references retrieved by the research will be checked until a kappa value with a score of at least 0.6 indicates a good reviewer agreement. If the kappa value is below 0.6, the reviewer will analyze the reasons for the insufficient kappa value within the whole review team and reassess the inclusion criteria. In a first step, all identified records will be assessed at title/abstract stage. In case that insufficient information is provided, the records will be passed on to the full text stage. Afterwards, the eligibility of records being retained after title/abstract screening will be checked at full text stage. A list of articles excluded at full text stage with the reason for exclusion will be provided.

At title/abstract and at full text stage, two reviewers working independently from each other will apply the inclusion criteria to all articles. Inconsistencies in rating decisions will be documented and the reasons will be discussed in the review team.

#### 4.4.4 Study quality assessment

The aim of this systematic map is to provide a broad overview of the current progress in modifying the plant genome as well as the occurrence of off-target effects observed due to the use of genome editing techniques. Therefore, a full critical appraisal of included studies will not be performed. In order to facilitate the decision whether a systematic review would be worthwhile for being performed on a specific section of the map, data being indicative for the validity of an included study will be extracted (e.g. search for off-targets, off-target detection method).

#### 4.4.5 Data coding strategy

For each study, one reviewer will extract data and the extracted data will be cross-checked by a second reviewer to minimize human error.

The following superordinate categories will be considered:

1. Bibliographic information (Reference type, authors, year of publication, title, abstract, keywords, periodical, issue number, page range, volume, DOI/ISBN, corresponding author and the name of the country the corresponding author is located).
2. Information answering secondary question one about traits modified by genome editing (genome editing technique, plant species, sequence identifier, trait, kind of modification, progress in research towards application, key topic).
3. Information answering secondary question two about the occurrence of off-target effects due to the use of genome editing (search for off-targets, prediction of potential off-targets (in silico), the prediction method used, identification of potential off-targets, detection of off-targets (biased/ unbiased), detection method, number of identified off-targets (biased/ unbiased)).

More detailed information about the data extraction strategy and the data extraction mode are shown in additional file 4. In a pilot scheme 5% of the studies retained for data extraction (at least 20) will be checked a priori by all reviewers aiming to assess repeatability of the extraction process.

#### 4.4.6 Study mapping and presentation

Included studies will be catalogued in a searchable database as well as in an Excel file. The database will be freely accessible on the project webpage [www.dialog-gea.de](http://www.dialog-gea.de). In addition, eligible studies will be characterized using descriptive statistics on key trends, including:

- (i) Frequency distribution of countries which are working on genome editing.
- (ii) Frequency of different plants used for the application of genome editing.
- (iii) Frequency distribution of the use of each genome editing technique.
- (iv) Frequency for the additional analysis of off-target effects when using a genome editing technique.

- (v) Frequency distribution of the method used to predict/ detect off-target effects.
- (vi) How often off-targets were detected dependent on the used detection method.

Knowledge gaps (un- or underrepresented subtopics that warrant further primary research) and knowledge clusters (well-represented subtopics that are amenable to full synthesis by a systematic review) will be identified e.g. by cross-tabulating key meta-data variables in heat maps. Furthermore, all results gained within this systematic map will be summarized in a narrative report. Moreover, additional files will include:

- (i) An EndNote database of all studies included in the systematic map.
- (ii) An overview of all studies excluded at full text stage with the reason for exclusion.

#### **Additional files**

<https://environmentalevidencejournal.biomedcentral.com/articles/10.1186/s13750-018-0130-6#Sec19>

## **4.5 References**

1. Ronald P. Plant genetics, sustainable agriculture and global food security. *Genetics*. 2011; 188:11–20. doi:10.1534/genetics.111.128553.
2. Becker H. *Pflanzenzüchtung*. 2nd ed. Stuttgart: UTB; 2011.
3. FAO - COMMITTEE ON AGRICULTURE (17th Session). 06.03.2003. <http://www.fao.org/docrep/meeting/006/y8704e.htm>. Accessed 5 Mar 2018.
4. Sovová T, Kerins G, Demnerová K, Ovesná J. Genome Editing with Engineered Nucleases in Economically Important Animals and Plants: State of the Art in the Research Pipeline. *Current Issues in Molecular Biology*. 2017:41–62. doi:10.21775/cimb.021.041.
5. Schlegel RHJ. *Dictionary of plant breeding*. 2nd ed. Boca Raton, Fla.: CRC Press; 2009.
6. Barton KA, Binns AN, Matzke AJM, Chilton M-D. Regeneration of intact tobacco plants containing full length copies of genetically engineered T-DNA, and transmission of T-DNA to R1 progeny. *Cell*. 1983;32:1033–43. doi:10.1016/0092-8674(83)90288-X.
7. Herrera-Estrella L, Depicker A, van Montagu M, Schell J. Expression of chimaeric genes transferred into plant cells using a Ti-plasmid-derived vector. *Nature*. 1983;303:209–13. doi:10.1038/303209a0.
8. Hartung F, Schiemann J. Precise plant breeding using new genome editing techniques: Opportunities, safety and regulation in the EU. *Plant J*. 2014;78:742–52. doi:10.1111/tpj.12413.
9. Sprink T, Eriksson D, Schiemann J, Hartung F. Regulatory hurdles for genome editing: Process- vs. product-based approaches in different regulatory contexts. *Plant Cell Rep*. 2016;35:1493–506. doi:10.1007/s00299-016-1990-2.
10. Govindan G, Ramalingam S. Programmable Site-Specific Nucleases for Targeted Genome Engineering in Higher Eukaryotes. *J Cell Physiol*. 2016;231:2380–92. doi:10.1002/jcp.25367.
11. SAM High Level Group of Scientific Advisors. *New techniques in Agricultural Biotechnology*. 2017. ISBN 978-92-79-66222-5. <https://doi.org/10.2777/57449> 8.

- <https://ec.europa.eu/research/sam/index.cfm?pg=agribiotechology#>. Accessed 18 July 2017.
12. European Food Safety Authority. Scientific opinion addressing the safety assessment of plants developed using Zinc Finger Nuclease 3 and other Site-Directed Nucleases with similar function. *EFSA J.* 2012;10(10):2943.
  13. van de Wiel CCM, Schaart JG, Lotz LAP, Smulders MJM. New traits in crops produced by genome editing techniques based on deletions. *Plant Biotechnol Rep.* 2017;11:1–8. doi:10.1007/s11816-017-0425-z.
  14. Podevin N, Davies HV, Hartung F, Nogué F, Casacuberta JM. Site-directed nucleases: A paradigm shift in predictable, knowledge-based plant breeding. *Trends Biotechnol.* 2013;31:375–83. doi:10.1016/j.tibtech.2013.03.004.
  15. Puchta H, Fauser F. Synthetic nucleases for genome engineering in plants: Prospects for a bright future. *Plant J.* 2014;78:727–41. doi:10.1111/tpj.12338.
  16. Rinaldo AR, Ayliffe M. Gene targeting and editing in crop plants: A new era of precision opportunities. *Mol Breeding.* 2015;35:1126. doi:10.1007/s11032-015-0210-z.
  17. Roth N, Klimesch J, Dukowicz-Schulze S, Pacher M, Mannuss A, Puchta H. The requirement for recombination factors differs considerably between different pathways of homologous double-strand break repair in somatic plant cells. *Plant J.* 2012;72:781–90. doi:10.1111/j.1365-313X.2012.05119.x.
  18. Djukanovic V, Smith J, Lowe K, Yang M, Gao H, Jones S, et al. Male-sterile maize plants produced by targeted mutagenesis of the cytochrome P450-like gene (MS26) using a re-designed I-CreI homing endonuclease. *Plant J.* 2013;76:888–99. doi:10.1111/tpj.12335.
  19. D'Halluin K, Vanderstraeten C, van Hulle J, Rosolowska J, van den Brande I, Pennewaert A, et al. Targeted molecular trait stacking in cotton through targeted double-strand break induction. *Plant Biotechnol J.* 2013;11:933–41. doi:10.1111/pbi.12085.
  20. Kim Y-GG, Cha J, Chandrasegaran S. Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A.* 1996;93:1156–60.
  21. Boch J, Bonas U. Xanthomonas AvrBs3 family-type III effectors: Discovery and function. *Annu Rev Phytopathol.* 2010;48 (2010):419–36. doi:10.1146/annurev-phyto-080508-081936.
  22. Khandagale K, Nadaf A. Genome editing for targeted improvement of plants. *Plant Biotechnol Rep.* 2016;10:327–43. doi:10.1007/s11816-016-0417-4.
  23. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* 2012;337:816–21. doi:10.1126/science.1225829.
  24. Subburaj S, Tu L, Jin Y-T, Bae S, Seo PJ, Jung YJ, Lee G-J. Targeted genome editing, an alternative tool for trait improvement in horticultural crops. *Hortic. Environ. Biotechnol.* 2016;57:531–43. doi:10.1007/s13580-016-0281-8.
  25. Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science.* 2014;346:1258096. doi:10.1126/science.1258096.
  26. Hilscher J, Bürstmayr H, Stoger E. Targeted modification of plant genomes for precision crop breeding. *Biotechnol J.* 2017;12:1–14. doi:10.1002/biot.201600173.

27. Lusser M, Parisi C, Plan D, Rodriguez-Cerezo E. New plant breeding techniques. 2011. State-of-the-art and prospects for commercial development. 2011. <http://ipts.jrc.ec.europa.eu/publications/pub.cfm?id=4100>. Accessed 22 Aug 2017.
28. Nationale Akademie der Wissenschaften Leopoldina, Deutsche Forschungsgemeinschaft, acatech – Deutsche Akademie der Technikwissenschaften, Union der deutschen Akademien der Wissenschaften. Chancen und Grenzen des genome editing/The opportunities and limits of genome editing. 2015; ISBN: 978-3-8047-3493-7
29. Duan Y-B, Li J, Qin R-Y, Xu R-F, Li H, Yang Y-C, et al. Identification of a regulatory element responsible for salt induction of rice OsRAV2 through ex situ and in situ promoter analysis. *Plant Mol Biol*. 2016;90:49–62. doi:10.1007/s11103-015-0393-z.
30. Demorest ZL, Coffman A, Baltes NJ, Stoddard TJ, Clasen BM, Luo S, et al. Direct stacking of sequence-specific nuclease-induced mutations to produce high oleic and low linolenic soybean oil. *BMC Plant Biol*. 2016;16:225. doi:10.1186/s12870-016-0906-1.
31. Chandrasekaran J, Brumin M, Wolf D, Leibman D, Klap C, Pearlsman M, et al. Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol Plant Pathol*. 2016;17:1140–53. doi:10.1111/mpp.12375.
32. Gocal GFW, Schöpke C, Beetham PR. Oligo-Mediated Targeted Gene Editing. In: *Advances in New Technology for Targeted Modification of Plant Genomes*. 2015. p. 73–89. doi:10.1007/978-1-4939-2556-8\_5.
33. Shi J, Gao H, Wang H, Lafitte HR, Archibald RL, Yang M, et al. ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. *Plant Biotechnol J*. 2017;15:207–16. doi:10.1111/pbi.12603.
34. Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, Qiu J-L. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat Biotechnol*. 2014;32:947–51. doi:10.1038/nbt.2969.
35. Barakate A, Stephens J. An Overview of CRISPR-Based Tools and Their Improvements: New Opportunities in Understanding Plant-Pathogen Interactions for Better Crop Protection. *Front Plant Sci*. 2016;7:765. doi:10.3389/fpls.2016.00765.
36. Xue C, Seetharam AS, Musharova O, Severinov K, J. Brouns SJ, Severin AJ, Sashital DG. CRISPR interference and priming varies with individual spacer sequences. *Nucleic Acids Res*. 2015;43:10831–47. doi:10.1093/nar/gkv1259.
37. Tycko J, Myer VE, Hsu PD. Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity. *Mol Cell*. 2016;63:355–70. doi:10.1016/j.molcel.2016.07.004.
38. Agapito-Tenfen SZ, Wikmark O-G. Current status of emerging technologies for plant breeding. Biosafety and knowledge gaps of site directed nucleases and oligonucleotide-directed mutagenesis. *GenØk - Centre for Biosafety*. 2015.
39. Zhang X-H, Tee LY, Wang X-G, Huang Q-S, Yang S-H. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol Ther Nucleic Acids*. 2015;4:e264. doi:10.1038/mtna.2015.37.
40. Kanchiswamy CN, Maffei M, Malnoy M, Velasco R, Kim J-S. Fine-Tuning Next-Generation Genome Editing Tools. *Trends Biotechnol*. 2016;34:562–74. doi:10.1016/j.tibtech.2016.03.007.

41. Kohl C, McIntosh EJ, Unger S, Haddaway NR, Kecke S, Schiemann J, Wilhelm R. Online tools supporting the conduct and reporting of systematic reviews and systematic maps: A case study on CADIMA and review of existing tools. *Environ Evid.* 2018;7:2420. doi:10.1186/s13750-018-0115-5

## 5 Systematic map

---

### **What is the available evidence for the range of applications of genome editing as a new tool for plant trait modification and the potential occurrence of associated off-target effects: A systematic map**

---

Chapter 5 has been published:

**Modrzejewski, Dominik**<sup>1</sup>; Hartung, Frank<sup>1</sup>; Sprink, Thorben<sup>1</sup>; Krause, Dörthe<sup>1</sup>; Kohl, Christian<sup>1</sup>; Wilhelm, Ralf<sup>1</sup> (2019): What is the available evidence for the range of applications of genome editing as a new tool for plant trait modification and the potential occurrence of associated off-target effects: a systematic map. In: *Environ Evid* 8 (1), S. 11. DOI: 10.1186/s13750-019-0171-5.

<sup>1</sup>Institute for Biosafety in Plant Biotechnology, Federal Research Centre for Cultivated Plants, Julius Kühn-Institut, Erwin-Baur-Straße 27, 06484 Quedlinburg, Germany

#### **Authors' contributions**

The review process was coordinated by **DM**. **DM**, FH, TS and DK conducted the screening of articles and data extraction. **DM** drafted the report. FH, TS, CK and RW assisted in editing and revising the draft. All authors read and approved the final manuscript.

## 5.1 Abstract

### Background

Within the last decades, genome editing techniques such as CRISPR/Cas, TALENs, Zinc-Finger Nucleases, Meganucleases, Oligonucleotide-Directed Mutagenesis and base editing have been developed enabling a precise modification of DNA sequences. Such techniques provide options for simple, time-saving and cost-effective applications compared to other breeding techniques and hence genome editing has already been promoted for a wide range of plant species. Although the application of genome editing induces less unintended modifications (off-targets) in the genome compared to classical mutagenesis techniques, off-target effects are a prominent point of criticism as they are supposed to cause unintended effects, e.g. genomic instability or cell death. To address these aspects, this map aims to answer the following question: What is the available evidence for the range of applications of genome editing as a new tool for plant trait modification and the potential occurrence of associated off-target effects? This primary question will be considered by two secondary questions: One aims to overview the market-oriented traits being modified by genome editing in plants and the other explores the occurrence of off-target effects.

### Methods

A literature search in nine bibliographic databases, Google Scholar, and 47 web pages of companies and governmental agencies was conducted using predefined and tested search strings in English language. Articles were screened on title/abstract and full text level for relevance based on predefined inclusion criteria. The relevant information of included studies were mapped using a predefined data extraction strategy. Besides a descriptive summary of the relevant literature, a spreadsheet containing all extracted data is provided.

### Results

Altogether, 555 relevant articles from journals, company web pages and web pages of governmental agencies were identified containing 1328 studies/applications of genome editing in model plants and agricultural crops in the period January 1996 to May 2018. Most of the studies were conducted in China followed by the USA. Genome editing was already applied in 68 different plants. Although most of the studies were basic research, 99 different market-oriented applications were identified in 28 different crops leading to plants with improved food and feed quality, agronomic value like growth characteristics or increased yield, tolerance to biotic and abiotic stress, herbicide tolerance or industrial benefits. 252 studies explored off-target effects. Most of the studies were conducted using CRISPR/Cas. Several studies firstly investigated whether sites in the genome show similarity to the target sequence and secondly analyzed these potential off-target sites by sequencing. In around 3% of the analyzed potential off-target sites, unintended mutations were detected. Only a few studies conducted off-target analysis using unbiased detection methods (e.g. whole genome sequencing). No off-target effects that could be correlated to the genome editing process were identified in these studies.

### Conclusions

The rapid adoption in plant breeding was demonstrated by a considerable number of market oriented applications (crops and traits) described in publications worldwide. Studies investigating

off-target effects are very heterogeneous in their structure and design. Therefore, an in-depth assessment regarding their weight of evidence is mandatory.

## 5.2 Background

Technological progress in agriculture and plant breeding has contributed significantly to a stable food supply and formed the basis for high yields as well as the production of high-quality agricultural products [1]. However, in an ever-changing world, new challenges will be encountered within the next decades. Aside the demands of the growing global population and limited fossil resources, climate change is a driver of breeding efforts as it is associated with increased extreme weather events like droughts or floods as well as changing dynamics of pests and diseases. Agriculture needs to ensure and increase the world agricultural production to serve extended demands with limited environmental resources like soil and water [2]. In contrast, intensification of agriculture causes considerable impact on nature, as naturally diverse landscapes are replaced by arable land for the cultivation of few plant species. Biodiversity is threatened through habitat loss and pesticide use, which in turn is considered to increase disease and pest pressure [1]. Additionally, these impacts of agriculture on the environment are becoming increasingly important in societal debates.

To meet all these challenges, improved crop varieties may be developed and integrated into a sustainable farming system considering their economic, environmental and social impacts [2]. Examples for a sustainable farming system are the preservation of natural soil fertility through suitable crop rotations, fertilization and plant protection according to the principles of integrated cultivation or a balanced and species-appropriate animal husbandry [3]. Additionally, plant breeding is of crucial importance to manage environmental impacts on cultivation systems by providing varieties resistant to plant diseases or pests, tolerant to abiotic stress and more broadly support a “greener production” – in time. This may reduce pesticide use and result in less intense management efforts (e.g. irrigation) [4]. Further yield improvement can increase the yield per hectare and may open land management options e.g. balance with areas for nature conservation [1]. Nevertheless, there is evidence that increased yields may lead to a rebound-effect meaning that yield improvement conserves the rate of land clearance, but the effect is smaller than it could be [5]. One example for the rebound-effect in agriculture is the Green Revolution [6]. As a result of this revolution, yields increased, saving ecosystems from conversion to agricultural land. However, the effect was much smaller than expected. One explanation for this effect may be that increased productivity due to new technologies also increases the profitability of agriculture compared with alternative land use [6].

Plant breeding essentially relies on the utilization of genetic variation within the breeding material that can be used for crossing and selection to develop improved varieties. New genetic variation occurs naturally by spontaneous mutations that enable some individual plants in a population to adapt to changing environmental conditions. However, as the mutation rate is fairly low and random, plant breeders and scientists have artificially induced mutations for several decades [7]. The first generation of mutation breeding used chemical and physical mutagens to generate a plurality of nonspecific mutations. The increased mutation rate results in plants with a few positive, a lot of neutral and several negative characteristics. Thus, laborious backcrossing and selection steps are necessary in order to select for a target trait in a desired genetic background (maternal

variety). Nevertheless, to date 3282 mutant varieties from 225 different plant species have been generated through undirected mutagenesis and officially but voluntarily registered [8].

**Glossary: [1, 9, 10]**

**Backcrossing:**

Backcrossing is a crossing of a hybrid with one of its parents in order to achieve offspring that are genetically closer to the selected parent. This way, desired heterologous traits from the hybrid can be transferred into the genetic background of a parental line. Since crossing recombines all genes, many backcrosses are necessary to achieve considerable dilution of unwanted genes from the hybrid.

**Mutagen:**

A mutagen is an agent that increases the mutation rate within an organism or cell, e.g. X-rays, gamma-rays or chemicals like ethyl methane sulfonate (EMS).

**Mutation breeding:**

A plant breeding approach using mutagens to enhance genetic variation. The resulting random mutations can generate new gene variations with positive traits that can be selected for further breeding. However, several of these mutations are negative and diminish the viability of the plant.

**Off-target effect due to genome editing:**

Unintended cleavage and mutations at untargeted genomic sites with similar but not identical sequences compared to the target site.

**Selection:**

A process in breeding by which the breeder chooses only those individuals that show desired trait(s).

In 1983, the first recombinant DNA was delivered to plant cells using *Agrobacterium tumefaciens* [11, 12]. From this time on, it is possible to work at a single gene level with genetic material from any organism, generating plants that cannot be bred conventionally. Nevertheless, the induced mutations using chemical and physical mutagens as well as the “classical” transgenic approach show limited efficiencies and unintended side effects due to the random targeting [13].

In recent years, genome editing techniques have been developed enabling a more precise modification of DNA sequences in a site-directed manner [10]. To date, genome editing comprises three molecular approaches that efficiently induce targeted alterations in genomes: (i) Site-directed nucleases (SDN), including Meganucleases (MN), Zinc-Finger Nucleases (ZFN), Transcription Activator-Like Effector Nucleases (TALENs) and Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated protein (CRISPR/Cas), (ii) Oligonucleotide-directed mutagenesis (ODM) and (iii) base editing (BE). A detailed description of the single techniques is summarized in Additional file 1. Site-directed nucleases induce double-strand breaks (DSBs) in the DNA which are subsequently repaired by the autochthonous cellular mechanisms. The type of repair can be categorized in three main types [10, 14–16]:

- (i) No template is added and the DSB is repaired by autochthonous cellular mechanisms (in most cases non-homologous end joining (NHEJ)) resulting in small insertion-deletion (indel) mutations. This approach is defined as SDN1.
- (ii) A repair template is added which, except for a few nucleotides, is identical to the sequences in which the DSB is introduced. Then, the DSB is repaired via homology-directed repair (HDR), causing nucleotide substitution or, depending on the template used, targeted indels of a specific size. This approach is defined as SDN2.
- (iii) The repair template harbors a recombinant DNA sequence additional to the homologous sequences in which the DSB is made and the break is repaired via HDR, resulting in more complex alterations, i.e. the insertion of foreign genes. This approach is defined as SDN3.

The Oligonucleotide-Directed Mutagenesis technique does not induce a DSB. Instead the introduced oligonucleotide binds to the targeted DNA sequences and this sequence is then modified by the cellular mechanism of mismatch repair [10]. Base editing is a recently developed approach enabling a targeted switch of DNA bases in a given frame into another without any DSB [17].

Genome editing offers substantial advantages compared to previous mutation breeding techniques and conventional genetic engineering in terms of speed and precision. It provides the opportunity to selectively mutate or modify one or a few genes (SDN1, SDN2). In addition, it is now possible to precisely modify or selectively replace (SDN3) entire genes from both closely as well as distantly related organisms [10]. By the use of conventional genetic engineering, traces of recombinant DNA, from the used gene shuttle (e.g. bacteria, virus or plasmid), persist in the modified organism leading to clearly characterized genetically modified organisms. In contrast, by applying genome editing, it is possible to modify crops without inserting foreign DNA sequences at all [18]. Therefore, some countries like USA, Canada, Brasilia, Argentina and others reduced the regulatory burden for plant breeders [19]. Due to the simplicity, time-saving and cost-effective application of genome editing, it has already been applied in a wide range of cultivars. Genome editing has been used for:

- (i) Analyzing gene functions (e.g. effect of the RAV2 gene for salt stress in rice [20]).
- (ii) Improvement of product quality (e.g. improved oil quality in soybean [21]).
- (iii) Development of disease resistant varieties (e.g. virus resistant cucumber [22]).
- (iv) Developing of herbicide tolerant varieties (e.g. resistance to the herbicide chlorsulfuron in canola [23]).
- (v) Improved adaption to abiotic stress, (e.g. drought tolerance in maize [24]).

Even in plants like hexaploid wheat, that were so far largely inaccessible for targeted genetic alterations, the simultaneous mutation of all six alleles was successfully performed [25]. These successful applications are opening up new dimensions for the scientific plant breeding and agricultural community.

Compared to randomly induced mutations by chemicals or irradiation, the number of unintended mutations (off-target effects) is broadly reduced by genome editing techniques [10]. Nevertheless, their application does not completely or *per se* exclude the occurrence of off-target effects. Off-

target effects caused by genome editing most likely occur in DNA sequences that are similar (not identical) to the targeted one but are located at another site in the genome. Mainly, they occur due to the lack of exclusiveness and/or length of the recognition site [10]. When analyzing off-target effects, one distinguishes between biased and unbiased detection methods [26]. To date, the predominant approach for identifying off-target effects is the biased approach consisting of two steps: (i). Using sequence alignment programs, sites in the genome with high similarity to the target sequence are identified, which are designated as potential off-target sites. Several different tools like BLAST [27], Cas-OFFinder [28] or CRISPR-P [29, 30] are used to identify these potential off-target sites. (ii). The identified individual DNA sequences (potential off-target sites) are then analyzed for undesired mutations (off-target effects) using various detection methods like mismatch-sensitivity endonuclease assays, Sanger sequencing or targeted deep sequencing [31]. All detection methods have their specific advantages and disadvantages which are addressed in several reviews (e.g. [26, 32]). In contrast to the biased detection methods, unbiased ones are used to identify off-target effects in a completely unrestricted way. Therefore, it requires genome-wide sequencing to identify off-target mutations anywhere in the genome and *de novo* define off-target sites [26, 32]. Depending on the detection method being used, the results of identifying off-target mutations vary widely. Although genome editing techniques induce much less off-target effects compared to classical mutagenesis techniques, these are an important point of criticism as they may possibly cause genomic instability, cytotoxicity and cell death [33–35].

This systematic map facilitates an objective debate by informing interested stakeholder communities in a transparent and retraceable manner about the status of research, the progress of genome editing in plants and the available evidence for the potential occurrence of associated off-target effects. Furthermore, risk assessors and decision makers are depending on the provision of a reliable body of evidence to support conclusions about potential risks being associated with the application of genome editing. Thus, an overview of the available evidence on the occurrence of off-target effects could be of crucial importance.

### **Stakeholder engagement**

The systematic map question, the secondary questions and the scope was designed by the review team reflecting discussions with policy makers, authorities, regulators and academia requesting a broad overview on the available evidence about the application of genome editing in plants and the potential occurrence of off-target effects. Throughout the review process there was no stakeholder engagement. As indicated in the systematic map protocol the results of the map were discussed on a conference with different stakeholders from this field, including besides others plant breeders, federal authorities, academia, farmer organizations and processing industry. Stakeholder remarks are taken into account when preparing a systematic review based on the results of this map.

## **5.3 Objectives of the map**

As genome editing techniques are a promising tool to revolutionize plant breeding, they are of particular relevance to scientists, breeders, farmers but also to decision and policy makers with regards to the broader agricultural management and future challenges. Therefore, we wanted to provide a comprehensive and transparent overview of the available evidence base concerning the effects of genome editing in plants. The main objectives were:

- (i) Overview of the traits modified by genome editing in model plants as well as in crops produced for agricultural production.
- (ii) Overview of the available evidence about the occurrence of off-target effects due to the use of genome editing techniques in model plants as well as in crops produced for agricultural production.
- (iii) Identification of the geographical distribution of genome editing activities in plants worldwide.
- (iv) Identification of the volume of the available literature, evidence clusters and key characteristics of the evidence base to inform interested stakeholder communities.
- (v) Identification of knowledge gaps concerning the occurrence of off-target effects in order to inform decision makers which future research might be needed for a risk assessment.
- (vi) Assessment whether a specific section of the available evidence base is suitable for an in-depth analysis by a systematic review.

The primary question of the systematic map was: *“What is the available evidence for the range of applications of genome editing as a new tool for plant trait modification and the potential occurrence of associated off-target effects”?*

To answer this primary question, it was subdivided into two secondary questions related to (1) the traits modified by genome editing and (2) the occurrence of off-target effects due to the use of genome editing.

### **Secondary question one**

*“What are the traits modified by genome editing in model plants as well as in crops produced for agricultural production?”*

**Population:** Any model plant or crop produced for agricultural production.

**Intervention:** One of the following genome editing techniques was used to induce an alteration in the plant genome: Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated protein (CRISPR/Cas), Transcription Activator-Like Effector Nucleases (TALENs), Meganucleases (MN), Zinc-Finger Nucleases (ZFN), Oligonucleotide-Directed Mutagenesis (ODM), base editing (BE)

**Outcome:** The alteration of the genome (i.e. insertion, deletion or substitution of nucleotides) induced by the use of a genome editing technique.

### **Secondary question two**

*“What is the available evidence for the potential occurrence of associated off-target effects due to the use of genome editing in model plants as well as in crops produced for agricultural production?”*

**Population:** Any model plant or crop produced for agricultural production.

**Intervention:** One of the following genome editing techniques was used to induce an alteration in the plant genome: Clustered Regularly Interspaced Short Palindromic

Repeats/CRISPR associated protein (CRISPR/Cas), Transcription Activator-Like Effector Nucleases (TALENs), Meganucleases (MN), Zinc-Finger Nucleases (ZFN), Oligonucleotide-Directed Mutagenesis (ODM), base editing (BE).

**Outcome:** The alteration of the genome (i.e. insertion, deletion or substitution of nucleotides) induced by the use of a genome editing technique. Additionally, the occurrence of off-target effects was assessed.

## 5.4 Methods

The methods used to conduct this systematic map were based on the Collaboration for Environmental Evidence (CEE) systematic review guidelines [36]. Detailed information about the methods used to perform this systematic map are presented in the published protocol [37]. A brief summary of these methods is provided here.

### Search for articles

The search string was composed of two parts: The first part defined the population of interest comprising less specific terms like crop, plant or seed as well as specific model plants and crops including their English and Latin names. The second part defined the intervention, i.e. the genome editing technique applied to induce an alteration in the plant genome (CRISPR, TALENs, ZFN, MN, ODM or BE). To test the comprehensiveness of the search strategy a scoping search was carried out and the identified records were tested against an *a priori* defined test library with articles of known relevance. Details on search settings and subscriptions can be found in Additional file 2. The following bibliographic databases were searched whereby the search string was adapted to the specific needs of each database to which it was applied to:

- (i) Scopus
- (ii) PubMed
- (iii) Science direct
- (iv) Agris
- (v) Web of Science (WoS)
- (vi) Biological Abstracts
- (vii) BIOSIS Previews
- (viii) CAB Abstracts
- (ix) SciELO Citation Index

In addition, Google Scholar (<https://scholar.google.com>) was searched using 30 different combinations of different (model) plants and genome editing terms. The first 20 hits organized by relevance, of each search term were examined for relevance. Furthermore, a total amount of 47 web pages of companies working with genome editing and the USDA database “Am I regulated?” (<https://www.aphis.usda.gov/aphis/ourfocus/biotechnology/am-i-regulated>) were searched to identify grey literature. Finally, the bibliographies of 107 review articles identified by the literature search were screened to identify further relevant papers. In May 2018, the search string was applied

in order to identify articles published after 1996, when the first study about a genome editing technique was published.

## **Article screening and study eligibility criteria**

### **Screening process**

Before applying the selection criteria at title/abstract and then at full text level, a consistency check was conducted by all four participating reviewers aiming to determine the inter-reviewer agreement. The level of agreement was tested formally using a kappa test [38]. 100 references retrieved by the search were randomly explored at title/abstract level leading to a kappa value of 0.48. After discussing all disagreements, a second consistency check was carried out using another randomly allocated 100 references resulting in a kappa value of 0.71. After title/abstract screening, potentially relevant articles were checked at full text level. A list of unobtainable articles (Additional file 3) and articles excluded at full text level with the reason for exclusion (Additional file 4) are provided. For conducting the consistency check as well as for the screening process at title/abstract and full text level the open-access and non-profit database CADIMA was used [39]. Two members of the review team are authors of a few articles retrieved by the review process. However, as none of these papers comprise primary data, their articles were excluded at title/abstract level. Nevertheless, the two coauthors routinely screened literature covering the issues addressed in this map.

### **Eligibility criteria**

An article had to meet all the following inclusion criteria in order to enter into the systematic map:

Eligible populations: Any model plant or crop produced for agricultural production as well as higher fungi was used.

Eligible interventions: At least one of the following genome editing techniques was used to induce an alteration in the plant genome: Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated protein (CRISPR/Cas), Transcription Activator-like Effector Nuclease (TALENs), Zinc-Finger Nuclease (ZFN), Meganuclease (MN), Oligonucleotide-Directed Mutagenesis (ODM), base editing (BE).

Eligible outcome for secondary question one: An alteration in the plant genome was reported (insertion, deletion or substitution) due to the use of a genome editing technique.

Eligible outcome for secondary question two: The occurrence of off-target effects was assessed.

Eligible type of data: Only those references were included which comprise primary data referring to the use of a genome editing technique to induce a sequence alteration in the plant genome.

Eligible languages: References in German and English languages were included. Articles in other languages were included when besides title and abstract, further parts of the article, like figures or tables, were in English or German and the provided information allowed for a definite judgment of their relevance.

### **Study validity assessment**

The aim of this systematic map was to provide a broad overview of the progress on genome editing in plants as well as the examination of associated off-target effects. The validity of the included studies (critical appraisal) was not assessed. However, in order to facilitate the decision on the potential of a subsequent systematic review, data being indicative for the validity of an included study were extracted.

### **Data coding strategy**

Articles in which several genome editing techniques were applied, different plants were used or different genes were addressed have been subdivided in distinct studies. While articles were screened for relevance at title/abstract and full text level, the relevant data were finally extracted at study level.

The data of each included study was extracted in one row in the excel file for the following superordinate categories:

1. Bibliographic information (reference type, authors, year of publication, title, abstract, keywords, periodical, issue number, page range, volume, DOI/ISBN, corresponding author and the name of the country the corresponding author is located).
2. Information answering secondary question one about traits modified by genome editing (genome editing technique, plant species, sequence identifier, trait, type of alteration, progress in research, key topic).
3. Information answering secondary question two about the occurrence of off-target effects due to the use of genome editing (search for off-target effects, prediction of potential off-target effects (in silico), the prediction method used, identification of potential off-target sites, detection of off-target effects (biased/ unbiased), detection method, amount of identified off-target effects (biased/ unbiased)).

Data were extracted according to the systematic map protocol [37]. Data from a subset of 20 studies were extracted by two reviewers independently to assess the consistency of the extraction process across the reviewers. Discrepancies were discussed and clarified within the whole review team. Then, the data of the other studies were extracted by one reviewer and cross-checked by a second one to minimize the introduction of human error. In case of missing data, “no information” has been noted under the respective category. If data were in another language than English or German, it is stated as “language”.

### **Data mapping method**

The overview of research activities is provided in a narrative report and visualized in tables and figures. In addition, relevant studies and the extracted data are catalogued in Additional file 5 (to answer secondary question one) and Additional file 6 (to answer secondary question two) as well as provided in a searchable database that is freely accessible on the web page <https://www.dialog-gea.de/de/service/repositorium>.

### **Deviations from the systematic map protocol**

The methods used in this map deviate from the protocol in the following aspects:

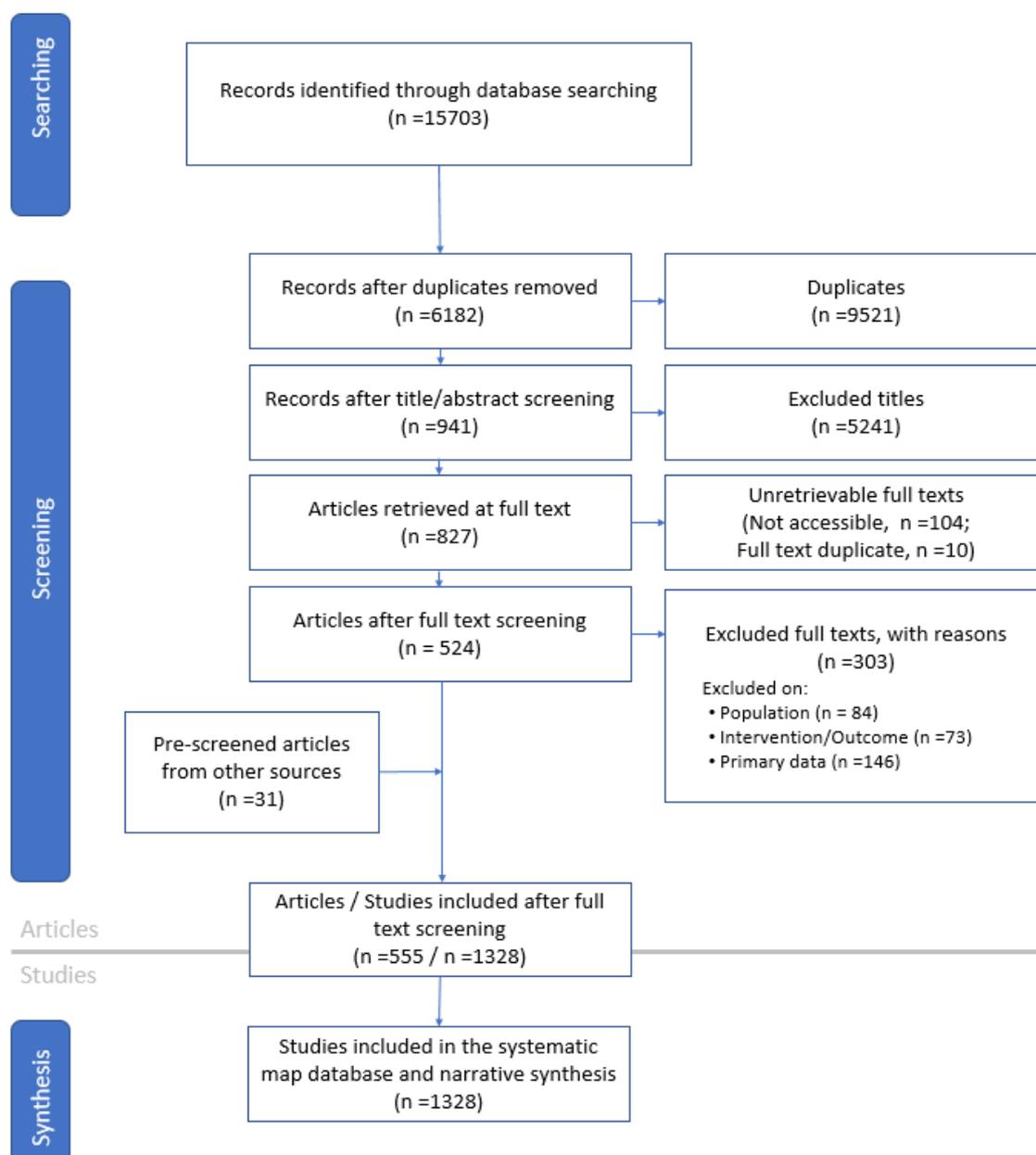
- (i) Contrary to what was stated in the protocol, the recently developed base editing method was not excluded. To identify all base editing studies the second part of the search string (intervention) was extended by the terms “base editing” and “base editing”.
- (ii) Articles in other languages than English or German were included if further parts than title and abstract of the article like figures or tables were in English or German and the provided information allowed for a definite judgment of their relevance.
- (iii) In the course of data extraction we noticed that it was not possible to properly categorize the progress in research to the three classes indicated in the systematic map protocol [37]. Therefore, we decided to classify the studies as either basic research or market-oriented application. To be flagged as market-oriented, a study had to meet three criteria:
  - (1) Genome editing was applied in an agricultural crop.
  - (2) A trait was addressed that may be of interest for commercialization (market-oriented trait).
  - (3) The targeted trait is expressed in the edited plant grown.

All studies that did not meet all three criteria were classified as basic research.
- (iv) Contrary to what was stated in the protocol, no EndNote database of all studies included in the systematic map was attached to this systematic map. Instead, one excel file for each secondary question was provided as additional file containing all included studies and the extracted data.

## 5.5 Results

### Review descriptive statistics

Figure 5.1 presents the systematic mapping process of articles and studies in a flow diagram. From January 1996 until May 2018, in total 15703 records were identified from ten bibliographic databases, Google Scholar, the targeted search on 47 company web pages, the USDA-database “Am I regulated?” and the screening of 107 review articles. After removing duplicates (n=9521), 6182 articles remained and were screened on title/abstract level. Main reasons for exclusion at this stage were the application of genome editing in animals and the absence of primary data. 941 articles passed the inclusion criteria on title/abstract level or were rated as “unclear” and remained included for full text screening. The application of the inclusion criteria at full text level resulted in 524 relevant articles. Searching for grey literature on company websites and websites from governmental authorities identified another 31 further relevant documents and web pages that were considered during data extraction. Out of these 555 records, a total amount of 1328 studies were extracted and formed the basis for this systematic map. A list of all studies and the extracted data to answer secondary question one is provided in Additional file 5. Additional file 6 provides a list of all studies and the extracted data to answer secondary question two. A ROSES reporting form is included in Additional file 7.

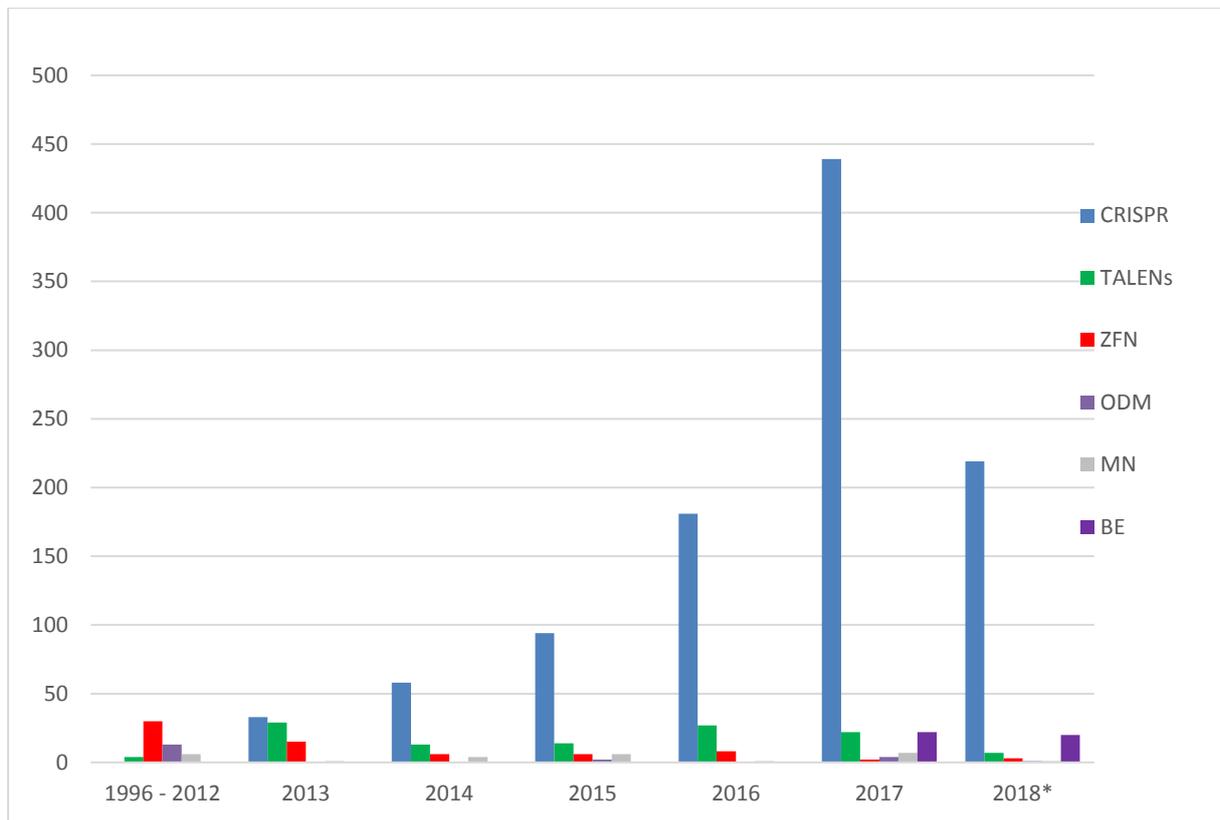


**Figure 5.1** Flow diagram of the systematic mapping process explaining the selection of relevant articles and studies. This diagram follows ROSES guidance [40].

### 5.5.1 General overview of the application of genome editing in model plants and crops

#### Studies per year

In the mapping period between January 1996 and May 2018, a total amount of 1328 studies were identified. As shown in Figure 5.2, the number of studies using TALENs, ZFN, ODM, MN and BE has remained on a relatively low level. In contrast, the number of studies on CRISPR/Cas has risen sharply soon after the system was applied for the first time in plants in 2013. Nearly 85% of the studies were published since 2015 indicating the rapid dissemination and development of these techniques within the last few years.



**Figure 5.210** The number of genome editing applications published per year.

\*Only January – May 2018; CRISPR/Cas: Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated protein; TALENs: Transcription activator-like effector nucleases; ZFN: Zinc-Finger Nucleases; ODM: Oligo-Directed Mutagenesis; MN: Meganucleases; BE: Base editing.

As shown in Table 5.1, 26 studies used MN, 27 studies used ODM, 42 studies used BE, 73 studies used ZFN, 128 studies used TALENs and 1032 studies used CRISPR/Cas. When using CRISPR/Cas, the most frequently used nuclease was Cas9 (n=986) followed with a large distance by Cas12a (also known as Cpf1) (n=46).

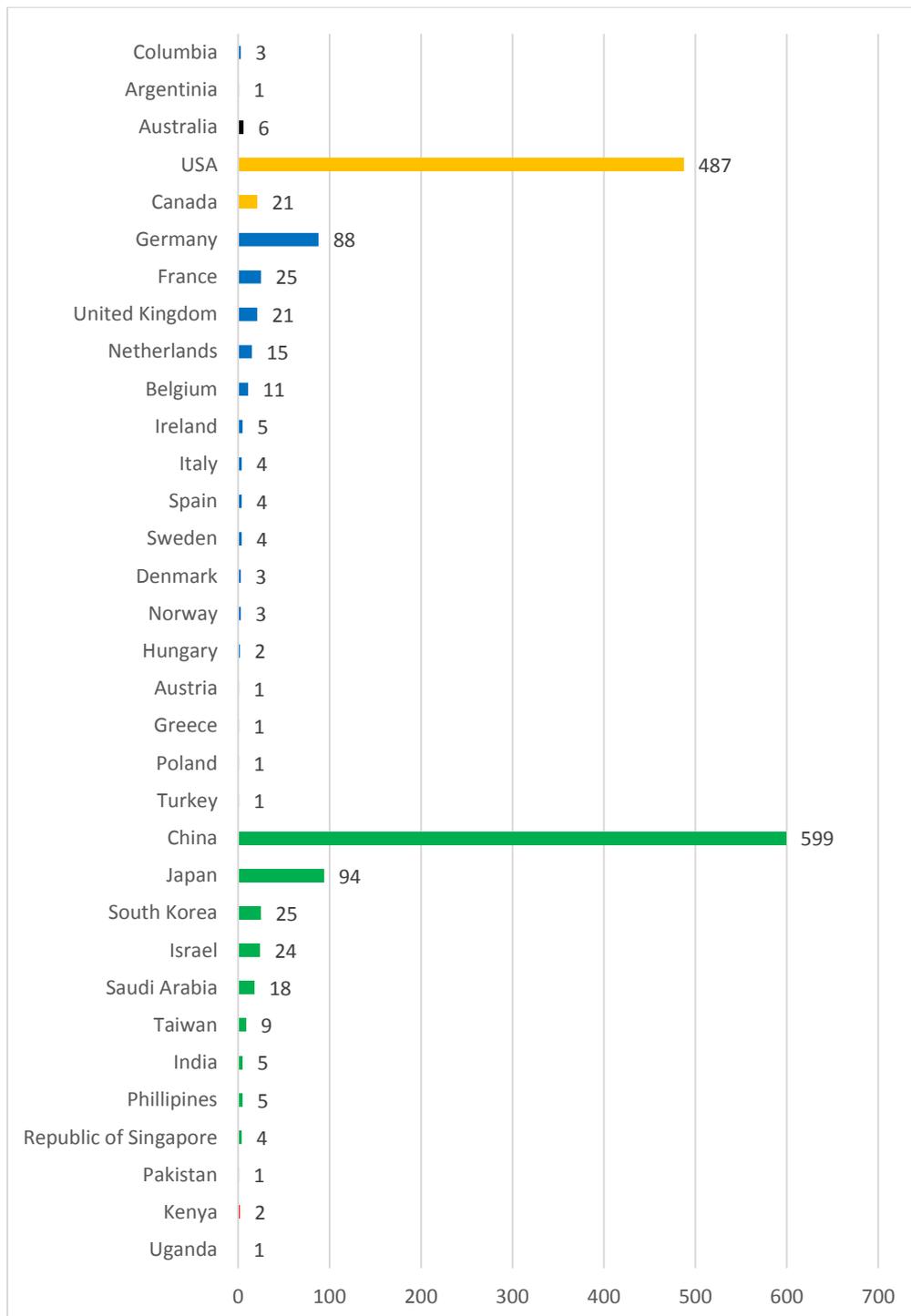
**Table 5.1** Heat map showing number of studies performed with the different genome editing techniques (rows) and the year the studies were published (columns) (1996 - May 2018).

|                | CRISPR/Cas | TALENs | ZFN | ODM | MN | BE | # studies |
|----------------|------------|--------|-----|-----|----|----|-----------|
| 1996           | 0          | 0      | 0   | 0   | 0  | 0  | 0         |
| 1997           | 0          | 0      | 0   | 0   | 0  | 0  | 0         |
| 1998           | 0          | 0      | 0   | 0   | 0  | 0  | 0         |
| 1999           | 0          | 0      | 0   | 4   | 0  | 0  | 4         |
| 2000           | 0          | 0      | 0   | 1   | 0  | 0  | 1         |
| 2001           | 0          | 0      | 0   | 0   | 0  | 0  | 0         |
| 2002           | 0          | 0      | 0   | 0   | 0  | 0  | 0         |
| 2003           | 0          | 0      | 0   | 6   | 1  | 0  | 7         |
| 2004           | 0          | 0      | 0   | 1   | 0  | 0  | 1         |
| 2005           | 0          | 0      | 2   | 0   | 0  | 0  | 2         |
| 2006           | 0          | 0      | 1   | 1   | 0  | 0  | 2         |
| 2007           | 0          | 0      | 0   | 0   | 0  | 0  | 0         |
| 2008           | 0          | 0      | 4   | 0   | 0  | 0  | 4         |
| 2009           | 0          | 0      | 10  | 0   | 2  | 0  | 12        |
| 2010           | 0          | 0      | 5   | 0   | 1  | 0  | 6         |
| 2011           | 0          | 2      | 10  | 0   | 0  | 0  | 12        |
| 2012           | 0          | 2      | 1   | 0   | 2  | 0  | 5         |
| 2013           | 33         | 29     | 15  | 0   | 1  | 0  | 78        |
| 2014           | 58         | 13     | 6   | 0   | 4  | 0  | 81        |
| 2015           | 94         | 14     | 6   | 2   | 6  | 0  | 122       |
| 2016           | 181        | 27     | 8   | 0   | 1  | 0  | 217       |
| 2017           | 439        | 22     | 2   | 4   | 7  | 22 | 496       |
| 2018*          | 219        | 7      | 3   | 1   | 1  | 20 | 251       |
| No information | 8          | 12     | 0   | 7   | 0  | 0  | 27        |
| # studies      | 1032       | 128    | 73  | 27  | 26 | 42 | 1328      |

\*only January – May 2018; CRISPR/Cas: Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated protein; TALENs: Transcription activator-like effector nucleases; ZFN: Zinc-Finger Nucleases; ODM: Oligo-Directed Mutagenesis; MN: Meganucleases; BE: Base editing.

### Geographical distribution of genome editing studies

The number of studies per country was calculated based on the country the corresponding author is located at. In case of grey literature, the study was accounted based on the country the company is located at. Multiple assignments are possible if the corresponding author is affiliated with institutions in different countries or if more than one author was indicated being corresponding author. Therefore, the sum of studies accounted for different countries is higher (n=1494) than the total amount of studies identified (n=1328). Asia is the leading continent when applying and publicizing genome editing in plants (n=784 studies, 53%) followed by North America (n=508 studies, 34%), Europe (n=189 studies, 13%), Australia (n=6 studies, <1%), South America (n=4 studies, <1%) and Africa (n=3 studies, <1%) (Figure 5.3). In total, publications from 33 countries were identified. As shown in Figure 5.3, China has a substantial lead in the number of studies (n=599 studies, 40%) followed by the USA (n=487 studies, 33%), Japan (94 studies, 6%) and Germany (n=88 studies, 6%).



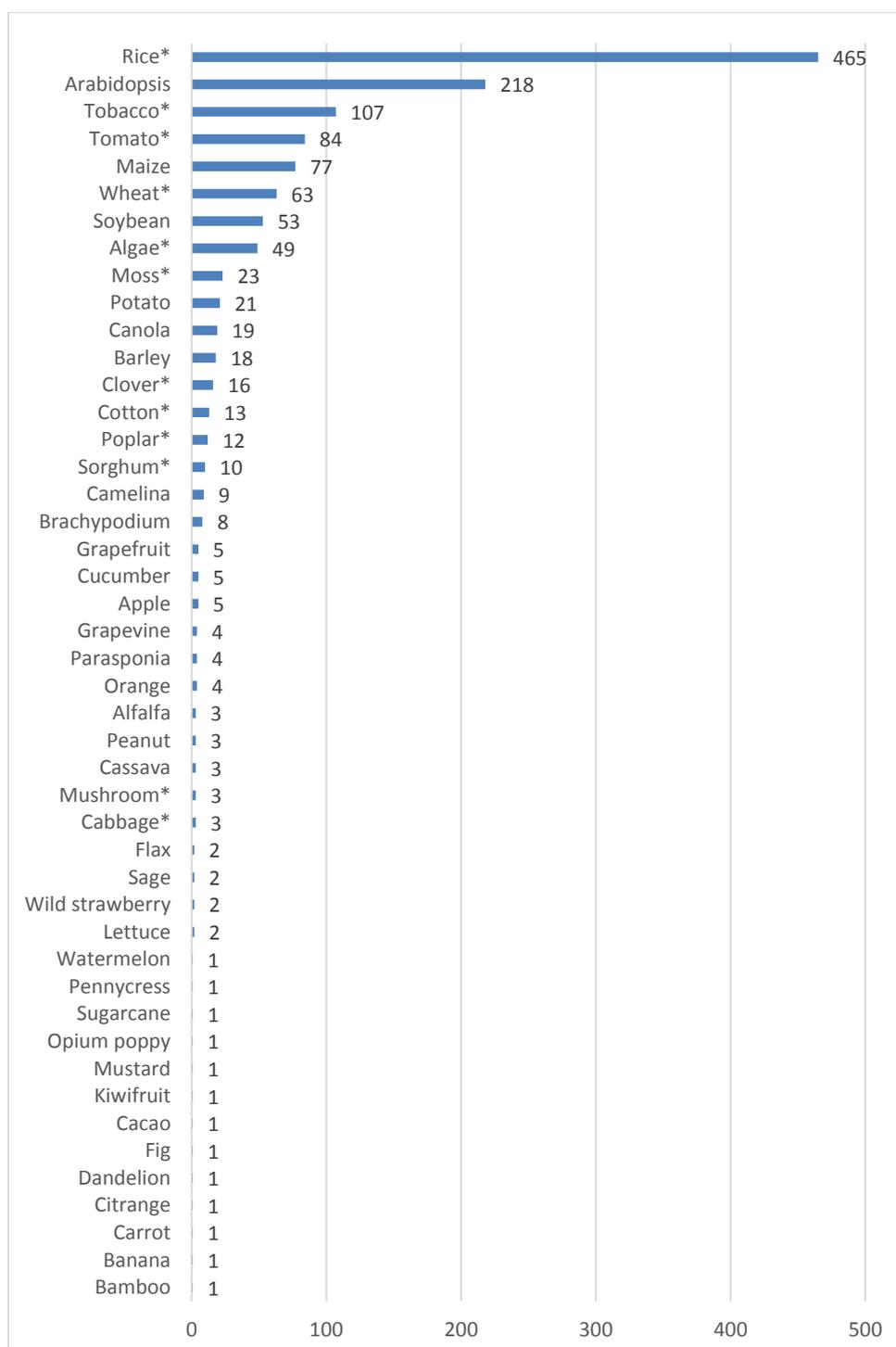
**Figure 5.3** Number of studies per country\* in the systematic map database (Grouped by continent; January 1996 - May 2018).

\* Identified by the corresponding author(s); South America: Grey; Australia: Black; North America: Orange; Europe: Blue; Asia: Green; Africa: Red

### Genome editing applications in model plants and crops

Around two third of the studies ( $n=907$ ; 68%) were conducted on agricultural crops and one third ( $n=421$ ; 32%) on model organisms. However, it is worth noting that rice is an important crop plant but it is also used as model species because of its genome size and because embryogenic rice cultures can be easily prepared, transformed and rapidly regenerated into fertile plants. In total, 51 species and subspecies of different agricultural crops and 17 species/subspecies of model plants

were under investigation, with the majority of studies focusing on rice (n=465) followed by the model organisms *Arabidopsis* (n=218) and tobacco (n=107). Besides to these, tomato (n=84) is most commonly studied followed by maize (n=77), wheat (n=63) and soybean (n=53) (Figure 5.4).

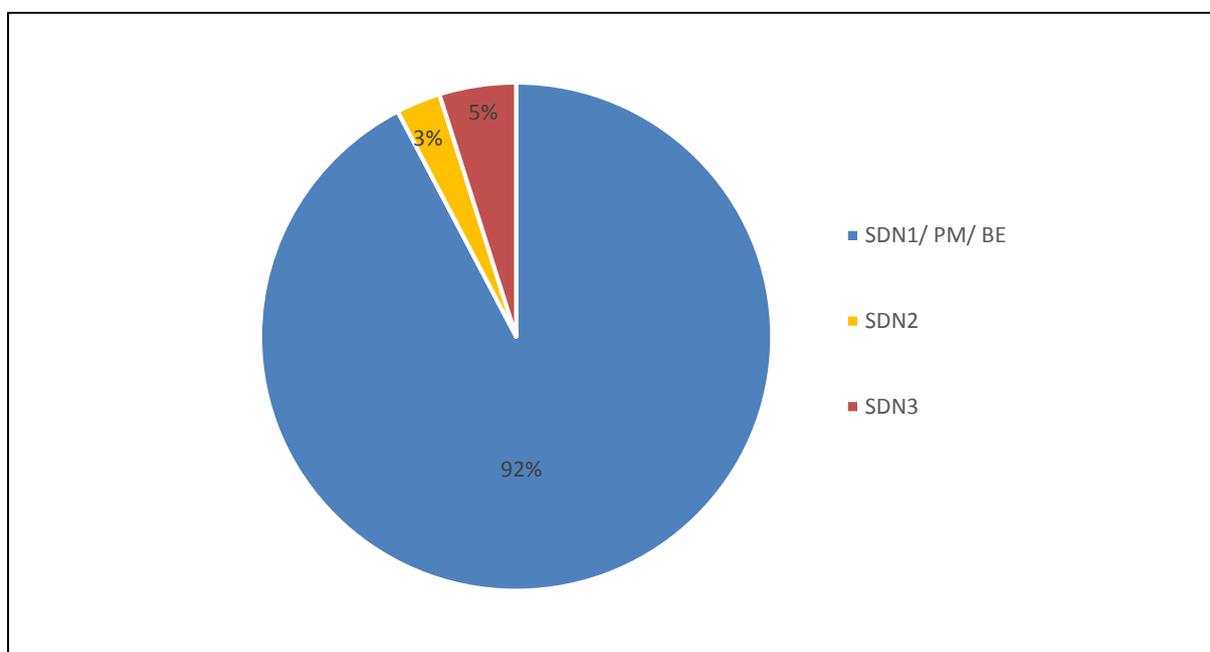


**Figure 5.4** Total amount of genome editing applications in crops and model plants (1996 - May 2018) \*Several species or subspecies were used when applying genome editing.

### Type of alteration in the plant genome

As shown in Figure 5.5, the majority of studies (n=1223; 92%) describe the induction of point mutations or indels comparable to spontaneous mutations or undirected mutagenesis. This was mainly achieved with SDN1 (n=1154) where no repair template was added and the DSB was

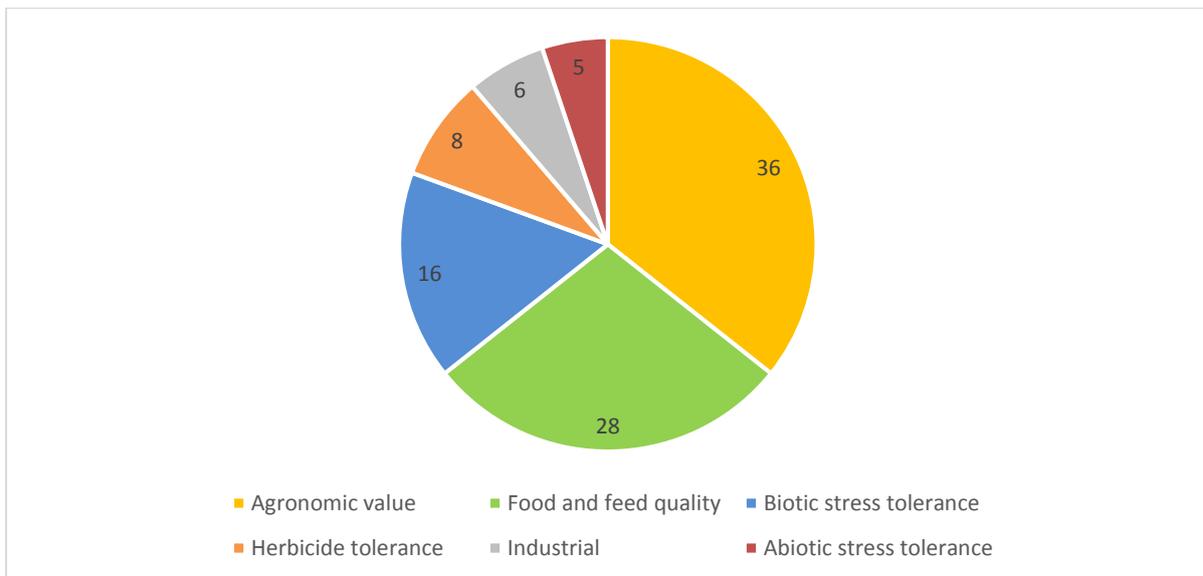
repaired by NHEJ. Additionally, the induction of point mutations (PM) using the ODM technique (n=27) or BE (n=42) leads to point mutations comparable to SDN1. Only 36 studies (3%) added a repair template that, except for a few nucleotides, was identical to the targeted sequence in which the DSB was introduced leading to a DSB repair via homology-directed repair (SDN2). In 68 studies (5%), a repair template was added that harbors a recombinant DNA sequence additional to the homologous sequences and the DSB was repaired via homology-directed repair (SDN3).



**Figure 5.5** Distribution of the type of alteration introduced by the application of genome editing (January 1996 - May 2018). SDN: Site-directed nucleases; PM: Point mutation; BE: Base editing

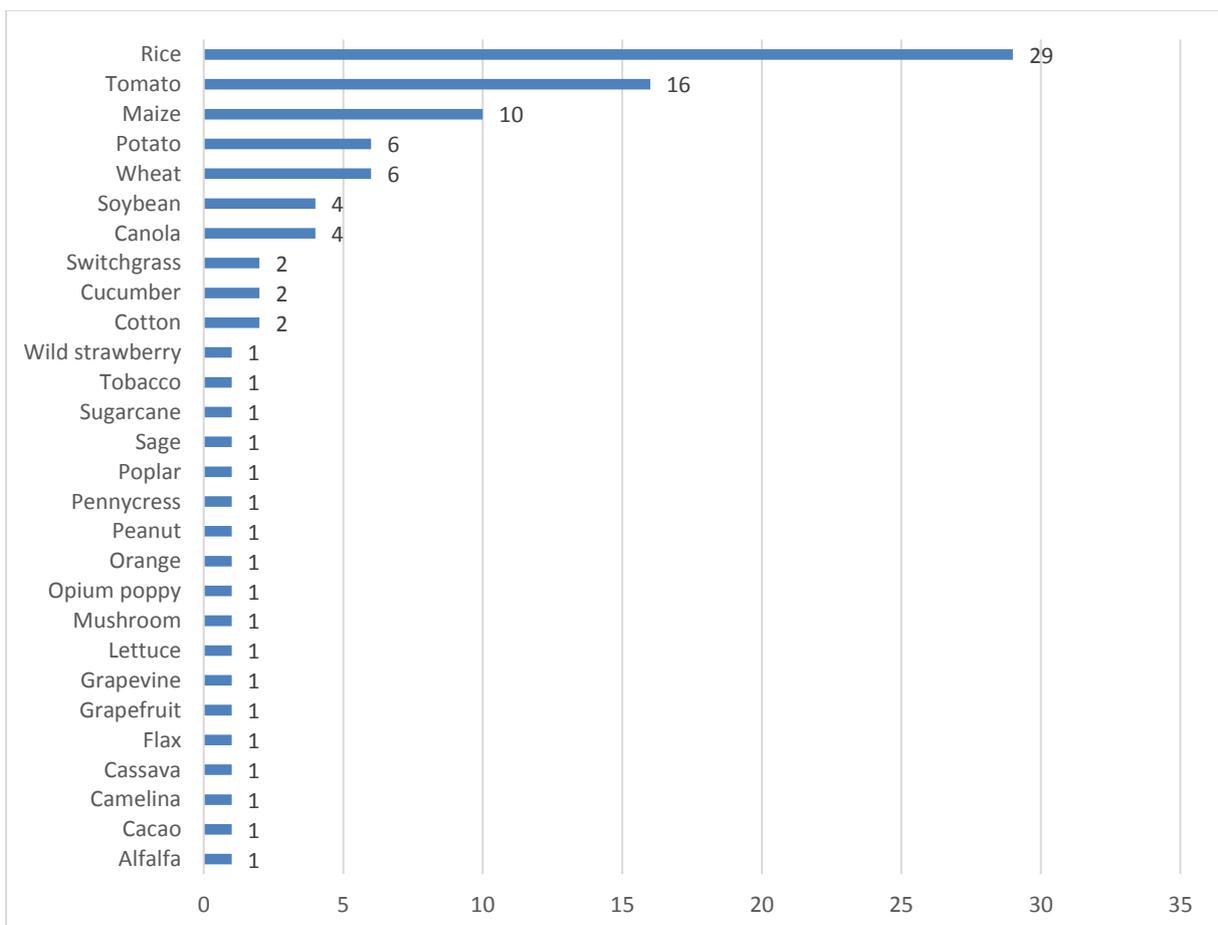
### 5.5.2 Secondary question 1: “What are the traits modified by genome editing in model plants as well as in crops produced for agricultural production?”

In total, 193 studies were allocated as market-oriented applications. However, different scientists studied the same crop species and trait. Considering this, a total amount of 99 different applications in 28 different plant species remained. These market-oriented applications build the basis to answer secondary question 1. Figure 5.6 categorizes the market-oriented applications to different groups of traits. Most of the market-oriented applications (n=36) are related to an improved agronomic value (Table 5.2), followed by 28 applications with an improved food and feed quality (Table 5.3). For biotic stress tolerance, 16 different applications were identified (Table 5.4), for herbicide tolerance eight applications (Table 5.5), for industrial utilization six applications (Table 5.6) and for abiotic stress tolerance five applications (Table 5.7). For more detailed information about the analyzed traits, see the respective tables.



**Figure 5.6** Distribution of market-oriented applications of crops with nutritionally, agriculturally or industrially relevant traits (January 1996 - May 2018).

As shown in Figure 5.7, most of the market-oriented applications were applied in rice ( $n=29$ ), followed by tomato ( $n=16$ ), maize ( $n=10$ ), potato ( $n=6$ ), wheat ( $n=6$ ), soybean ( $n=4$ ) and canola ( $n=4$ ). In 21 other agricultural relevant crops, one or two market-oriented applications were identified.



**Figure 5.7** Total amount of genome editing applications in crops with market-oriented traits (January 1996 - May 2018).

**Table 5.2** Total amount of genome editing applications in crops with market-oriented traits (1996 - May 2018).

| Plant    | Developer, Producer, Country   | Trait                   | Specification  | Technological specification | Reference |
|----------|--|-------------------------|--|-----------------------------|-----------|
| Canola   | Christian-Albrechts-University of Kiel, Germany  | Increased yield         | Increased shatter resistance to avoid seed loss during mechanical harvest                  | CRISPR/Cas9<br>SDN1         | [41]      |
| Canola   | Huazhong Agricultural University, China  | Increased yield         | Increased seeds number per husk, higher seed weight  | CRISPR/Cas9<br>SDN1         | [42]      |
| Cotton   | Anhui Agricultural University, China; Chinese Academy of Agricultural Sciences, China          | Growth characteristics  | Improved root growth under high and low nitrogen conditions                                | CRISPR/Cas9<br>SDN1         | [43]      |
| Cucumber | Chinese Academy of Agricultural Sciences, China  | Growth characteristics  | Only female flowers  | CRISPR/Cas9<br>SDN1         | [44]      |
| Lettuce  | University of California, USA  | Increased yield         | Germination at high temperature  | CRISPR/Cas9<br>SDN1         | [45]      |
| Maize    | Benson Hill Biosystems, USA  | Increased yield         | Increased photosynthesis efficiency  | Meganuclease<br>SDN3        | [46]      |
| Maize    | University of Wisconsin, USA   | Growth characteristics  | Early flowering under long day conditions  | CRISPR/Cas9<br>SDN1         | [47]      |
| Maize    | DuPont Pioneer, USA;   | Growth characteristics  | Male sterility   | CRISPR/Cas9<br>SDN1         | [48, 49]  |
|          | University of Science and Technology Beijing, China; Beijing Solidwill Sci-Tech Co. Ltd, China |                         |  | CRISPR/Cas9<br>SDN1         | [4]       |
|          | Chinese Academy of Sciences, China   |                         |  | CRISPR/Cas9<br>SDN1         | [50]      |
| Maize    | Syngenta Seeds, USA  | Growth characteristics  | Haploid induction  | TALENs<br>SDN1              | [51]      |
| Potato   | Collectis Plant Science, USA   | Storage characteristics | Improved cold storage and processing traits (reduced sugars/ reduced levels of acrylamide) | TALENs<br>SDN1              | [52]      |
| Rice     | Chinese Academy of Sciences, China   | Increased yield         | Altered grain number per panicle   | CRISPR/Cas9<br>SDN1         | [53]      |
|          | National Rice Research Institute, China  |                         |  | CRISPR/Cas9<br>SDN1         | [54]      |
| Rice     | Chinese Academy of Sciences, China   | Increased yield         | Seed size/ increased seed weight   | CRISPR/Cas9<br>SDN1         | [53]      |
|          | Anhui Academy of Agricultural Sciences, China  |                         |  | CRISPR/Cas9<br>SDN1         | [55]      |

| Plant | Developer, Producer, Country  | Trait                   | Specification  | Technological specification | Reference |
|-------|---|-------------------------|--|-----------------------------|-----------|
|       | Fudan University, China   |                         |  | CRISPR/Cas9<br>SDN1         | [56]      |
|       | Yangzhou University, China  |                         |  | CRISPR/Cas9<br>SDN1         | [57]      |
|       | Agronomy College of Henan Agricultural University, China  |                         |  | CRISPR/Cas9<br>SDN1         | [58]      |
|       | Chinese Academy of Agricultural Sciences, China; Yangzhou University, China                     |                         |  | CRISPR/Cas9<br>SDN1         | [59]      |
| Rice  | Chinese Academy of Sciences, China  | Growth characteristics  | Increased plant height, improved tiller-production, erect panicle, increased biomass | CRISPR/Cas9;<br>SDN1        | [53, 60]  |
|       | Wuhan Institute of Bioengineering; Huazhong Agricultural University, China                      |                         |  | CRISPR/Cas9<br>SDN1         | [60]      |
|       | Sichuan Agricultural University   |                         |  | CRISPR/Cas9<br>SDN1         | [61]      |
|       | Chinese Academy of Agricultural Sciences, China; Yangzhou University, China                     |                         |  | CRISPR/Cas9<br>SDN1         | [59]      |
| Rice  | Chinese Academy of Agricultural Sciences, China; Jangsu Academy of Agricultural Sciences, China | Growth characteristics  | Early maturing   | CRISPR/Cas9<br>SDN1         | [62]      |
| Rice  | Kyung Hee University, South Korea   | Growth characteristics  | Male sterility   | CRISPR/Cas9<br>SDN1         | [63]      |
|       | Shanghai Jiao Tong University, China  |                         |  | CRISPR/Cas9<br>SDN1         | [64]      |
|       | South China Agricultural University, China  |                         |  | CRISPR/Cas9<br>SDN1         | [65, 66]  |
|       | Sichuan Agricultural University, China  |                         |  | CRISPR/Cas9<br>SDN1         | [67, 68]  |
| Rice  | Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China            | Increased yield         | Regulation of pollen tube growth   | CRISPR/Cas9<br>SDN1         | [69]      |
| Rice  | China Agricultural University, China  | Storage characteristics | Increased seed storage   | TALENs<br>SDN1              | [70]      |

| Plant       | Developer, Producer, Country  | Trait                  | Specification  | Technological specification | Reference |
|-------------|---|------------------------|--|-----------------------------|-----------|
| Rice        | China National Rice Research Institute, China; China Three Gorges University, China   | Increased yield        | Increased seed setting rate  | CRISPR/Cas9 SDN1            | [71]      |
| Rice        | Anhui Academy of Agricultural Sciences, China   | Increased yield        | Longer panicle   | CRISPR/Cas9 SDN1            | [55]      |
| Rice        | Nanjing Agricultural University, China  | Increased yield        | Grain yield, Regulation of seed development  | CRISPR/Cas9 SDN1            | [72]      |
| Rice        | Chinese Academy of Sciences, China  | Growth characteristics | Decreased plant height   | CRISPR/Cas9 SDN1            | [73]      |
|             | Syngenta Biotechnology, China   |                        |  | CRISPR/Cas9 SDN1            | [74]      |
| Rice        | Wuhan Institute of Bioengineering, China; Huazhong Agricultural University, China   | Increased yield        | Increased nitrogen utilization efficiency  | CRISPR/Cas9 SDN1            | [60]      |
| Rice        | Hunan Normal University, China  | Growth characteristics | Regulation of seed dormancy, stomatal opening, plant developmental, abiotic stress tolerance and leaf senescence | CRISPR/Cas9 SDN1            | [75]      |
| Soybean     | Chinese Academy of Agricultural Sciences, China   | Growth characteristics | Late flowering   | CRISPR/Cas9 SDN1            | [76]      |
| Switchgrass | Iowa State University, USA  | Growth characteristics | Bushy phenotype  | CRISPR/Cas9 SDN1            | [77]      |
| Tomato      | National Food Research Institute, Japan   | Increased yield        | Regulating fruit ripening  | CRISPR/Cas9 SDN1            | [78]      |
| Tomato      | University of Minnesota, USA  | Growth characteristics | Bigger seedlings   | TALENs SDN1                 | [79]      |
| Tomato      | Cold Spring Harbor Laboratory, USA; Max Planck Institute for Plant Breeding Research, Germany; Université Paris-Sclay, France | Growth characteristics | Early flowering  | CRISPR/Cas9 SDN1            | [80]      |
| Tomato      | University of Florida, USA  | Growth characteristics | Easy separation of fruit and stem  | CRISPR/Cas9 SDN1            | [81]      |
| Tomato      | Cold Spring Harbor Laboratory, USA  | Increased yield        | Fruit size   | CRISPR/Cas9 SDN1            | [82]      |
| Tomato      | Cold Spring Harbor Laboratory, USA  | Increased yield        | Highly branched inflorescence and formation of multiple flowers  | CRISPR/Cas9 SDN1            | [82]      |

| Plant           | Developer, Producer, Country   | Trait                  | Specification                         | Technological specification | Reference |
|-----------------|--|------------------------|---------------------------------------|-----------------------------|-----------|
| Tomato          | Weizmann Institute of Science, Israel  | Growth characteristics | Yellow fruit color                    | CRISPR/Cas9 SDN1            | [83]      |
|                 | Weizmann Institute of Science, Israel  |                        |                                       | CRISPR/Cas9 SDN 3           | [84]      |
| Tomato          | Weizmann Institute of Science, Israel  | Growth characteristics | Orange fruit color                    | CRISPR/Cas9 SDN3            | [84]      |
| Tomato          | Academy of Agriculture and Forestry Sciences; Chinese Academy of Sciences, China | Growth characteristics | Pink fruit color                      | CRISPR/Cas9 SDN1            | [85]      |
| Wheat           | Kansas State University, USA   | Increased yield        | Bigger grains, increased grain weight | CRISPR/Cas9 SDN1            | [86]      |
|                 | Chinese Academy of Sciences, China   |                        |                                       | CRISPR/Cas9 SDN1            | [87]      |
| Wild strawberry | University of Maryland, USA  | Growth characteristics | Faster seedling growth                | CRISPR/Cas9 SDN1            | [88]      |

Explanations:

TALENS: Transcription Activator-Like Effector Nucleases

CRISPR/Cas9: Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated protein 9

SDN: Site-Directed Nucleases

**Table 5.3** Genome editing in plants for improved food and feed quality (1996 - May 2018).

| Plant           | Developer, Producer, Country                    | Trait           | Specification   | Technological specification | Reference |
|-----------------|---|-----------------|---|-----------------------------|-----------|
| Alfalfa         | Calyxt, Inc., USA                               | Product quality | Reduced lignin content                                      | TALENS SDN1                 | [89]      |
| Camelina sativa | Montana State University, USA                   | Product quality | Increased levels of oleic acid and $\alpha$ -linolenic acid | CRISPR/Cas9 SDN1            | [90]      |
|                 | University Nebraska, USA                        |                 |   | CRISPR/Cas9 SDN1            | [91]      |
|                 | Université Paris-Saclay, France                 |                 |   | CRISPR/Cas9 SDN1            | [92]      |
|                 | Kansas State University, USA                    |                 |   | CRISPR/Cas9 SDN1            | [93]      |
| Canola          | Tamagawa University, Japan                      | Product quality | Altered fatty acid composition                              | CRISPR/Cas9 SDN1            | [94]      |
| Maize           | Du Pont Pioneer, USA;                           | Product quality | Waxy corn, improved starch production                       | CRISPR/Cas9 SDN1            | [95]      |
|                 | Chinese Academy of Agricultural Sciences, China |                 |   | CRISPR/Cas9 SDN1            | [96]      |
| Maize           | Agrivida, USA                                   | Product quality | Higher levels of starch in their leaves and stalks          | Meganucleases SDN1          | [97]      |

| Plant       | Developer, Producer, Country   | Trait           | Specification   | Technological specification | Reference |
|-------------|--|-----------------|---|-----------------------------|-----------|
| Maize       | Dow AgroScience, USA   | Product quality | Reduced phytate production<br>+ herbicide tolerance                     | ZFN<br>SDN3                 | [98, 99]  |
| Mushroom    | Penn State University, USA   | Product quality | Non-browning mushroom   | CRISPR/Cas9<br>SDN1         | [100]     |
| Opium poppy | Cankiri Karatekin University, Turkey; Dokuz Eylul University, Turkey                       | Product quality | Reduced morphine and thebaine content                                   | CRISPR/Cas9<br>SDN1         | [101]     |
| Peanut      | Guangdong Academy of Agricultural Sciences, China  | Product quality | Increased oleic acid content, decreased linoleic acid content           | TALENs<br>SDN1              | [102]     |
| Potato      | Calyxt, USA  | Product quality | Non-browning potato   | TALENs<br>SDN1              | [103]     |
| Potato      | Simplot Plant Science, USA   | Product quality | Reduced black spottiness  | TALENs<br>SDN1              | [104]     |
| Potato      | RIKEN Center for Sustainable Resource Science, Japan; Chiba University, Japan              | Product quality | Reduction of harmful ingredients (Glycoalkaloids)                       | TALENs<br>SDN1              | [105]     |
|             | Kobe University, Japan;  |                 | Complete abolition of Glycoalkaloids (bitter taste)                     | CRISPR/Cas9<br>SDN1         | [106]     |
| Rice        | Chinese Academy of Sciences, China   | Product quality | Fragrant rice   | TALENs<br>SDN1              | [107]     |
|             | Chinese Academy of Agricultural Sciences, China; Yangzhou University, China                |                 |   | CRISPR/Cas9<br>SDN1         | [59]      |
| Rice        | Chinese Academy of Agricultural Sciences, China; University of California, USA             | Product quality | Increased contents benefitting human health (increased amylose content) | CRISPR/Cas9<br>SDN1         | [108]     |
| Rice        | Huazhong Agricultural University, China  | Product quality | Reduced contents harming human health (arsenic content)                 | CRISPR/Cas9<br>SDN1         | [109]     |
|             | Sun Yat-sen University, China  |                 |   | CRISPR/Cas9<br>SDN1         | [110]     |
| Rice        | National Agriculture and Food Research Organization, Japan                                 | Product quality | Altered fatty acid composition  | CRISPR/Cas9<br>SDN1         | [111]     |
| Rice        | Université Montpellier, France   | Product quality | Reduced contents harming human health (cesium content)                  | CRISPR/Cas9<br>SDN1         | [112]     |
| Rice        | Hunan Agricultural University, Hunan Hybrid Rice Research Center, Normal University, China | Product quality | Reduced contents harming human health (Cadmium content in plants)       | CRISPR/Cas9<br>SDN1         | [113]     |

| Plant         | Developer, Producer, Country   | Trait           | Specification  | Technological specification                | Reference      |
|---------------|--|-----------------|--|--|----------------|
| Rice          | Chinese Academy of Sciences, Shanghai, China; Purdue University, West Lafayette, USA | Product quality | Waxy rice  | CRISPR/Cas9<br>SDN1                        | [114]          |
| Sage          | Second Military Medical University, China  | Product quality | Reduced phenolic acid content  | CRISPR/Cas9<br>SDN1                        | [115]          |
| Soybean       | Cellectis plant science Inc., USA/Calyxt, USA  | Product quality | High oleic content, low linoleic content                                 | TALENs<br>SDN1                             | [21, 116–118]  |
| Tomato        | Agricultural Research Organization, Israel; Tokushima University, Japan              | Product quality | Seedless tomato  | CRISPR/Cas9<br>SDN1<br>CRISPR/Cas9<br>SDN1 | [119]<br>[120] |
| Tomato        | University of Tsukuba, Japan;  | Product quality | Increased contents benefitting human health (increased GABA content)     | CRISPR/Cas9<br>SDN1                        | [121]          |
|               | China Agricultural University, China   |                 |  | CRISPR/Cas9<br>SDN1                        | [122]          |
| Tomato        | China Agricultural University, China   | Product quality | Increased contents benefitting human health (increased lycopene content) | CRISPR/Cas9<br>SDN1                        | [123]          |
| Tomato        | Xinjiang Academy of Agricultural Science, China                                      | Product quality | Improved shelf life  | CRISPR/Cas9<br>SDN1                        | [124]          |
| Wheat         | Calyxt, Inc., USA  | Product quality | Increased nutritional value  | TALENs<br>SDN1                             | [125]          |
| Wheat         | Instituto de Agricultura Sostenible (IAS-CSIC), Spain; University of Minnesota, USA  | Product quality | Reduced gluten content   | CRISPR/Cas9<br>SDN1                        | [126]          |
| Wheat (durum) | Instituto de Agricultura Sostenible (IAS-CSIC), Spain; University of Minnesota, USA  | Product quality | Reduced gluten content   | CRISPR/Cas9<br>SDN1                        | [126]          |

Explanations:

TALENs: Transcription Activator-Like Effector Nucleases

CRISPR/Cas9: Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated protein 9

ZFN: Zinc-Finger Nucleases

SDN: Site-Directed Nucleases

**Table 5.4** Genome editing in plants for increased resistance to biotic stress (1996 - May 2018).

| Plant | Developer, Producer, Country       | Trait             | Specification                         | Technological specification | Reference |
|-------|------------------------------------|-------------------|---------------------------------------|-----------------------------|-----------|
| Cacao | Pennsylvania State University, USA | Fungal resistance | Resistance to Phytophthora tropicalis | CRISPR/Cas9<br>SDN1         | [127]     |

| Plant      | Developer, Producer, Country   | Trait                | Specification  | Technological specification   | Reference  |
|------------|--|----------------------|--|-------------------------------|------------|
| Cucumber   | Volcani Center, Israel   | Virus resistance     | Immunity to Cucumber vein yellowing virus (Ipomovirus) infection and resistance to the potyviruses Zucchini yellow mosaic virus and Papaya ring spot mosaic virus-W. | CRISPR/Cas9 SDN1              | [22]       |
| Grapefruit | University of Florida, USA   | Bacterial resistance | Resistance to citrus canker  | CRISPR/Cas9 SDN1              | [128, 129] |
| Grapevine  | Northwest A&F University and Ministry of Agriculture, China  | Fungal resistance    | Resistance to Botrytis cinerea   | CRISPR/Cas9 SDN1              | [130]      |
| Maize      | Du Pont Pioneer, USA   | Fungal resistance    | Resistance to Northern Leaf Blight (NLB)   | CRISPR/Cas9 (Cisgenesis) SDN3 | [131]      |
| Orange     | Chinese Academy of Agricultural Sciences and National Center for Citrus Variety Improvement; Southwest University, China | Bacterial resistance | Resistance to citrus canker  | CRISPR/Cas9 SDN1              | [132]      |
| Rice       | Chinese Academy of Agriculture, China  | Fungal resistance    | Resistance to rice blast   | CRISPR/Cas9; SDN1             | [133]      |
| Rice       | Iowa State University, USA   | Bacterial resistance | Resistance to bacterial blight   | CRISPR/Cas9; SDN1             | [134]      |
|            | IRD-CIRAD- Université, France  |                      |  | TALENs SDN1                   | [135]      |
|            | Iowa State University, USA   |                      |  | TALENs SDN1                   | [136]      |
|            | National University of Singapore, Singapore  |                      |  | TALENs SDN1                   | [137]      |
|            | Chinese Academy of Sciences, China   |                      |  | TALENs SDN1                   | [138]      |
| Rice       | National Center for Plant Gene Research, China; Sichuan Agricultural University, China                                   | Fungal resistance    | Resistance to powdery mildew   | CRISPR/Cas9; SDN1             | [139]      |
|            | Iowa State University, USA   |                      |  | TALENs SDN1                   | [140]      |
| Rice       | Shanghai Jiao Tong University, China; Yunnan Academy of Agricultural Sciences, China                                     | Bacterial resistance | Resistance to the pathogen Xoc RS105   | TALENs SDN1                   | [141]      |

| Plant  | Developer, Producer, Country   | Trait                                      | Specification  | Technological specification | Reference |
|--------|--|--|--|-----------------------------|-----------|
| Rice   | Sichuan Agricultural University, China   | Bacterial resistance/<br>Fungal resistance | Resistance to bacterial blight and rice blight   | CRISPR/Cas9<br>SDN1         | [61]      |
| Rice   | International Rice Research Institute (IRRI), Philippines                          | Virus resistance                           | Resistance to rice tungro disease (RTD)  | CRISPR/Cas9<br>SDN1         | [142]     |
| Tomato | Max Planck Institute for Developmental Biology, Germany; Norwich Research Park, UK | Fungal resistance                          | Resistance to powdery mildew   | CRISPR/Cas9<br>SDN1         | [143]     |
| Tomato | King Abdullah University of Science and Technology, Saudi Arabia                   | Virus resistance                           | Resistance to tomato yellow leaf virus   | CRISPR/Cas9<br>SDN1         | [144]     |
| Tomato | University of California, USA  | Bacterial resistance                       | Resistance to different pathogens including <i>P. syringae</i> , <i>P. capsici</i> and <i>Xanthomonas</i> spp. | CRISPR/Cas9<br>SDN1         | [145]     |
| Wheat  | Chinese Academy of Sciences, China   | Fungal resistance                          | Resistance to powdery mildew   | TALENs<br>SDN1              | [25]      |
|        | Chinese Academy of Sciences, China   |  |  | CRISPR/Cas9<br>SDN1         | [146]     |
|        | Calyxt, Inc., USA  |  |  | TALENs<br>SDN1              | [147]     |

Explanations:

TALENs: Transcription Activator-Like Effector Nucleases

CRISPR/Cas9: Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated protein 9

SDN: Site-Directed Nucleases

**Table 5.5** Genome editing for generating herbicide tolerant plants (1996 - May 2018).

| Plants  | Developer, Producer, Country                         | Trait*              | Specification | Technological specification     | Reference  |
|---------|--|---------------------|---------------|---------------------------------|------------|
| Canola  | Cibus, Canada; Cibus, USA;                           | Herbicide tolerance | -             | ODM                             | [23]       |
|         | Bayer BioScience N.V., Belgium                       |                     |               |                                 | [148]      |
| Cassava | Donald Danforth Plant Science Center, St. Louis, USA | Herbicide tolerance | -             | CRISPR/Cas9<br>SDN3             | [149]      |
| Cotton  | Bayer CropScience N.V., Belgium                      | Herbicide tolerance | -             | Mega-nucleases<br>SDN3          | [150]      |
| Flax    | Cibus, USA   | Herbicide tolerance | -             | CRISPR/Cas9<br>SDN1             | [151]      |
| Maize   | DuPont Pioneer, USA                                  | Herbicide tolerance | -             | CRISPR/Cas9<br>SDN1, SDN2, SDN3 | [48, 49]   |
|         | Dow AgroScience, USA                                 |                     |               | ZFN<br>SDN3                     | [99, 152]  |
|         | Pioneer Hi-Bred International, USA                   |                     |               | ODM                             | [153, 154] |

| Plants  | Developer, Producer, Country  | Trait*              | Specification | Technological specification | Reference  |
|---------|---|---------------------|---------------|-----------------------------|------------|
| Potato  | Michigan State University, USA  | Herbicide tolerance | -             | CRISPR/Cas9,TALENs SDN2     | [155]      |
|         | Chinese Academy of Sciences, China  |                     |               | CRISPR/Cas9 SDN2            | [156]      |
|         | Chinese Academy of Sciences, China; Huazhong Agricultural University, China; University of California San Diego, USA; |                     |               | CRISPR/Cas9 SDN2            | [157]      |
| Rice    | Zhejiang University, China  | Herbicide tolerance | -             | TALENs SDN2                 | [158]      |
|         | Tohoku University, Japan  |                     |               | ODM                         | [189]      |
|         | Kobe University, Japan; University of Tsukuba, Japan  |                     |               | BE                          | [160, 161] |
|         | King Abdullah University of Science and Technology, Saudi Arabia  |                     |               | CRISPR/Cas9 SDN2            | [162]      |
| Soybean | DuPont Pioneer, USA   | Herbicide tolerance | -             | CRISPR/Cas9 SDN2            | [163]      |
|         |   |                     |               | CRISPR/Cas9 SDN3            | [164]      |

Explanations:

\* No detailed breakdown regarding chemical agents

TALENs: Transcription Activator-Like Effector Nucleases

CRISPR/Cas9: Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated protein 9

ZFN: Zinc-Finger Nuclease

ODM: Oligo-Directed Mutagenesis

SDN: Site-Directed Nucleases

BE: Base editing

**Table 5.6** Genome editing in plants for industrial utilization (1996 - May 2018).

| Plant       | Developer, Producer, Country                        | Trait           | Specification                                   | Technological specification | Reference  |
|-------------|---|-----------------|---|-----------------------------|------------|
| Pennycress  | Illinois State University, USA                      | Product quality | Altered oil composition                         | CRISPR/Cas9 SDN1            | [165]      |
| Poplar      | University of Georgia, USA                          | Product quality | Stem wood discoloration due to lignin reduction | CRISPR/Cas9 SDN1            | [166]      |
| Potato      | Swedish University of Agricultural Sciences, Sweden | Product quality | Improved starch quality                         | CRISPR/Cas9 SDN1            | [167]      |
| Sugarcane   | University of Florida, USA                          | Product quality | Reduced lignin content                          | TALENs SDN1                 | [168, 169] |
| Switchgrass | Noble Research Institute, USA                       | Product quality | Reduced lignin content                          | CRISPR/Cas9                 | [170]      |
| Tobacco     | North Carolina State University, USA                | Product quality | Reduced nicotine content                        | Meganucleases SDN1          | [171]      |

Explanations:

CRISPR/Cas9: Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated protein 9

TALENs: Transcription Activator-Like Effector Nucleases

SDN: Site-Directed Nucleases

**Table 5.7** Genome editing in plants to improve tolerance to abiotic stress (1996 - May 2018).

| Plant   | Developer, Producer, Country  | Trait                      | Specification | Technological specification | Reference |
|---------|---|----------------------------|---------------|-----------------------------|-----------|
| Maize   | Ghent University, Belgium; Center for Plant Systems Biology, Belgium; Jomo Kenyatta University of Agriculture and Technology, Kenya | Drought tolerance          | -             | CRISPR/Cas9<br>SDN1         | [172]     |
|         | DuPont Pioneer, USA   |                            |               | CRISPR/Cas9<br>SDN3         | [24, 164] |
| Rice    | Anhui Academy of Agricultural Sciences, China   | Salt tolerance             | -             | CRISPR/Cas9<br>SDN1         | [20]      |
| Rice    | Huazhong Agricultural University, China   | Arsenic tolerance          | -             | CRISPR/Cas9<br>SDN1         | [109]     |
|         | Sun Yat-sen University, China   |                            |               | CRISPR/Cas9<br>SDN1         | [110]     |
| Soybean | USDA-ARS, USA   | Drought and salt tolerance | -             | CRISPR/Cas9<br>SDN1         | [173]     |
| Wheat   | Montana State University, USA   | Drought tolerance          | -             | CRISPR/Cas9<br>SDN1         | [174]     |

Explanations:

CRISPR/Cas9: Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated protein 9

SDN: Site-Directed Nucleases

### 5.5.3 Secondary question 2: “What is the available evidence for the potential occurrence of associated off-target effects due to the use of genome editing in model plants as well as in crops produced for agricultural production?”

In total, 252 studies from 161 articles were identified in which the occurrence of off-target effects was assessed. Table 5.8 maps the number of analyzed off-target effects for different genome editing techniques and different plant species. Most of the off-target analysis were conducted in CRISPR/Cas studies (n=228) followed by TALENs studies (n=9), BE studies (n=9) and ZFN studies (n=4). Solely in one ODM and in one MN study off-target effects were investigated. Most off-target effects were analyzed in rice (n=93), followed by tomato (n=28), *Arabidopsis* (n=23) and soybean (n=15) (Table 5.8).

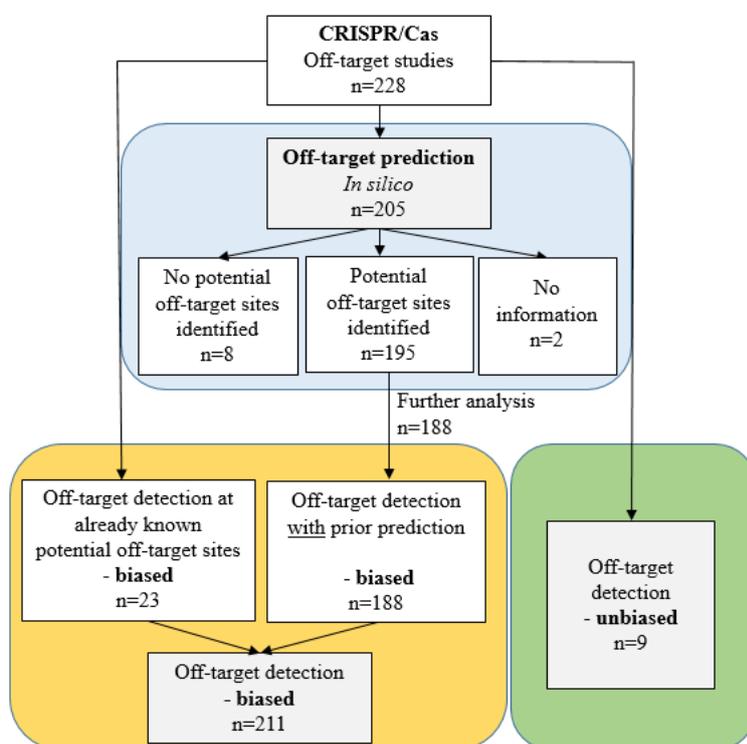
**Table 5.8** Overview of off-target studies in relation to different genome editing techniques and plant species (January 1996 - May 2018).

|                 | CRISPR/Cas | TALENs | ZFN | ODM | MN | BE | # studies |
|-----------------|------------|--------|-----|-----|----|----|-----------|
| Rice            | 82         | 2      | 0   | 1   | 0  | 8  | 93        |
| Tomato          | 25         | 1      | 1   | 0   | 0  | 1  | 28        |
| Arabidopsis     | 23         | 2      | 2   | 0   | 0  | 0  | 27        |
| Soybean         | 13         | 2      | 0   | 0   | 0  | 0  | 15        |
| Moss            | 13         | 0      | 0   | 0   | 0  | 0  | 13        |
| Wheat           | 11         | 0      | 0   | 0   | 0  | 0  | 11        |
| Tobacco         | 9          | 0      | 0   | 0   | 1  | 0  | 10        |
| Maize           | 7          | 0      | 1   | 0   | 0  | 0  | 8         |
| Algae           | 4          | 0      | 0   | 0   | 0  | 0  | 4         |
| Canola          | 4          | 0      | 0   | 0   | 0  | 0  | 4         |
| Cotton          | 4          | 0      | 0   | 0   | 0  | 0  | 4         |
| Clover          | 3          | 0      | 0   | 0   | 0  | 0  | 3         |
| Grapevine       | 3          | 0      | 0   | 0   | 0  | 0  | 3         |
| Orange          | 3          | 0      | 0   | 0   | 0  | 0  | 3         |
| Poplar          | 3          | 0      | 0   | 0   | 0  | 0  | 3         |
| Cucumber        | 2          | 0      | 0   | 0   | 0  | 0  | 2         |
| Grapefruit      | 2          | 0      | 0   | 0   | 0  | 0  | 2         |
| Lettuce         | 2          | 0      | 0   | 0   | 0  | 0  | 2         |
| Sage            | 2          | 0      | 0   | 0   | 0  | 0  | 2         |
| Barley          | 1          | 1      | 0   | 0   | 0  | 0  | 2         |
| Potato          | 1          | 1      | 0   | 0   | 0  | 0  | 2         |
| Alfalfa         | 1          | 0      | 0   | 0   | 0  | 0  | 1         |
| Cabbage         | 1          | 0      | 0   | 0   | 0  | 0  | 1         |
| Cacao           | 1          | 0      | 0   | 0   | 0  | 0  | 1         |
| Camelina        | 1          | 0      | 0   | 0   | 0  | 0  | 1         |
| Carrot          | 1          | 0      | 0   | 0   | 0  | 0  | 1         |
| Citrange        | 1          | 0      | 0   | 0   | 0  | 0  | 1         |
| Flax            | 1          | 0      | 0   | 0   | 0  | 0  | 1         |
| Kiwifruit       | 1          | 0      | 0   | 0   | 0  | 0  | 1         |
| Opium poppy     | 1          | 0      | 0   | 0   | 0  | 0  | 1         |
| Watermelon      | 1          | 0      | 0   | 0   | 0  | 0  | 1         |
| Wild strawberry | 1          | 0      | 0   | 0   | 0  | 0  | 1         |
| # studies       | 228        | 9      | 4   | 1   | 1  | 9  | 252       |

Explanation: CRISPR/Cas: Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated protein; TALENs: Transcription Activator-Like Effector Nucleases; ZFN: Zinc-Finger Nuclease; ODM: Oligo-Directed Mutagenesis; MN: Meganucleases; BE: Base editing.

### Off-target effects considered for CRISPR/Cas-systems

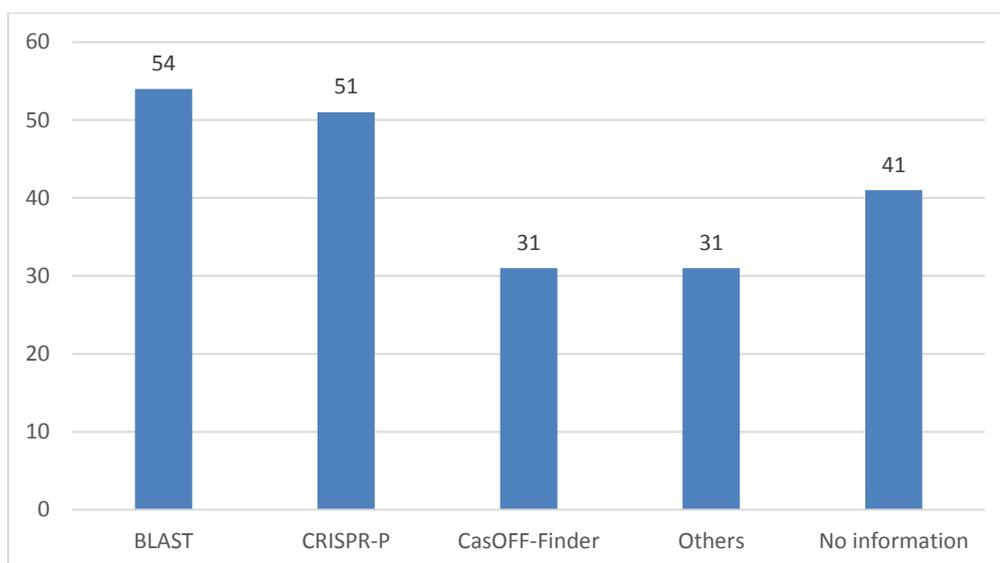
More than 90% of the studies, in which off-target effects were assessed, were conducted with CRISPR/Cas. Figure 5.8 provides an overview of the applied approaches to identify off-target effects. In total, 228 CRISPR/Cas studies dealt with the analysis of off-target effects. 205 studies predicted potential off-target sites and 195 of these identified potential off-target sites. Solely in 188 studies, these potential off-target sites were further analyzed for the occurrence of off-target effects using biased detection methods. In addition, 23 studies with already known potential off-target sites were assessed using biased detection methods. So, in total, 211 studies analyzed potential off-target sites using biased detection methods. Solely, nine studies searched for off-target mutations in a completely unrestricted way using unbiased detection methods. An overview of all identified CRISPR/Cas studies, in which off-target effects were addressed, is provided in Additional file 6 including all extracted data.



**Figure 5.8** Overview of the used approaches to assess off-target effects in CRISPR/Cas studies. All numbers represent the amount of studies conducted for different approaches. In some studies, several approaches were used to analyze off-targets.

### Prediction of potential off-target sites by CRISPR/Cas

Different prediction tools can be used to identify potential off-target sites. All of them have in common that based on sequence alignment programs DNA-sequences are identified in which unintended mutations could occur due to high similarity between the targeted sequence and the potential off-target site. However, the prediction of a potential off-target site is not equated with a real off-target mutation. It has to be shown in a follow up step by verifying the potential off-target site using biased detection methods. 205 CRISPR/Cas studies searched for potential off-target sites. As shown in Figure 5.9, many different prediction tools were used to identify these sites. Mainly, three tools were used to predict potential off-target sites. BLAST was used 54 times, CRISPR-P 51 times and CasOFF-Finder 31 times. 12 other prediction tools were used in 31 studies (for detailed information see Additional file 6). 41 studies did not provide any details about the used tool(s). Three studies used two different prediction tools to identify potential off-target sites.

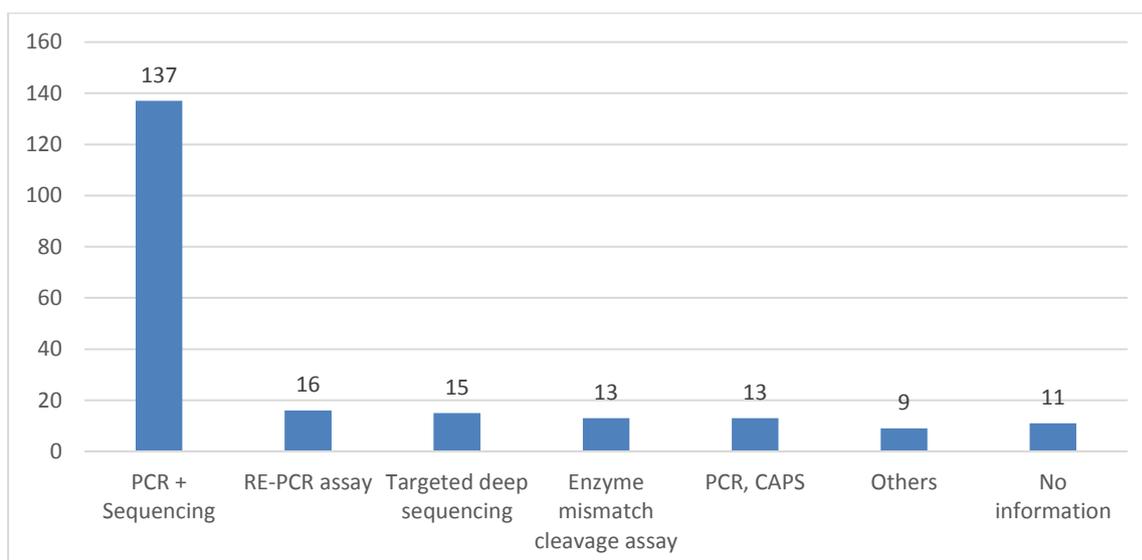


**Figure 5.9** Tools used to predict potential off-target sites in 205 CRISPR/Cas-studies. (1996- May 2018). Three studies employed two tools.

The number of predicted off-target sites varies widely between studies from zero to 4265. Several reasons for this broad heterogeneity exist, which were not extracted in detail within this map but will be elucidated in the discussion.

#### **Detection of off-target effects – biased**

Targeted sequencing to pre-selected sites was applied in 211 CRISPR/Cas studies to detect off-target effects. As shown in Figure 5.8, the predicted off-target sites were frequently analyzed for the occurrence of off-target effects using biased detection methods. In a few studies, potential off-target sites were already known and analyzed without using sequence alignment programs a priori. Figure 5.10 displays the different detection methods applied to identify off-target mutations. In the large majority of studies, off-target effects were detected using a PCR followed by sequencing (n=137). Only a few studies used the detection methods RE-PCR assay (n=16), Targeted deep sequencing (n=15), Enzyme mismatch cleavage assay (n=13) and CAPS analysis (n=13). In 11 studies, no information about the used detection method was provided.



**Figure 5.10** Biased detection methods used to identify off-target effects in CRISPR/Cas studies. Three studies employed two tools. PCR: Polymerase chain reaction; RE-PCR: Restriction enzyme- Polymerase chain reaction; CAPS: Cleaved Amplified Polymorphic Sequences.

Taking all CRISPR/Cas studies together, 1738 different potential off-target sites were analyzed using targeted sequencing. Off-target effects were identified in 55 of these sites, indicating that in around 3% of the analyzed sequences off-target mutations were detected. In another six studies, no information was provided about the amount of analyzed off-target sites but off-target mutations were not identified either.

Considering the different plant species, most of the CRISPR/Cas studies using biased detection methods were conducted in rice ( $n=77$ ), followed by tomato ( $n=23$ ), *Arabidopsis* ( $n=21$ ), different moss species ( $n=13$ ) and soybean ( $n=12$ ) (Table 5.9). In rice, a total amount of 291 potential off-target sites were analyzed and in 25 of these sites, off-target mutations were detected. In contrast, studies conducted in tomatoes solely reported the identification of one off-target mutation when analyzing 222 potential off-target sites.

**Table 5.9** Overview of the amount of studies analyzing potential off-target sites, the amount of analyzed potential off-target sites and the amount of identified off-target effects using biased detection methods in different plant species.

| Plant species | Number of studies analyzing preselected potential off-target sites | Number of analyzed potential off-target sites - biased | Number of identified off-target effects - biased |
|---------------|--|--|--|
| Rice          | 77   | 296  | 25   |
| Tomato        | 23   | 213  | 1  |
| Arabidopsis   | 21   | 229  | 4  |
| Moss          | 13   | 58   | 0  |
| Soybean       | 12   | 106  | 6  |
| Tobacco       | 9  | 50   | 1  |
| Wheat         | 8  | 85   | 6  |
| Maize         | 7  | 23   | 2  |
| Algae         | 4  | 357  | 0  |
| Canola        | 4  | 73   | 0  |
| Cotton        | 4  | 58   | 0  |

| Plant species   | Number of studies analyzing preselected potential off-target sites | Number of analyzed potential off-target sites - biased | Number of identified off-target effects - biased |
|-----------------|--|--|--|
| Orange          | 3  | 21   | 0  |
| Grapevine       | 3  | 13   | 0  |
| Poplar          | 3  | 6  | 0  |
| Grapefruit      | 2  | 15   | 0  |
| Cucumber        | 2  | 9  | 0  |
| Clover          | 2  | 3  | 0  |
| Lettuce         | 1  | 91   | 0  |
| Cacao           | 1  | 9  | 0  |
| Flax            | 1  | 8  | 0  |
| Barley          | 1  | 4  | 1  |
| Kiwifruit       | 1  | 4  | 0  |
| Camelina        | 1  | 3  | 0  |
| Citrange        | 1  | 3  | 0  |
| Watermelon      | 1  | 3  | 0  |
| Cabbage         | 1  | 2  | 1  |
| Carrot          | 1  | 1  | 1  |
| Potato          | 1  | 1  | 0  |
| Sage            | 1  | 1  | 0  |
| Wild strawberry | 1  | 1  | 0  |
| Alfalfa         | 1  | No information   | 1  |

In one study, a different approach was chosen to assess the occurrence of off-target effects [175]. In this study, a series of mismatches were introduced at the sgRNA followed by analyzing whether the targeted sequence was successfully mutated despite the mismatch(es) between the sgRNA and the targeted sequence. Additionally, in this study, the off-target patterns between two PAMs (NGG and NAG) were compared. 22 times the sgRNA was designed in a way that it contained one mismatch to the targeted sequence. 15 of these altered sgRNA induced a DSB at the targeted sequence. 14 times the sgRNA contained two mismatches to the targeted sequence and four of these sgRNA induced a DSB in the targeted sequence. Moreover, eight times the sgRNA contained three mismatches. In these cases, no mutation was identified in the targeted sequence [175]. According to the prediction of off-target effects, the summary provided here does not allow any conclusions to be drawn due to broad heterogeneity.

### **Heterogeneity in CRISPR/Cas-studies regarding the evidence how to predict and detect potential off-target effects**

The number of potential off-target sites called “identified” by the respective authors varies widely between zero and 4265. Several reasons were identified that could explain this broad heterogeneity:

- (i) In total, potential off-target sites were investigated in over 30 different plant species and subspecies. The different genome sizes and the different number of chromosome sets of the individual plants vary widely which influence the number of potential off-target sites.

- (ii) To identify potential off-target sites in CRISPR/Cas-studies 15 different prediction tools were applied.
- (iii) For the detection of off-target effects, various methods have been used, but all of them show their specific advantages and disadvantages and could affect the occurrence of off-target effects [26, 32].
- (iv) The number of hypothetically tolerated mismatches between the target sequence and the potential off-target sites which is an exercise in combinatorics. Ali et al. (2015) [176] identified a total amount of 4265 potential off-target sites in the model organism *Nicotiana benthamiana* for a CRISPR/Cas9 sgRNA. To identify candidate off-target sites, the genome was screened allowing one to seven mismatches to the target sequence. Hence, the more mismatches are tolerated in prediction the higher is the number of potential off-target sites (in the paper: one, two or three mismatches tolerated: No potential off-target sites, four mismatches tolerated: One potential off-target site, five mismatches tolerated: 60 potential off-target sites, six mismatches tolerated: 515 potential off-target sites, seven mismatches tolerated: 3689 potential off-target sites). This indicates that the number of potential off-target sites strongly depends on the number of hypothetically tolerated mismatches. Throughout the available literature, the number of tolerated mismatches predetermined by the researchers was very heterogeneous ranging up to 13 mismatches [177].
- (v) Individual studies also deviate between different structural assumptions when determining potential off-target sites. In most studies, a potential off-target was only counted as such if a PAM followed the potential off-target site. However, in some studies, potential off-target sites were assigned as such without being followed by a PAM (e.g. [44, 178]), although at these sites no DSB can be induced with the specific CRISPR-nucleases used.
- (vi) In some studies, the sgRNA was selected taking into account that no potential off-target effect should occur. Care was taken that, apart from the target sequence, no other sites in the genome possess a similar sequence that could result in an off-target mutation. Thus, the predicted number of potential off-target sites is lower or zero in these studies compared to studies, in which potential off-target sites were not considered a priori in the sequence selection.

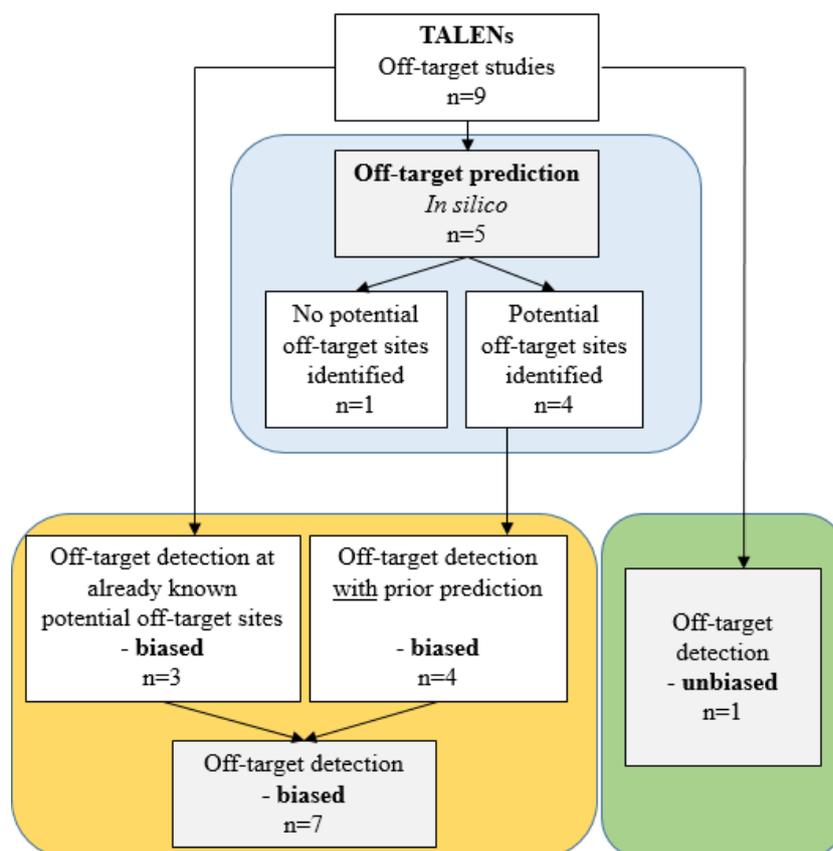
### **Detection of off-target effects – unbiased**

Nine CRISPR/Cas studies were identified using whole genome sequencing (WGS) as an unbiased detection method to identify genome-wide off-target effects (see Additional file 6). No off-target mutations were detected in any of these studies. One study compared both biased and unbiased detection methods. While using unbiased detection methods, no off-target effects were detected, but biased methods detected one off-target [179]. An explanation for this could be that WGS is able to detect higher frequency off-target effects only, but lacks the sensibility required to detect off-target mutations in bulk population [26].

### **Off-target effects considered for TALENs systems**

In the period until May 2018, nine TALENs studies were identified addressing off-target effects. Additional file 6 provides an overview of these studies including all extracted data according to the systematic map protocol. Figure 5.11 maps the different approaches used to analyze off-target effects. Five studies predicted potential off-target sites. Two times, the TAL Effector Nucleotide Targeter 2.0 was used and one time each of the tools PROGNONS, kmasker and Arabidopsis

Information Resource PatMatch. The number of identified potential off-target sites varies widely between zero and 18. In one study, no precise information was given and it was just indicated that many potential off-target sites were identified. The four studies that identified potential off-target sites investigated these for the occurrence of off-target effects using biased detection methods. In addition, three studies with already known potential off-target sites assessed these sites. Different tools were used to examine whether off-target effects occurred. Two studies used enzyme mismatch cleavage assay and one study each PCR + Sequencing, PCR + CAPS and RE-PCR assay. In two studies, no detailed information about the applied detection method was provided. In total, 31 potential off-target sites were analyzed for the occurrence of off-target effects and one of these sites contained an off-target mutation. In one study, an unbiased search for off-target effects was conducted [180]. Using Whole Genome Sequencing (WGS), three off-target mutations were identified which were not present in the wild-type sample. However, the off-target sequences showed no similarity to the TALENs binding sites. Therefore, the occurrence of these mutations cannot be ruled out to be spontaneous ones or sequencing errors [180].



**Figure 5.11** Overview of the used approaches to assess off-target effects in TALENs studies. All numbers represent the amount of studies conducted for different approaches.

### Off-target effects considered for Zinc-finger nucleases systems

Four studies dealt with the analysis of off-target effects when applying ZFN in plants. Detailed information is provided in additional file 6. All studies identified putative off-target sites that are most closely related to the target sequences. Two times, the database PLantGDB was used to identify potential off-target sites and two times no detailed information was provided. In total, 10 potential off-target sites were detected which were then screened for the occurrence of off-target effects. No off-target mutations were detected in any of these analyzed, potential off-target sites.

### Off-target effects considered for Meganuclease systems

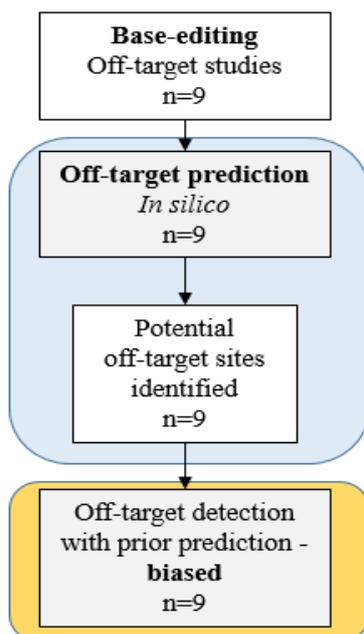
Solely one MN study was identified that analyzed off-target effects (Additional file 6). In this study, a homologous sequence to the target one was identified that differs solely in two nucleotides. However, when screening this potential off-target site no modification was identified.

### Off-target effects considered for Oligonucleotide-directed mutagenesis systems

One study addressed the occurrence of off-target effects in rice using the ODM technique (Additional file 6). Beside the targeted Acetolactate synthase (ALS) sequence, the whole coding region of ALS gene was sequenced but no further mutation was detected.

### Off-target effects considered for base editing systems

In total, nine BE studies analyzed the occurrence of off-target effects (Additional file 6). However, one study could only rarely be evaluated due to language barriers [181]. The approach used to identify off-target effects was similar in all studies. In a first step, potential off-target sites were predicted using different tools (2x CRISPR-GE tool, 2x CRISPR-P, 1x CasOff-Finder, 3 x no information). According to the other genome editing techniques, the number of potential off-target sites varies widely between one and nine. In a second step, all predicted potential off-target sites were sequenced for the occurrence of off-target effects using PCR + Sequencing method (n=6), Targeted Deep Sequencing (n=2) or Enzyme mismatch cleavage assay (n=1) (Figure 5.12). One off-target mutation was identified.



**Figure 5.12** Overview of the used approaches to assess off-target effects in BE studies. All numbers represent the amount of studies conducted for different approaches.

## 5.6 Discussion

### 5.6.1 Market-oriented applications of genome editing

This map documents the state of evidence for the application of genome editing as a new tool for the modification of plant traits and the associated potential occurrence of off-target effects. The publication rate of primary research has risen sharply since the CRISPR/Cas technique was first

applied in plants in 2013. It is worth mentioning that in total primary studies from 33 countries were identified but nearly three quarter of these studies originate from either China (40%) or the USA (33%). For comparison, Japan and Germany published around 6% of the studies each and no other country contributed more than 2% of the total. Summarizing the number of published studies by continent, more than 50% of the studies were conducted in an Asian country (53%), around one third in North American countries (34%) and only 13% in European countries. As the genome editing techniques, especially CRISPR/Cas, were just recently developed, the large majority of the existing applications represent basic research. Nevertheless, almost 100 different applications aimed to produce beneficial agricultural traits in 28 different agricultural crops. We determined that the majority of such “market-oriented” applications have been carried out in economically important crops such as maize, rice, wheat and soybeans, but less economically important crops such as cucumber, lettuce, peanut or grapefruit have been worked on as well. The market-oriented applications address several breeding objectives including yield improvement, improved growth characteristics, improved food and feed quality, tolerances to biotic and abiotic stress, herbicide tolerance and industrial utilization. This indicates that genome editing is able to address beneficial traits for the agricultural value chain and could contribute to food security and the environmental management. To confirm this hypothesis a systematic review would be required to quantify such effects.

### 5.6.2 Off-target effects of genome editing

252 studies investigated the occurrence of off-target effects following a genome editing application in plants. Most of these studies were performed with the genome editing technique CRISPR/Cas (n=228), whereas only a few studies used TALENs (n=9), BE (n=9), ZFN (n=4), MN (n=1) or ODM (n=1). Reasons for these findings are that the total number of studies in which CRISPR/Cas was applied is much larger compared to the other genome editing techniques. Additionally, CRISPR/Cas is more susceptible to off-target effects compared to other nuclease techniques such as ZFN and TALENs, as it works as a monomer, whereas ZFN and TALENs work as dimers [182]. In addition, the sgRNA used in CRISPR/Cas-applications is able to tolerate several mismatches leading to the induction of a DSB at a site in the DNA that is similar but not identical to the targeted sequence [32]. In most CRISPR/Cas studies biased detection methods were used, meaning that potential off-target sites were first predicted using bioinformatics programs followed by targeted sequencing of these sites for the occurrence of off-target effects. The biased method suggests a very broad data basis, but as shown in the result section, individual studies are very heterogeneous in their structure and design. Heterogeneity was identified regarding the plant species, the CRISPR-variant, the prediction tools and detection methods used, the amount of tolerated mismatches and the chosen sgRNA. In order to allow any conclusions to be drawn about the occurrence of off-target effects a more in-depth analysis e.g. by a systematic review is mandatory.

### 5.6.3 Knowledge gaps

Regarding secondary question 2, several topics have been identified representing knowledge gaps where no studies or only a small number of studies exist. A knowledge gap exists for the analysis of off-target effects in TALENs, ZFN, MN, ODM and BE. Only nine studies have been identified analyzing the occurrence of off-target effects in TALENs and BE studies. The amount of off-target studies for ZFN (n=4), ODM (n=1) and MN (n=1) were even lower. For a more in-depth off-target

analysis regarding these techniques, further research and more primary studies are needed. However, the CRISPR/Cas technique is applied much more frequently because of efficiency, time saving and cost-effectiveness compared to TALENs, ZFN, ODM and MN [182]. Therefore, it is questionable what additional research effort is justified. BE differs as it has great potential to support plant breeding, though it is not broadly established yet. It is based on the CRISPR/Cas9 system but enables exchanging individual base pairs at specific sites without inducing a DSB. Our map shows that only a few BE studies addressed the occurrence of off-target effects so far. The likely reason is that BE is a “young” method in plants, and its first publication was in 2017. Further research about BE including the analysis of off-target effects is highly recommended. Another knowledge gap affecting all genome editing techniques is the genome-wide analysis of off-target effects in a completely unrestricted way. Only nine studies addressed this aspect using the CRISPR/Cas technique and one using the TALENs technique. No study analyzed off-target effects using unbiased detection methods in ZFN, ODM, MN and BE studies.

#### 5.6.4 Knowledge clusters

The analysis of off-target effects in CRISPR/Cas studies using biased detection methods represents a knowledge cluster. This cluster is amenable for a critical appraisal serving the interest of researchers and decision-makers. The following parameters could affect the occurrence of off-targets and are worth to be analyzed in-depth:

- (i) Genome composition, size and ploidy level of different plant species.
- (ii) Amount of tolerated mismatches at the potential off-target sites.
- (iii) Quality of available biased detection methods.
- (iv) CRISPR variants (e.g. Nickase).
- (v) Chosen sgRNA.
- (vi) Methodology of SDN delivery.

#### 5.6.5 Limitations of the systematic map

References were only searched in German and English language. Publications in an Asian language, in which only the abstract was available in English, were also identified but could not be included in this map for the extraction of detailed data. Since most of the studies were found in scientific journals, a bias in the pool of articles found is that publications from countries that probably use genome editing but do not publish in either English or German language are not represented. For example, it can be assumed that in South American countries, considerably more research is done than the identified literature suggests. Identifying grey literature as well as manually identifying scientific journals publishing in the local language(s) was also not possible due to language barriers.

Additionally, the full text of 104 articles, that have been rated as relevant on title/abstract level, were un-retrievable and were therefore not included in the systematic map (see Additional file 3).

It has been demonstrated that the individual studies addressing off-target effects differ widely in design and implementation. Therefore, no reliable conclusions about the occurrence of off-target

effects can be drawn based on the results of this map. A critical appraisal of the individual studies in form of a systematic review is recommended.

## 5.7 Conclusions

### 5.7.1 Implications for policy/management

This systematic map identified substantial bodies of evidence regarding the applications of different genome editing techniques in plants as well as the occurrence of off-target effects. Until May 2018, almost 100 market-oriented applications were identified including improved food and feed quality, yield improvement, altered growth characteristics, resistance against biotic and abiotic stress, herbicide tolerance and industrial utilization. The wide range of different applications addressing all parts in the agricultural value chain and the application in many different plant species indicates that genome editing became a promising tool to breed varieties that are better adapted to the needs of agriculture and to enable a valuable contribution to food security and the environment.

A decisive factor that impacts the use of genome editing is the acceptance by consumers and retailers for products derived from innovative plant breeding techniques. A prominent point of criticism in this context is the occurrence of off-target effects. Since plant breeding is the contextual background, it is worth to compare the occurrence of off-target effects in the context of naturally occurring mutations and routinely used breeding techniques like regular crossing or undirected mutagenesis using chemical mutagens or irradiation [10].

### 5.7.2 Implication for research

Results of this systematic map determined that different approaches were used to analyze off-target effects depending on the plant species and with regard to the CRISPR-variant, the prediction tools and detection methods used, the amount of tolerated mismatches and the chosen sgRNA. A critical appraisal in the course of a systematic review could help to identify parameters in order to further reduce the occurrence of off-target effects. The identified knowledge cluster for the detection of off-target effects in CRISPR-studies using biased detection methods would be suitable to address this.

The results of this map further identified a knowledge gap regarding the analysis of off-target effects using unbiased detection tools. To increase this data basis and to evaluate whether an unbiased off-target analysis has an added value compared to biased of target analysis, more studies should assess this aspect, in future.

#### Additional files

<https://environmentalevidencejournal.biomedcentral.com/articles/10.1186/s13750-019-0171-5#Sec49>

## 5.8 References

1. Becker H. Pflanzenzüchtung. 2nd ed. Stuttgart: UTB; 2011.

2. Ronald P. Plant genetics, sustainable agriculture and global food security. *Genetics*. 2011;188:11–20. doi:10.1534/genetics.111.128553.
3. FAO - Committee on Agriculture (17th Session). 06.03.2003. <http://www.fao.org/docrep/meeting/006/y8704e.htm>. Accessed 25 Sep 2018.
4. Xie K, Wu S, Li Z, Zhou Y, Zhang D, Dong Z, et al. Map-based cloning and characterization of *Zea mays* male sterility33 (*ZmMs33*) gene, encoding a glycerol-3-phosphate acyltransferase. *Theor Appl Genet*. 2018;131:1363–78. doi:10.1007/s00122-018-3083-9.
5. Phalan B. What Have We Learned from the Land Sparing-sharing Model? *Sustainability*. 2018;10:1760. doi:10.3390/su10061760.
6. Stevenson JR, Villoria N, Byerlee D, Kelley T, Maredia M. Green Revolution research saved an estimated 18 to 27 million hectares from being brought into agricultural production. *Proc Natl Acad Sci U S A*. 2013;110:8363–8. doi:10.1073/pnas.1208065110.
7. Songstad DD, Petolino JF, Voytas DF, Reichert NA. Genome Editing of Plants. *Critical Reviews in Plant Sciences*. 2017;36:1–23. doi:10.1080/07352689.2017.1281663.
8. International Atomic Energy Agency (IAEA). Mutant Variety Database. <https://mvd.iaea.org/#!/Search?page=1&size=15&sortby=Name&sort=ASC>. Accessed 25 Sep 2018.
9. Schlegel RHJ. Dictionary of plant breeding. 2nd ed. Boca Raton, Fla.: CRC Press; 2009.
10. SAM High Level Group of Scientific Advisors. New techniques in Agricultural Biotechnology. 2017. ISBN 978-92-79-66222-5. <https://doi.org/10.2777/57449> 8. <https://ec.europa.eu/research/sam/index.cfm?pg=agribiotechology#>. Accessed 18 July 2017.
11. Barton KA, Binns AN, Matzke AJM, Chilton M-D. Regeneration of intact tobacco plants containing full length copies of genetically engineered T-DNA, and transmission of T-DNA to R1 progeny. *Cell*. 1983;32:1033–43. doi:10.1016/0092-8674(83)90288-X.
12. Herrera-Estrella L, Depicker A, van Montagu M, Schell J. Expression of chimaeric genes transferred into plant cells using a Ti-plasmid-derived vector. *Nature*. 1983;303:209–13. doi:10.1038/303209a0.
13. Govindan G, Ramalingam S. Programmable Site-Specific Nucleases for Targeted Genome Engineering in Higher Eukaryotes. *J Cell Physiol*. 2016;231:2380–92. doi:10.1002/jcp.25367.
14. European Food Safety Authority (EFSA). Scientific opinion addressing the safety assessment of plants developed using Zinc Finger Nuclease 3 and other Site-Directed Nucleases with similar function. *EFSA J*. 2012;10(10):2943.
15. van de Wiel CCM, Schaart JG, Lotz LAP, Smulders MJM. New traits in crops produced by genome editing techniques based on deletions. *Plant Biotechnol Rep*. 2017;11:1–8. doi:10.1007/s11816-017-0425-z.
16. Podevin N, Davies HV, Hartung F, Nogué F, Casacuberta JM. Site-directed nucleases: A paradigm shift in predictable, knowledge-based plant breeding. *Trends Biotechnol*. 2013;31:375–83. doi:10.1016/j.tibtech.2013.03.004.

17. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*. 2016;533:420–4. doi:10.1038/nature17946.
18. Nationale Akademie der Wissenschaften Leopoldina, Deutsche Forschungsgemeinschaft, acatech – Deutsche Akademie der Technikwissenschaften, Union der deutschen Akademien der Wissenschaften. *The opportunities and limits of genome editing*. 2015. ISBN: 978-3-8047-3493-7
19. Bömeke O, Kahrman J, Matthies A. Detaillierte Übersicht zum regulatorischen Status der neuen molekularbiologischen Techniken (NMT) in ausgewählten Drittstaaten. [https://www.bmel.de/DE/Landwirtschaft/Pflanzenbau/Gentechnik/\\_Texte/Neue\\_molekular\\_biological\\_Techniken.html](https://www.bmel.de/DE/Landwirtschaft/Pflanzenbau/Gentechnik/_Texte/Neue_molekular_biological_Techniken.html). Accessed 21 Feb 2019.
20. Duan Y-B, Li J, Qin R-Y, Xu R-F, Li H, Yang Y-C, et al. Identification of a regulatory element responsible for salt induction of rice OsRAV2 through ex situ and in situ promoter analysis. *Plant Mol Biol*. 2016;90:49–62. doi:10.1007/s11103-015-0393-z.
21. Demorest ZL, Coffman A, Baltes NJ, Stoddard TJ, Clasen BM, Luo S, et al. Direct stacking of sequence-specific nuclease-induced mutations to produce high oleic and low linolenic soybean oil. *BMC Plant Biol*. 2016;16:225. doi:10.1186/s12870-016-0906-1.
22. Chandrasekaran J, Brumin M, Wolf D, Leibman D, Klap C, Pearlsman M, et al. Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol Plant Pathol*. 2016;17:1140–53. doi:10.1111/mpp.12375.
23. Gocal GFW, Schöpke C, Beetham PR. Oligo-Mediated Targeted Gene Editing. In: *Advances in New Technology for Targeted Modification of Plant Genomes*; 2015. p. 73–89. doi:10.1007/978-1-4939-2556-8\_5.
24. Shi J, Gao H, Wang H, Lafitte HR, Archibald RL, Yang M, et al. ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. *Plant Biotechnol J*. 2017;15:207–16. doi:10.1111/pbi.12603.
25. Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, Qiu J-L. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat Biotechnol*. 2014;32:947–51. doi:10.1038/nbt.2969.
26. Martin F, Sánchez-Hernández S, Gutiérrez-Guerrero A, Pinedo-Gomez J, Benabdellah K. Biased and Unbiased Methods for the Detection of Off-Target Cleavage by CRISPR/Cas9: An Overview. *Int J Mol Sci* 2016. doi:10.3390/ijms17091507.
27. Altschul S, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. *J Mol Biol*. 1990;215:403–10. doi:10.1006/jmbi.1990.9999.
28. Bae S, Park J, Kim J-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*. 2014;30:1473–5. doi:10.1093/bioinformatics/btu048.
29. Lei Y, Lu L, Liu H-Y, Li S, Xing F, Chen L-L. CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. *Mol Plant*. 2014;7:1494–6. doi:10.1093/mp/ssu044.
30. Liu H, Ding Y, Zhou Y, Jin W, Xie K, Chen L-L. CRISPR-P 2.0: An Improved CRISPR-Cas9 Tool for Genome Editing in Plants. *Mol Plant*. 2017;10:530–2. doi:10.1016/j.molp.2017.01.003.

31. Tycko J, Myer VE, Hsu PD. Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity. *Mol Cell*. 2016;63:355–70. doi:10.1016/j.molcel.2016.07.004.
32. Zischewski J, Fischer R, Bortesi L. Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnol Adv*. 2017;35:95–104. doi:10.1016/j.biotechadv.2016.12.003.
33. Agapito-Tenfen SZ, Wikmark O-G. Current status of emerging technologies for plant breeding. Biosafety and knowledge gaps of site directed nucleases and oligonucleotide-directed mutagenesis. 2015. GenØk - Centre for Biosafety.
34. Zhang X-H, Tee LY, Wang X-G, Huang Q-S, Yang S-H. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol Ther Nucleic Acids*. 2015;4:e264. doi:10.1038/mtna.2015.37.
35. Kanchiswamy CN, Maffei M, Malnoy M, Velasco R, Kim J-S. Fine-Tuning Next-Generation Genome Editing Tools. *Trends Biotechnol*. 2016;34:562–74. doi:10.1016/j.tibtech.2016.03.007.
36. Collaboration for Environmental Evidence. Guidelines for Systematic Review and Evidence Synthesis in Environmental Management. 2013. [www.environmentalevidence.org/Documents/Guidelines/Guidelines4.2.pdf](http://www.environmentalevidence.org/Documents/Guidelines/Guidelines4.2.pdf). Accessed 25 Feb 2019.
37. Modrzejewski D, Hartung F, Sprink T, Krause D, Kohl C, Schiemann J, Wilhelm R. What is the available evidence for the application of genome editing as a new tool for plant trait modification and the potential occurrence of associated off-target effects: a systematic map protocol. *Environ Evid*. 2018;7:11. doi:10.1186/s13750-018-0130-6.
38. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*. 1960;20:37–46. doi:10.1177/001316446002000104.
39. Kohl C, McIntosh EJ, Unger S, Haddaway NR, Kecke S, Schiemann J, Wilhelm R. Online tools supporting the conduct and reporting of systematic reviews and systematic maps: a case study on CADIMA and review of existing tools. *Environ Evid*. 2018;7:2420. doi:10.1186/s13750-018-0115-5.
40. Haddaway N, Macura B, Whaley P, Pullin A. ROSES Flow Diagram for Systematic Maps. Version 1.0: Figshare; 2018.
41. Braatz J, Harloff H-J, Mascher M, Stein N, Himmelbach A, Jung C. CRISPR-Cas9 Targeted Mutagenesis Leads to Simultaneous Modification of Different Homoeologous Gene Copies in Polyploid Oilseed Rape (*Brassica napus*). *Plant Physiol*. 2017;174:9. doi:10.1104/pp.17.00426.
42. Yang Y, Zhu K, Li H, Han S, Meng Q, Khan SU, et al. Precise editing of CLAVATA genes in *Brassica napus* L. regulates multilocular silique development. *Plant Biotechnol J*. 2018;16:1322–35. doi:10.1111/pbi.12872.
43. Wang Y, Meng Z, Liang C, Meng Z, Wang Y, Sun G, et al. Increased lateral root formation by CRISPR/Cas9-mediated editing of arginase genes in cotton. *Sci China Life Sci*. 2017;60:524–7. doi:10.1007/s11427-017-9031-y.
44. Hu B, Li D, Liu X, Qi J, Gao D, Zhao S, et al. Engineering Non-transgenic Gynoecious Cucumber Using an Improved Transformation Protocol and Optimized CRISPR/Cas9 System. *Mol Plant*. 2017;10:1575–8. doi:10.1016/j.molp.2017.09.005.

45. Bertier LD, Ron M, Huo H, Bradford KJ, Britt AB, Michelmore RW. High-Resolution Analysis of the Efficiency, Heritability, and Editing Outcomes of CRISPR/Cas9-Induced Modifications of NCED4 in Lettuce (*Lactuca sativa*). *G3* (Bethesda). 2018;8:1513–21. doi:10.1534/g3.117.300396.
46. United States Department of Agriculture (USDA). 2015. [https://www.aphis.usda.gov/biotechnology/downloads/reg\\_loi/15-013-01air.pdf](https://www.aphis.usda.gov/biotechnology/downloads/reg_loi/15-013-01air.pdf). Accessed 25 Aug 2018.
47. Huang C, Sun H, Xu D, Chen Q, Liang Y, Wang X, et al. ZmCCT9 enhances maize adaptation to higher latitudes. *Proc Natl Acad Sci U S A*. 2018;115:E334–E341. doi:10.1073/pnas.1718058115.
48. Svitashv S, Young JK, Schwartz C, Gao H, Falco SC, Cigan AM. Targeted Mutagenesis, Precise Gene Editing, and Site-Specific Gene Insertion in Maize Using Cas9 and Guide RNA. *Plant Physiol*. 2015;169:931–45. doi:10.1104/pp.15.00793.
49. Svitashv S, Schwartz C, Lenderts B, Young JK, Mark Cigan A. Genome editing in maize directed by CRISPR-Cas9 ribonucleoprotein complexes. *Nat Commun*. 2016;7:1–7. doi:10.1038/ncomms13274.
50. Li J, Zhang H, Si X, Tian Y, Chen K, Liu J, et al. Generation of thermosensitive male-sterile maize by targeted knockout of the ZmTMS5 gene. *J Genet Genomics*. 2017;44:465–8. doi:10.1016/j.jgg.2017.02.002.
51. Kelliher T, Starr D, Richbourg L, Chintamanani S, Delzer B, Nuccio ML, et al. MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. *Nature*. 2017;542:105–9. doi:10.1038/nature20827.
52. Clasen BM, Stoddard TJ, Luo S, Demorest ZL, Li J, Cedrone F, et al. Improving cold storage and processing traits in potato through targeted gene knockout. *Plant Biotechnol J*. 2016;14:169–76. doi:10.1111/pbi.12370.
53. Li M, Li X, Zhou Z, Wu P, Fang M, Pan X, et al. Reassessment of the Four Yield-related Genes Gn1a, DEP1, GS3, and IPA1 in Rice Using a CRISPR/Cas9 System. *Front Plant Sci*. 2016;7:1–13. doi:10.3389/fpls.2016.00377.
54. Shen L, Wang C, Fu Y, Wang J, Liu Q, Zhang X, et al. QTL editing confers opposing yield performance in different rice varieties. *J Integr Plant Biol* 2016. doi:10.1111/jipb.12501.
55. Xu R, Yang Y, Qin R, Li H, Qiu C, Li L, et al. Rapid improvement of grain weight via highly efficient CRISPR/Cas9-mediated multiplex genome editing in rice. *J Genet Genomics*. 2016;43:529–32. doi:10.1016/j.jgg.2016.07.003.
56. Hu Z, Lu S-J, Wang M-J, He H, Le Sun, Wang H, et al. A Novel QTL q TGW3 Encodes the GSK3/SHAGGY-Like Kinase OsGSK5/OsSK41 that Interacts with OsARF4 to Negatively Regulate Grain Size and Weight in Rice. *Mol Plant*. 2018;11:736–49. doi:10.1016/j.molp.2018.03.005.
57. Shen Lan, Li Jian, Fu Yaping, Wang Junjie, Hua Yufeng, Jiao Xiaozhen, Yan Changjie, Wang Kejian. Orientation Improvement of Grain Length and Grain Number in Rice by Using CRISPR/Cas9 System 2017. doi:10.16819/j.1001-7216.2017.7029.
58. Ji X, Li F, Yan Y, Sun HZ, Zhang J, Li JZ, et al. CRISPR/Cas9 System-Based Editing of Phytochrome-Interacting Factor OsPIL15 2017. doi:10.3864/j.issn.0578-1752.2017.15.002.

59. Shen L, Hua Y, Fu Y, Li J, Liu Q, Jiao X, et al. Rapid generation of genetic diversity by multiplex CRISPR/Cas9 genome editing in rice. *Sci China Life Sci.* 2017;60:506–15. doi:10.1007/s11427-017-9008-8.
60. Lu K, Wu B, Wang J, Zhu W, Nie H, Qian J, et al. Blocking amino acid transporter OsAAP3 improves grain yield by promoting outgrowth buds and increasing tiller number in rice. *Plant Biotechnol J.* 2018;50:1416. doi:10.1111/pbi.12907.
61. Liao Y, Bai Q, Xu P, Wu T, Guo D, Peng Y, et al. Mutation in Rice Abscisic Acid2 Results in Cell Death, Enhanced Disease-Resistance, Altered Seed Dormancy and Development. *Front. Plant Sci.* 2018;9:1248. doi:10.3389/fpls.2018.00405.
62. Li X, Zhou W, Ren Y, Tian X, Lv T, Wang Z, et al. High-efficiency breeding of early-maturing rice cultivars via CRISPR/Cas9-mediated genome editing. *J Genet Genomics.* 2017;44:175–8. doi:10.1016/j.jgg.2017.02.001.
63. Lee S-K, Eom J-S, Hwang S-K, Shin D, An G, Okita TW, Jeon J-S. Plastidic phosphoglucomutase and ADP-glucose pyrophosphorylase mutants impair starch synthesis in rice pollen grains and cause male sterility. *J Exp Bot.* 2016;67:5557–69. doi:10.1093/jxb/erw324.
64. Li Q, Zhang D, Chen M, Liang W, Wei J, Qi Y, Yuan Z. Development of japonica Photo-Sensitive Genic Male Sterile Rice Lines by Editing Carbon Starved Anther Using CRISPR/Cas9. *J Genet Genomics.* 2016;43:415–9. doi:10.1016/j.jgg.2016.04.011.
65. Xie Y, Niu B, Long Y, Li G, Tang J, Zhang Y, et al. Suppression or knockout of SaF/SaM overcomes the Sa-mediated hybrid male sterility in rice. *J Integr Plant Biol.* 2017;59:669–79. doi:10.1111/jipb.12564.
66. Zhou H, He M, Li J, Chen L, Huang Z, Zheng S, et al. Development of Commercial Thermo-sensitive Genic Male Sterile Rice Accelerates Hybrid Rice Breeding Using the CRISPR/Cas9-mediated TMS5 Editing System. *Sci Rep.* 2016;6:1–12. doi:10.1038/srep37395.
67. Zou T, He Z, Qu L, Liu M, Zeng J, Liang Y, et al. Knockout of OsACOS12 caused male sterility in rice. *Mol Breeding.* 2017;37:437. doi:10.1007/s11032-017-0722-9.
68. Zou T, Xiao Q, Li W, Luo T, Yuan G, He Z, et al. OsLAP6/OsPKS1, an orthologue of Arabidopsis PKSA/LAP6, is critical for proper pollen exine formation. *Rice (N Y).* 2017;10:615. doi:10.1186/s12284-017-0191-0.
69. Liu L, Zheng C, Kuang B, Wei L, Yan L, Wang T. Receptor-Like Kinase RUPO Interacts with Potassium Transporters to Regulate Pollen Tube Growth and Integrity in Rice. *PLoS Genet.* 2016;12:e1006085. doi:10.1371/journal.pgen.1006085.
70. Ma L, Zhu F, Li Z, Zhang J, Li X, Dong J, Wang T. TALEN-Based Mutagenesis of Lipoxxygenase LOX3 Enhances the Storage Tolerance of Rice (*Oryza sativa*) Seeds. *PLoS ONE.* 2015;10:e0143877. doi:10.1371/journal.pone.0143877.
71. Qian W, Wu C, Fu Y, Hu G, He Z, Liu W. Novel rice mutants overexpressing the brassinosteroid catabolic gene CYP734A4. *Plant Mol Biol.* 2017;93:197–208. doi:10.1007/s11103-016-0558-4.
72. Yuan J, Chen S, Jiao W, Wang L, Wang L, Ye W, et al. Both maternally and paternally imprinted genes regulate seed development in rice. *New Phytol.* 2017;216:373–87. doi:10.1111/nph.14510.

73. Lu Y, Zhu J-K. Precise Editing of a Target Base in the Rice Genome Using a Modified CRISPR/Cas9 System. *Mol Plant*. 2017;10:523–5. doi:10.1016/j.molp.2016.11.013.
74. Wang Y, Geng L, Yuan M, Wei J, Jin C, Li M, et al. Deletion of a target gene in Indica rice via CRISPR/Cas9. *Plant Cell Rep*. 2017;36:1333–43. doi:10.1007/s00299-017-2158-4.
75. Huang Y, Guo Y, Liu Y, Zhang F, Wang Z, Wang H, et al. 9-cis-Epoxycarotenoid Dioxygenase 3 Regulates Plant Growth and Enhances Multi-Abiotic Stress Tolerance in Rice. *Front. Plant Sci*. 2018;9:1248. doi:10.3389/fpls.2018.00162.
76. Cai Y, Chen L, Liu X, Guo C, Sun S, Wu C, et al. CRISPR/Cas9-mediated targeted mutagenesis of GmFT2a delays flowering time in soya bean. *Plant Biotechnol J*. 2017;1–10. doi:10.1111/pbi.12758.
77. Liu Y, Merrick P, Zhang Z, Ji C, Yang B, Fei S-z. Targeted mutagenesis in tetraploid switchgrass (*Panicum virgatum* L.) using CRISPR/Cas9. *Plant Biotechnol J*. 2018;16:381–93. doi:10.1111/pbi.12778.
78. Ito Y, Nishizawa-Yokoi A, Endo M, Mikami M, Toki S. CRISPR/Cas9-mediated mutagenesis of the RIN locus that regulates tomato fruit ripening. *Biochem Biophys Res Commun*. 2015;467:76–82. doi:10.1016/j.bbrc.2015.09.117.
79. Lor VS, Starker CG, Voytas DF, Weiss D, Olszewski NE. Targeted mutagenesis of the tomato PROCERA gene using transcription activator-like effector nucleases. *Plant Physiol*. 2014;166:1288–91. doi:10.1104/pp.114.247593.
80. Soyk S, Müller NA, Park SJ, Schmalenbach I, Jiang K, Hayama R, et al. Variation in the flowering gene SELF PRUNING 5G promotes day-neutrality and early yield in tomato. *Nat Genet*. 2017;49:162–8. doi:10.1038/ng.3733.
81. United States Department of Agriculture (USDA). 2018. [https://www.aphis.usda.gov/biotechnology/downloads/reg\\_loi/18-051-01\\_air\\_response\\_signed.pdf](https://www.aphis.usda.gov/biotechnology/downloads/reg_loi/18-051-01_air_response_signed.pdf). Accessed 25 Aug 2018.
82. Rodríguez-Leal D, Lemmon ZH, Man J, Bartlett ME, Lippman ZB. Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing. *Cell*. 2017;171:470–480.e8. doi:10.1016/j.cell.2017.08.030.
83. Filler Hayut S, Melamed Bessudo C, Levy AA. Targeted recombination between homologous chromosomes for precise breeding in tomato. *Nat Commun*. 2017;8:15605. doi:10.1038/ncomms15605.
84. Dahan-Meir T, Filler-Hayut S, Melamed-Bessudo C, Bocobza S, Czosnek H, Aharoni A, Levy AA. Efficient in planta gene targeting in tomato using geminiviral replicons and the CRISPR/Cas9 system. *The Plant Journal*. 2018;95:5–16. doi:10.1111/tpj.13932.
85. Deng L, Wang H, Sun C, Li Q, Jiang H, Du M, et al. Efficient generation of pink-fruited tomatoes using CRISPR/Cas9 system. *J Genet Genomics*. 2018;45:51–4. doi:10.1016/j.jgg.2017.10.002.
86. Wang W, Pan Q, He F, Akhunova A, Chao S, Trick H, Akhunov E. Transgenerational CRISPR-Cas9 Activity Facilitates Multiplex Gene Editing in Allopolyploid Wheat. *The CRISPR Journal*. 2018;1:65–74. doi:10.1089/crispr.2017.0010.
87. Zhang Y, Li D, Zhang D, Zhao X, Cao X, Dong L, et al. Analysis of the functions of TaGW2 homoeologs in wheat grain weight and protein content traits. *The Plant Journal*. 2018;94:857–66. doi:10.1111/tpj.13903.

88. Zhou J, Wang G, Liu Z. Efficient genome editing of wild strawberry genes, vector development and validation. *Plant Biotechnol J*. 2018;166:1292. doi:10.1111/pbi.12922.
89. United States Department of Agriculture (USDA). 2017. [https://www.aphis.usda.gov/biotechnology/downloads/reg\\_loi/17-038-02\\_air\\_inquiry\\_cbidel.pdf](https://www.aphis.usda.gov/biotechnology/downloads/reg_loi/17-038-02_air_inquiry_cbidel.pdf). Accessed 25 Aug 2018.
90. Ozseyhan ME, Kang J, Mu X, Lu C. Mutagenesis of the FAE1 genes significantly changes fatty acid composition in seeds of *Camelina sativa*. *Plant Physiol Biochem*. 2018;123:1–7. doi:10.1016/j.plaphy.2017.11.021.
91. Jiang WZ, Henry IM, Lynagh PG, Comai L, Cahoon EB, Weeks DP. Significant enhancement of fatty acid composition in seeds of the allohexaploid, *Camelina sativa*, using CRISPR/Cas9 gene editing. *Plant Biotechnol J*. 2017;15:648–57. doi:10.1111/pbi.12663.
92. Morineau C, Bellec Y, Tellier F, Gissot L, Kelemen Z, Nogué F, Faure J-D. Selective gene dosage by CRISPR-Cas9 genome editing in hexaploid *Camelina sativa*. *Plant Biotechnol J*. 2017;15:729–39. doi:10.1111/pbi.12671.
93. Aznar-Moreno JA, Durrett TP. Simultaneous Targeting of Multiple Gene Homeologs to Alter Seed Oil Production in *Camelina sativa*. *Plant Cell Physiol*. 2017;58:1260–7. doi:10.1093/pcp/pcx058.
94. Okuzaki A, Ogawa T, Koizuka C, Kaneko K, Inaba M, Imamura J, Koizuka N. CRISPR/Cas9-mediated genome editing of the fatty acid desaturase 2 gene in *Brassica napus*. *Plant Physiol Biochem* 2018. doi:10.1016/j.plaphy.2018.04.025.
95. United States Department of Agriculture (USDA). 2015. [https://www.aphis.usda.gov/biotechnology/downloads/reg\\_loi/15-352-01\\_air\\_inquiry\\_cbidel.pdf](https://www.aphis.usda.gov/biotechnology/downloads/reg_loi/15-352-01_air_inquiry_cbidel.pdf). Accessed 25 Aug 2018.
96. Qi X, Le Dong, Liu C, Mao L, Liu F, Zhang X, et al. Systematic identification of endogenous RNA polymerase III promoters for efficient RNA guide-based genome editing technologies in maize. *The Crop Journal*. 2018;6:314–20. doi:10.1016/j.cj.2018.02.005.
97. United States Department of Agriculture (USDA). 2015. [https://www.aphis.usda.gov/biotechnology/downloads/reg\\_loi/15-078-02\\_air\\_inquiry.pdf](https://www.aphis.usda.gov/biotechnology/downloads/reg_loi/15-078-02_air_inquiry.pdf). Accessed 25 Aug 2018.
98. United States Department of Agriculture (USDA). 2010. [https://www.aphis.usda.gov/biotechnology/downloads/reg\\_loi/DOW\\_Email\\_%20to\\_Susan\\_%20Kohler\\_032010.pdf](https://www.aphis.usda.gov/biotechnology/downloads/reg_loi/DOW_Email_%20to_Susan_%20Kohler_032010.pdf). Accessed 25 Aug 2018.
99. Shukla VK, Doyon Y, Miller JC, DeKolver RC, Moehle EA, Worden SE, et al. Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature*. 2009;459:437–41. doi:10.1038/nature07992.
100. United States Department of Agriculture (USDA). 2015. [https://www.aphis.usda.gov/biotechnology/downloads/reg\\_loi/15-321-01\\_air\\_inquiry.pdf](https://www.aphis.usda.gov/biotechnology/downloads/reg_loi/15-321-01_air_inquiry.pdf). Accessed 25 Aug 2018.
101. Alagoz Y, Gurkok T, Zhang B, Unver T. Manipulating the Biosynthesis of Bioactive Compound Alkaloids for Next-Generation Metabolic Engineering in Opium Poppy Using CRISPR-Cas 9 Genome Editing Technology. *Sci Rep*. 2016;6:30910–8. doi:10.1038/srep30910.

102. Wen S, Liu H, Li X, Chen X, Hong Y, Li H, et al. TALEN-mediated targeted mutagenesis of fatty acid desaturase 2 (FAD2) in peanut (*Arachis hypogaea* L.) promotes the accumulation of oleic acid. *Plant Mol Biol*. 2018;97:177–85. doi:10.1007/s11103-018-0731-z.
103. United States Department of Agriculture (USDA). 2016. [https://www.aphis.usda.gov/biotechnology/downloads/reg\\_loi/16-090-01\\_air\\_inquiry\\_cbidel.pdf](https://www.aphis.usda.gov/biotechnology/downloads/reg_loi/16-090-01_air_inquiry_cbidel.pdf). Accessed 25 Aug 2018.
104. United States Department of Agriculture (USDA). 2016. [https://www.aphis.usda.gov/biotechnology/downloads/reg\\_loi/16-320-01\\_air\\_inquiry.pdf](https://www.aphis.usda.gov/biotechnology/downloads/reg_loi/16-320-01_air_inquiry.pdf). Accessed 25 Aug 2018.
105. Sawai S, Ohyama K, Yasumoto S, Seki H, Sakuma T, Yamamoto T, et al. Sterol side chain reductase 2 is a key enzyme in the biosynthesis of cholesterol, the common precursor of toxic steroidal glycoalkaloids in potato. *Plant Cell*. 2014;26:3763–74. doi:10.1105/tpc.114.130096.
106. Nakayasu M, Akiyama R, Lee HJ, Osakabe K, Osakabe Y, Watanabe B, et al. Generation of  $\alpha$ -solanine-free hairy roots of potato by CRISPR/Cas9 mediated genome editing of the St16DOX gene. *Plant Physiol Biochem* 2018. doi:10.1016/j.plaphy.2018.04.026.
107. Shan Q, Zhang Y, Chen K, Zhang K, Gao C. Creation of fragrant rice by targeted knockout of the OsBADH2 gene using TALEN technology. *Plant Biotechnol J*. 2015;13:791–800. doi:10.1111/pbi.12312.
108. Sun Y, Jiao G, Liu Z, Zhang X, Li J, Guo X, et al. Generation of High-Amylose Rice through CRISPR/Cas9-Mediated Targeted Mutagenesis of Starch Branching Enzymes. *Front Plant Sci*. 2017;8:1–15. doi:10.3389/fpls.2017.00298.
109. Ye Y, Li P, Xu T, Zeng L, Cheng D, Yang M, et al. OsPT4 Contributes to Arsenate Uptake and Transport in Rice. *Front. Plant Sci*. 2017;8:311. doi:10.3389/fpls.2017.02197.
110. Wang F-Z, Chen M-X, Yu L-J, Xie L-J, Yuan L-B, Qi H, et al. OsARM1, an R2R3 MYB Transcription Factor, Is Involved in Regulation of the Response to Arsenic Stress in Rice. *Front Plant Sci*. 2017;8:1868. doi:10.3389/fpls.2017.01868.
111. Abe K, Araki E, Suzuki Y, Toki S, SAIKA H. Production of high oleic/low linoleic rice by genome editing. *Plant Physiol Biochem* 2018. doi:10.1016/j.plaphy.2018.04.033.
112. Nieves-Cordones M, Mohamed S, Tanoi K, Kobayashi NI, Takagi K, Vernet A, et al. Production of low-Cs<sup>+</sup> rice plants by inactivation of the K<sup>+</sup> transporter OsHAK1 with the CRISPR-Cas system. *Plant J*. 2017;92:43–56. doi:10.1111/tbj.13632.
113. Tang L, Mao B, Li Y, Lv Q, Zhang L, Chen C, et al. Knockout of OsNramp5 using the CRISPR/Cas9 system produces low Cd-accumulating indica rice without compromising yield. *Sci Rep*. 2017;7:14438. doi:10.1038/s41598-017-14832-9.
114. Zhang J, Zhang H, Botella JR, Zhu J-K. Generation of new glutinous rice by CRISPR/Cas9-targeted mutagenesis of the Waxy gene in elite rice varieties. *J Integr Plant Biol*. 2018;60:369–75. doi:10.1111/jipb.12620.
115. Zhou Z, Tan H, Li Q, Chen J, Gao S, Wang Y, et al. CRISPR/Cas9-mediated efficient targeted mutagenesis of RAS in *Salvia miltiorrhiza*. *Phytochemistry*. 2018;148:63–70. doi:10.1016/j.phytochem.2018.01.015.

116. United States Department of Agriculture (USDA). 2014. [https://www.aphis.usda.gov/biotechnology/downloads/reg\\_loi/collectis\\_air\\_fad2k0\\_soy\\_cbidel.pdf](https://www.aphis.usda.gov/biotechnology/downloads/reg_loi/collectis_air_fad2k0_soy_cbidel.pdf). Accessed 25 Aug 2018.
117. United States Department of Agriculture (USDA). 2015. [https://www.aphis.usda.gov/biotechnology/downloads/reg\\_loi/15-071-01air.pdf](https://www.aphis.usda.gov/biotechnology/downloads/reg_loi/15-071-01air.pdf). Accessed 25 Aug 2018.
118. Haun W, Coffman A, Clasen BM, Demorest ZL, Lowy A, Ray E, et al. Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. *Plant Biotechnol J*. 2014;12:934–40. doi:10.1111/pbi.12201.
119. Klap C, Yeshayahou E, Bolger AM, Arazi T, Gupta SK, Shabtai S, et al. Tomato facultative parthenocarpy results from SIAGAMOUS-LIKE 6 loss of function. *Plant Biotechnol J*. 2017;15:634–47. doi:10.1111/pbi.12662.
120. Ueta R, Abe C, Watanabe T, Sugano SS, Ishihara R, Ezura H, et al. Rapid breeding of parthenocarpic tomato plants using CRISPR/Cas9. *Sci Rep*. 2017;7:507. doi:10.1038/s41598-017-00501-4.
121. Lee J, Nonaka S, Takayama M, Ezura H. Utilization of a Genome-Edited Tomato (*Solanum lycopersicum*) with High Gamma Aminobutyric Acid Content in Hybrid Breeding. *J. Agric. Food Chem*. 2018;66:963–71. doi:10.1021/acs.jafc.7b05171.
122. Nonaka S, Arai C, Takayama M, Matsukura C, Ezura H. Efficient increase of  $\gamma$ -aminobutyric acid (GABA) content in tomato fruits by targeted mutagenesis. *Sci Rep*. 2017;7:7057. doi:10.1038/s41598-017-06400-y.
123. Li X, Wang Y, Chen S, Tian H, Fu D, Zhu B, et al. Lycopene Is Enriched in Tomato Fruit by CRISPR/Cas9-Mediated Multiplex Genome Editing. *Front. Plant Sci*. 2018;9:179. doi:10.3389/fpls.2018.00559.
124. Yu Q-h, Wang B, Li N, Tang Y, Yang S, Yang T, et al. CRISPR/Cas9-induced Targeted Mutagenesis and Gene Replacement to Generate Long-shelf Life Tomato Lines. *Sci Rep*. 2017;7:818. doi:10.1038/s41598-017-12262-1.
125. United States Department of Agriculture (USDA). 2017. [https://www.aphis.usda.gov/biotechnology/downloads/reg\\_loi/17-038-01\\_air\\_inquiry\\_cbidel.pdf](https://www.aphis.usda.gov/biotechnology/downloads/reg_loi/17-038-01_air_inquiry_cbidel.pdf). Accessed 25 Aug 2018.
126. Sánchez-León S, Gil-Humanes J, Ozuna CV, Giménez MJ, Sousa C, Voytas DF, Barro F. Low-gluten, nontransgenic wheat engineered with CRISPR/Cas9. *Plant Biotechnol J* 2017. doi:10.1111/pbi.12837.
127. Fister AS, Landherr L, Maximova SN, Gultinan MJ. Transient Expression of CRISPR/Cas9 Machinery Targeting TcNPR3 Enhances Defense Response in *Theobroma cacao*. *Front. Plant Sci*. 2018;9:47. doi:10.3389/fpls.2018.00268.
128. Jia H, Orbovic V, Jones JB, Wang N. Modification of the PthA4 effector binding elements in Type I CsLOB1 promoter using Cas9/sgRNA to produce transgenic Duncan grapefruit alleviating Xcc $\Delta$ pthA4:dCsLOB1.3 infection. *Plant Biotechnol J*. 2016;14:1291–301. doi:10.1111/pbi.12495.
129. Jia H, Zhang Y, Orbović V, Xu J, White FF, Jones JB, Wang N. Genome editing of the disease susceptibility gene CsLOB1 in citrus confers resistance to citrus canker. *Plant Biotechnol J*. 2017;15:817–23. doi:10.1111/pbi.12677.

130. Wang X, Tu M, Wang D, Liu J, Li Y, Li Z, et al. CRISPR/Cas9-mediated efficient targeted mutagenesis in grape in the first generation. *Plant Biotechnol J*. 2018;16:844–55. doi:10.1111/pbi.12832.
131. United States Department of Agriculture (USDA). 2017. [https://www.aphis.usda.gov/biotechnology/downloads/reg\\_loi/17-076-01\\_air\\_inquiry\\_cbidel.pdf](https://www.aphis.usda.gov/biotechnology/downloads/reg_loi/17-076-01_air_inquiry_cbidel.pdf). Accessed 25 Aug 2018.
132. Peng A, Chen S, Lei T, Xu L, He Y, Wu L, et al. Engineering canker-resistant plants through CRISPR/Cas9-targeted editing of the susceptibility gene CsLOB1 promoter in citrus. *Plant Biotechnol J*. 2017;15:1509–19. doi:10.1111/pbi.12733.
133. Wang F, Wang C, Liu P, Lei C, Hao W, GAO Y, et al. Enhanced Rice Blast Resistance by CRISPR/Cas9-Targeted Mutagenesis of the ERF Transcription Factor Gene OsERF922. *PLoS ONE*. 2016;11:e0154027. doi:10.1371/journal.pone.0154027.
134. Zhou J, Peng Z, Long J, Sosso D, Liu B, Eom J-S, et al. Gene targeting by the TAL effector PthXo2 reveals cryptic resistance gene for bacterial blight of rice. *Plant J*. 2015;82:632–43. doi:10.1111/tpj.12838.
135. Blanvillain-Baufumé S, Reschke M, Solé M, Auguy F, Doucoure H, Szurek B, et al. Targeted promoter editing for rice resistance to *Xanthomonas oryzae* pv. *oryzae* reveals differential activities for SWEET14-inducing TAL effectors. *Plant Biotechnol J*. 2017;15:306–17. doi:10.1111/pbi.12613.
136. Li T, Liu B, Spalding MH, Weeks DP, Yang B. High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat Biotechnol*. 2012;30:390–2. doi:10.1038/nbt.2199.
137. Wang J, Tian D, Gu K, Yang X, Wang L, Zeng X, Yin Z. Induction of Xa10-like Genes in Rice Cultivar Nipponbare Confers Disease Resistance to Rice Bacterial Blight. *Mol Plant Microbe Interact*. 2017;30:466–77. doi:10.1094/MPMI-11-16-0229-R.
138. Xie C, Zhang G, Zhang Y, Song X, Guo H, Chen X, Fang R. SRWD1, a novel target gene of DELLA and WRKY proteins, participates in the development and immune response of rice (*Oryza sativa* L.). *Science Bulletin*. 2017;62:1639–48. doi:10.1016/j.scib.2017.12.002.
139. Zhou X, Liao H, Chern M, Yin J, Chen Y, Wang J, et al. Loss of function of a rice TPR-domain RNA-binding protein confers broad-spectrum disease resistance. *Proc Natl Acad Sci U S A*. 2018;115:3174–9. doi:10.1073/pnas.1705927115.
140. United States Department of Agriculture (USDA). 2014. [https://www.aphis.usda.gov/biotechnology/downloads/reg\\_loi/air\\_isu\\_ting\\_rice.pdf](https://www.aphis.usda.gov/biotechnology/downloads/reg_loi/air_isu_ting_rice.pdf). Accessed 25 Aug 2018.
141. Cai L, Cao Y, Xu Z, Ma W, Zakria M, Zou L, et al. A Transcription Activator-Like Effector Tal7 of *Xanthomonas oryzae* pv. *oryzicola* Activates Rice Gene Os09g29100 to Suppress Rice Immunity. *Sci Rep*. 2017;7:5089. doi:10.1038/s41598-017-04800-8.
142. Macovei A, Sevilla NR, Cantos C, Jonson GB, Slamet-Loedin I, Čermák T, et al. Novel alleles of rice eIF4G generated by CRISPR/Cas9-targeted mutagenesis confer resistance to Rice tungro spherical virus. *Plant Biotechnol J*. 2018;47:417. doi:10.1111/pbi.12927.
143. Nekrasov V, Wang C, Win J, Lanz C, Weigel D, Kamoun S. Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. *Sci Rep*. 2017;7:482. doi:10.1038/s41598-017-00578-x.

144. Mahfouz M, Tashkandi M, Ali Z, Aljedaani F, Shami A. Engineering resistance against Tomato yellow leaf curl virus via the CRISPR/Cas9 system in tomato. *Plant Signal Behav.* 2018;13(10):e1525996. doi:10.1101/237735.
145. Toledo Thomazella DP de, Brail Q, Dahlbeck D, Staskawicz BJ. CRISPR-Cas9 mediated mutagenesis of a DMR6 ortholog in tomato confers broad-spectrum disease resistance. 2016:1–23. doi:10.1101/064824.
146. Zhang Y, Bai Y, Wu G, Zou S, Chen Y, Gao C, Tang D. Simultaneous modification of three homoeologs of TaEDR1 by genome editing enhances powdery mildew resistance in wheat. *Plant J.* 2017;91:714–24. doi:10.1111/tpj.13599.
147. United States Department of Agriculture (USDA). 2015. [https://www.aphis.usda.gov/biotechnology/downloads/reg\\_loi/15-238-01\\_air\\_inquiry\\_cbidel.pdf](https://www.aphis.usda.gov/biotechnology/downloads/reg_loi/15-238-01_air_inquiry_cbidel.pdf). Accessed 25 Aug 2018.
148. Ruitter R, van den Brande I, Stals E, Delauré S, Cornelissen M, D'Halluin K. Spontaneous mutation frequency in plants obscures the effect of chimeraplasty. *Plant Mol Biol.* 2003;53:675–89. doi:10.1023/b:plan.0000019111.96107.01.
149. Hummel AW, Chauhan RD, Cermak T, Mutka AM, Vijayaraghavan A, Boyher A, et al. Allele exchange at the EPSPS locus confers glyphosate tolerance in cassava. *Plant Biotechnol J.* 2018;16:1275–82. doi:10.1111/pbi.12868.
150. D'Halluin K, Vanderstraeten C, van Hulle J, Rosolowska J, van den Brande I, Pennewaert A, et al. Targeted molecular trait stacking in cotton through targeted double-strand break induction. *Plant Biotechnol J.* 2013;11:933–41. doi:10.1111/pbi.12085.
151. Sauer NJ, Narváez-Vásquez J, Mozoruk J, Miller RB, Warburg ZJ, Woodward MJ, et al. Oligonucleotide-Mediated Genome Editing Provides Precision and Function to Engineered Nucleases and Antibiotics in Plants. *Plant Physiol.* 2016;170:1917–28. doi:10.1104/pp.15.01696.
152. Ainley WM, Sastry-Dent L, Welter ME, Murray MG, Zeitler B, Amora R, et al. Trait stacking via targeted genome editing. *Plant Biotechnol J.* 2013;11:1126–34. doi:10.1111/pbi.12107.
153. Zhu T, Peterson DJ, Tagliani L, St. Clair G, Baszczynski CL, Bowen B. Targeted manipulation of maize genes in vivo using chimeric RNA/DNA oligonucleotides. *Proc. Natl. Acad. Sci.* 1999;96:8768–73. doi:10.1073/pnas.96.15.8768.
154. Zhu T, Mettenburg K, Peterson DJ, Tagliani L, Baszczynski CL. Engineering herbicide-resistant maize using chimeric RNA/DNA oligonucleotides. *Nat Biotechnol.* 2000;18:555–8. doi:10.1038/75435.
155. Butler NM, Baltés NJ, Voytas DF, Douches DS. Geminivirus-Mediated Genome Editing in Potato (*Solanum tuberosum* L.) Using Sequence-Specific Nucleases. *Front Plant Sci.* 2016;7:1–13. doi:10.3389/fpls.2016.01045.
156. Li J, Meng X, Zong Y, Chen K, Zhang H, Liu J, et al. Gene replacements and insertions in rice by intron targeting using CRISPR-Cas9. *NPLANTS.* 2016;2:1–6. doi:10.1038/nplants.2016.139.
157. Sun Y, Zhang X, Wu C, He Y, Ma Y, Hou H, et al. Engineering Herbicide-Resistant Rice Plants through CRISPR/Cas9-Mediated Homologous Recombination of Acetolactate Synthase. *Mol Plant.* 2016;9:628–31. doi:10.1016/j.molp.2016.01.001.

158. Wang M, Liu Y, Zhang C, Liu J, Liu X, Wang L, et al. Gene editing by co-transformation of TALEN and chimeric RNA/DNA oligonucleotides on the rice OsEPSPS gene and the inheritance of mutations. *PLoS ONE*. 2015;10:e0122755. doi:10.1371/journal.pone.0122755.
159. Okuzaki A, Toriyama K. Chimeric RNA/DNA oligonucleotide-directed gene targeting in rice. *Plant Cell Rep*. 2004;22:509–12. doi:10.1007/s00299-003-0698-2.
160. Shimatani Z, Kashojiya S, Takayama M, Terada R, Arazoe T, Ishii H, et al. Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat Biotechnol*. 2017;35:441–3. doi:10.1038/nbt.3833.
161. Shimatani Z, Fujikura U, Ishii H, Matsui Y, Suzuki M, Ueke Y, et al. Inheritance of co-edited genes by CRISPR-based targeted nucleotide substitutions in rice. *Plant Physiol Biochem* 2018. doi:10.1016/j.plaphy.2018.04.028.
162. Butt H, Eid A, Ali Z, Atia MAM, Mokhtar MM, Hassan N, et al. Efficient CRISPR/Cas9-Mediated Genome Editing Using a Chimeric Single-Guide RNA Molecule. *Front Plant Sci*. 2017;8:1441. doi:10.3389/fpls.2017.01441.
163. Li Z, Liu Z-B, Xing A, Moon BP, Koellhoffer JP, Huang L, et al. Cas9-Guide RNA Directed Genome Editing in Soybean. *Plant Physiol*. 2015;169:960–70. doi:10.1104/pp.15.00783.
164. Chilcoat D, Liu Z-B, Sander J. Use of CRISPR/Cas9 for Crop Improvement in Maize and Soybean. *Prog Mol Biol Transl Sci*. 2017;149:27–46. doi:10.1016/bs.pmbts.2017.04.005.
165. United States Department of Agriculture (USDA). 2018. [https://www.aphis.usda.gov/biotechnology/downloads/reg\\_loi/18-036-01\\_a1\\_air\\_inquiry\\_cbid1.pdf](https://www.aphis.usda.gov/biotechnology/downloads/reg_loi/18-036-01_a1_air_inquiry_cbid1.pdf). Accessed 25 Aug 2018.
166. Zhou X, Jacobs TB, Xue L-J, Harding SA, Tsai C-J. Exploiting SNPs for biallelic CRISPR mutations in the outcrossing woody perennial *Populus* reveals 4-coumarate:CoA ligase specificity and redundancy. *New Phytol*. 2015;208:298–301. doi:10.1111/nph.13470.
167. Andersson M, Turesson H, Nicolia A, Fält A-S, Samuelsson M, Hofvander P. Efficient targeted multiallelic mutagenesis in tetraploid potato (*Solanum tuberosum*) by transient CRISPR-Cas9 expression in protoplasts. *Plant Cell Rep*. 2017;36:117–28. doi:10.1007/s00299-016-2062-3.
168. Jung JH, Altpeter F. TALEN mediated targeted mutagenesis of the caffeic acid O-methyltransferase in highly polyploid sugarcane improves cell wall composition for production of bioethanol. *Plant Mol Biol*. 2016;92:131–42. doi:10.1007/s11103-016-0499-y.
169. Kannan B, Jung JH, Moxley GW, Lee S-M, Altpeter F. TALEN-mediated targeted mutagenesis of more than 100 COMT copies/alleles in highly polyploid sugarcane improves saccharification efficiency without compromising biomass yield. *Plant Biotechnol J*. 2018;16:856–66. doi:10.1111/pbi.12833.
170. Park J-J, Yoo CG, Flanagan A, Pu Y, Debnath S, Ge Y, et al. Defined tetra-allelic gene disruption of the 4-coumarate:coenzyme A ligase 1 (Pv4CL1) gene by CRISPR/Cas9 in switchgrass results in lignin reduction and improved sugar release. *Biotechnol Biofuels*. 2017;10:284. doi:10.1186/s13068-017-0972-0.

171. United States Department of Agriculture (USDA). 2011. [https://www.aphis.usda.gov/biotechnology/downloads/reg\\_loi/17-126-01\\_air\\_inquiry.pdf](https://www.aphis.usda.gov/biotechnology/downloads/reg_loi/17-126-01_air_inquiry.pdf). Accessed 25 Aug 2018.
172. Njuguna E, Coussens G, Aesaert S, Neyt P, Anami S, van Lijsebettens M. Modulation of energy homeostasis in maize and Arabidopsis to develop lines tolerant to drought, genotoxic and oxidative stresses. *AF* 2018. doi:10.21825/af.v30i2.8080.
173. United States Department of Agriculture (USDA). 2017. [https://www.aphis.usda.gov/biotechnology/downloads/reg\\_loi/17-219-01\\_air\\_inquiry.pdf](https://www.aphis.usda.gov/biotechnology/downloads/reg_loi/17-219-01_air_inquiry.pdf). Accessed 25 Aug 2018.
174. Kim D, Alptekin B, Budak H. CRISPR/Cas9 genome editing in wheat. *Funct Integr Genomics*. 2018;18:31–41. doi:10.1007/s10142-017-0572-x.
175. Meng X, Hu X, Liu Q, Song X, Gao C, Li J, Wang K. Robust genome editing of CRISPR-Cas9 at NAG PAMs in rice. *Sci China Life Sci*. 2018;61:122–5. doi:10.1007/s11427-017-9247-9.
176. Ali Z, Abul-faraj A, Piatek M, Mahfouz MM. Activity and specificity of TRV-mediated gene editing in plants. *Plant Signal Behav*. 2015;10:e1044191. doi:10.1080/15592324.2015.1044191.
177. Shen C, Que Z, Xia Y, Tang N, Li D, He R, Cao M. Knock out of the annexin gene *OsAnn3* via CRISPR/Cas9-mediated genome editing decreased cold tolerance in rice. *J. Plant Biol*. 2017;60:539–47. doi:10.1007/s12374-016-0400-1.
178. Tian S, Jiang L, Gao Q, Zhang J, Zong M, Zhang H, et al. Efficient CRISPR/Cas9-based gene knockout in watermelon. *Plant Cell Rep*. 2017;36:399–406. doi:10.1007/s00299-016-2089-5.
179. Zhang H, Zhang J, Wei P, Zhang B, Gou F, Feng Z, et al. The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnol J*. 2014;12:797–807. doi:10.1111/pbi.12200.
180. Forner J, Pfeiffer A, Langenecker T, Manavella PA, Manavella P, Lohmann JU. Germline-transmitted genome editing in *Arabidopsis thaliana* Using TAL-effector-nucleases. *PLoS ONE*. 2015;10:e0121056. doi:10.1371/journal.pone.0121056.
181. Ren B, Yan F, Kuang Y, Li N, Zhang D, Lin H, Zhou H. Specificity and inheritance of rBE3 and rBE4 endonuclease-induced gene modifications in rice. *J. Phys. A: Math. Theor*. 2017.
182. Khandagale K, Nadaf A. Genome editing for targeted improvement of plants. *Plant Biotechnol Rep*. 2016;10:327–43. doi:10.1007/s11816-016-0417-4.

## 6 Systematic review

---

### Which factors affect the occurrence of off-target effects caused by the use of CRISPR/Cas: A systematic review in plants

---

Chapter 6 has been published:

**Modrzejewski Dominik**, Hartung, Frank; Lehnert, Heike; Sprink, Thorben; Kohl, Christian; Keilwagen, Jens, Wilhelm, Ralf (2020): Which Factors Affect the Occurrence of Off-Target Effects Caused by the Use of CRISPR/Cas: A Systematic Review in Plants. *Front. Plant Sci.* 11:574959. DOI: 10.3389/fpls.2020.574959.

<sup>1</sup>Institute for Biosafety in Plant Biotechnology, Federal Research Centre for Cultivated Plants, Julius Kühn-Institut, Erwin-Baur-Straße 27, 06484 Quedlinburg, Germany

#### Authors' contributions

The review process was coordinated by **DM**. **DM**, FH and TS conducted the screening of articles. **DM** extracted metadata. HL carried out statistical analysis. All authors participated in the critical appraisal of studies. **DM** drafted the report. RW, FH, TS and CK assisted in editing and revising the final manuscript.

## 6.1 Abstract

CRISPR/Cas enables a targeted modification of DNA sequences. Despite their ease and efficient use, one limitation is the potential occurrence of associated off-target effects. This systematic review aims to answer the following research question: Which factors affect the occurrence of off-target effects caused by the use of CRISPR/Cas in plants? Literature published until March 2019 was considered for this review. Articles were screened for relevance based on pre-defined inclusion criteria. Relevant studies were subject to critical appraisal. All studies included in the systematic review were synthesized in a narrative report, but studies rated as *high* and *medium/high validity* were reported separately from studies rated as *low* and *medium/low* or *unclear validity*. In addition, we ran a binary logistic regression analysis to verify five factors that may affect the occurrence of off-target effects: 1. Number of mismatches 2. Position of mismatches 3. GC-content of the targeting sequence 4. Altered nuclease variants 5. Delivery methods. In total, 180 relevant articles were included in this review containing 468 studies therein. 79% of these studies were rated as having *high* or *medium/high validity*. Within these studies, 6416 potential off-target sequences were assessed for the occurrence of off-target effects. Results clearly indicate that an increased number of mismatches between the on-target and potential off-target sequence steeply decreases the likelihood of off-target effects. The observed rate of off-target effects decreased from 59% when there is one mismatch between the on-target and off-target sequences towards 0% when four or more mismatches exist. In addition, mismatch/es located within the first eight nucleotides proximal to the PAM significantly decreased the occurrence of off-target effects. There is no evidence that the GC-content significantly affects off-target effects. The database regarding the impact of the nuclease variant and the delivery method is very poor as the majority of studies applied the standard nuclease SpCas9 and the CRISPR/Cas system was stably delivered in the genome. Hence, a general significant impact of these two factors on the occurrence of off-target effects cannot be proved. This identified evidence gap needs to be filled by systematic studies exploring these individual factors in sufficient numbers.

## 6.2 Background

Site-directed nucleases (SDN), including Clustered Regularly Interspaced Short Palindromic Repeats with associated protein (e.g. CRISPR/Cas9), Transcription Activator-Like Effector Nucleases (TALENs), Zinc-Finger Nucleases (ZFN) and Meganucleases (MN) enable a modification of a specific DNA sequences in a site-directed manner. Compared to other SDN, CRISPR/Cas is the most widely applied method due to simplicity, accessibility, lower costs and versatility as well as the possibility of multiplexing (Dönmez et al., 2016). Most CRISPR/Cas systems require two elements for targeted genome cleavage: An endonuclease and a single guide RNA (sgRNA) which recognizes and binds to a specific DNA sequence that is typically 20 base pairs long for SpCas9 (Sprink et al., 2016). The to date most commonly used endonuclease Cas9 derived from *Streptococcus pyogenes* is able to bind to any DNA sequence paired with an sgRNA followed by a protospacer adjacent motif (PAM) site (NGG for *S.pyogenes*) and introduces a double strand break (DSB) three bases upstream of the PAM (Zhu et al., 2017). The DSB is subsequently repaired by the DNA repair system of the cell. CRISPR/Cas has already been applied all over the world in more than 60 plant species ranging from model organisms to agricultural crops

(Modrzejewski et al., 2019). Despite the obvious potentials of CRISPR/Cas in plant breeding, there is an ongoing debate about its precise targeting and to what extent the occurrence of off-target effects matters (Eckerstorfer et al., 2019b, Zhao, Wolt, 2017). In order to evaluate potential side effects and to further enhance the specificity of CRISPR/Cas system, a detailed evaluation on the occurrence of off-target effects is important (Martin et al., 2016). Off-target effects can be defined as unintended cleavage and mutations at untargeted genomic sites showing a similar but not an identical sequence compared to the target site (Modrzejewski et al., 2019). It is not exactly known why the Cas9 protein cleaves some off-target sites and others not. One important factor affecting cleavage of Cas9 could be chromatin structure as it has been shown that Cas9 cleaves more efficient in open chromatin regions (Hinz et al., 2015, Liu et al., 2019). In addition, a number of biophysical experiments highlighted that not only PAM proximal but also PAM distal mismatches can strongly inhibit off-target cleavage. The proposed mechanism for this intriguing phenomenon is a balance shift in the HNH domain towards an inactive state (Chen et al., 2017, Dagdas et al., 2017, Mitchell et al., 2020, Ricci et al., 2019). After binding of the target DNA and establishment of the RNA:DNA hybrid at the on-target, the HNH domain flips from an RNA-bound state to an inactive state and then rapidly into a docked-state that enables cleavage (Dagdas et al., 2017). PAM distal mismatches between the sgRNA and the DNA (especially position 16-20) can slow down or even inhibit this transition from inactive to docked-state, thereby prohibiting off-target cleavage (Chen et al., 2017, Dagdas et al., 2017). This biophysical results lead to the conclusion, that not only RNA:DNA binding affects the occurrence of off-target cleavage but also the internal reorganization of the HNH domain. The first effect seems to be located PAM proximal as we could show also in this review by analysis of a number of experimentally validated off-target effects. The second effect is located PAM distal and does not rely on RNA:DNA binding. In contrast to this, mismatches in the more upstream positions 15-10 do almost not affect the cleavage negatively, as they neither have a remarkable effect on RNA:DNA duplex stability nor on the HNH conformational changes (Chen et al., 2017).

Two different strategies have been used to detect off-target effects: biased and unbiased detection methods (Martin et al., 2016). To date, the majority of studies analyzed off-target effects using biased detection methods (Modrzejewski et al., 2019). This approach consists of two steps. First, the potential off-target sequences that are similar to the target sequence are identified using bioinformatics approaches like CAS-OFF-Finder (Altschul, 1990), CRISPR-P (Lei et al., 2014, Liu et al., 2017), CHOPCHOP (Montague et al., 2014) or CCTop (Stemmer et al., 2015). Second, the identified potential off-target sequences only are analyzed for undesired mutations (off-target effects). Several different detection methods are used, whereby off-target effects were mostly analyzed using PCR followed by sequencing (Modrzejewski et al., 2019). Nevertheless, a drawback of solely screening pre-selected potential off-target sequences is the risk to overlook mutations at further loci in the plant genome (Zischewski et al., 2017). In contrast, using the unbiased approach whole genome sequencing (WGS) is the most common detection method in plants allowing the identification of off-target effects in a less restricted way (Martin et al., 2016, Modrzejewski et al., 2019). However, due to the high costs today, solely a few clones are sequenced for each target in most instances, and hence low-frequency off-target effects might be missed because of low coverage (read depth) (Wu et al., 2014). Mutations detected by using unbiased detection methods do not necessarily originate from the application of a genome editing technique, but may also result from spontaneous mutations or through the regeneration of plants from cell

culture (somaclonal variation) (Scientific Advice Mechanism, 2017). Therefore, a suitable reference genome is needed to detect genetic differences. In order to trace these differences to a genome editing alteration, bioinformatics and statistical analyses are necessary to estimate whether these differences are considerably likely to be technology-induced genetic modifications (Bartsch et al., 2018). Compared to routinely used breeding techniques like undirected mutagenesis (e.g. by irradiation or chemicals) or somaclonal variation, CRISPR/Cas causes far less unintended changes (Jander et al., 2003, Miyao et al., 2012, Scientific Advice Mechanism, 2017). Nevertheless, a recently published systematic map on off-target effects in CRISPR/Cas-studies identified several publications, which reported the identification of off-target effects (Modrzejewski et al., 2019). It is estimated that monomeric CRISPR/Cas system is more prone for the occurrence of off-target effects compared to the dimeric ZFN or TALENs systems that use two neighboring target sequences and hence recognize a longer target sequence (Lee et al., 2016, Zischewski et al., 2017). In addition, the sgRNA can tolerate some mismatches and bulges and hence efforts have been made to increase the on-target efficiency and to decrease the occurrence of off-target effects.

Several factors are described in literature that may affect off-target effects. The main factor to decrease the occurrence of off-target effects is a careful selection of the target sequence (Hsu et al., 2013, Zhu et al., 2017, Zischewski et al., 2017). Design tools like CRISPR-P or CHOPCHOP provide the possibility to take potential off-target sequences into account when choosing the target sequence (Zhao, Wolt, 2017). It was considered that mismatches occurring within the seed sequence (8 up to 12 nucleotides proximal to the PAM) determine the editing efficiency. Therefore, mismatches within this region may reduce off-target effects (Endo et al., 2015, Hahn, Nekrasov, 2018, Hsu et al., 2013). However, it is not clear how many nucleotides proximal to the PAM are building up the seed sequence. Different researchers define the seed sequence to be eight (Russo et al., 2018), nine (Bertier et al., 2018), ten (Jiang et al., 2017), eleven (Jacobs et al., 2015) or up to twelve (LeBlanc et al., 2018) nucleotides proximal to the PAM.

Scientists also assume that the guanine-cytosine (GC) content of a chosen target sequence may influence off-target effects, as high GC-content stabilize the sRNA/genomic DNA hybridization (Fu et al., 2013). Yu et al. (2017) speculate that a low GC-content results in an reduced number of off-target effects (Yu et al., 2017). Russo et al. (2018) have chosen a target sequence with a GC-content of 50% as they assume that a GC-content higher than 70% may increase the risk of off-targeting (Russo et al., 2018).

Much effort is invested in examining other nucleases and developing improved ones. Next to the most widely used nuclease SpCas9 derived from *Streptococcus pyogenes* further Cas9 proteins from different bacterial or archaea species have been adapted. One example is the Cas9 derived from *Staphylococcus aureus* (SaCas9) that is capable to use a longer recognition sequence of 21- or 22- nucleotides and a different PAM (NNGRRT). Therefore fewer off-target sequences are predicted per se and its specificity is estimated to be higher (Kaya et al., 2016). In addition, the Cpf1 nuclease (also known as Cas12a) has been applied for targeted genome modification in plants. Compared to Cas9, the Cpf1 (from *Prevotella spec.* and *Francisella spec.*) recognizes a T-rich PAM (TTTV) at the 5' end instead of the 3' end of the protospacer and has the potential to decrease off-target effects due to their DNA recognition and cutting properties (Hahn, Nekrasov, 2018). In addition, protein-engineering approaches resulted in altered SpCas9 variants with potentially reduced off-target activity. Several SpCas9-variants like SpCas9-HF, eSpCas9 (1.0) or eSpCas9

(1.1) have been engineered and were already applied in plants (Raitskin et al., 2019, Zhang et al., 2017). These nuclease variants were described to nearly entirely avoid off-targeting (Tycko et al., 2016). Another promising approach is the application of CRISPR/nickase systems. In contrast to nucleases which induce a DSB a nickase system can be used to induce single strand breaks or as a paired nickase system to induce offset DSBs (Puchta, Fauser, 2014, Zhao, Wolt, 2017). This may reduce off-target effects as the recognition site is doubled from 20 to 40 nucleotides (Puchta, Fauser, 2014).

A further aspect that may have an impact on off-target effects is the delivery of the CRISPR/Cas system into the plant cell. The system can be supplied to the plant genome either as DNA (transiently or stably expressed), RNA or directly as Ribonucleoproteins (RNP) (Jansing et al., 2019). It is supposed that the occurrence of off-target effects depends on how long the CRISPR/Cas system is active in the plant cell (Jansing et al., 2019). Stable transformation leads to a permanent expression of the CRISPR/Cas system compared to the transient approach in which the CRISPR/Cas system is available only for a limited time. Therefore, it is supposed that a stable transformation leads to an increased on-target as well as an increased off-target activity (Jansing et al., 2019, Metje-Sprink et al., 2018, Zischewski et al., 2017). Supplying the CRISPR/Cas system as RNA or RNP may further reduce off-target effects, as it is degraded in the shortest period and the mode of action is only present in the edited cells but not in the regenerated plants (Jansing et al., 2019, Metje-Sprink et al., 2018, Woo et al., 2015). All delivery methods have their specific advantages and disadvantages and are more or less suitable for different plant species (Jansing et al., 2019).

All in all a broad range of factors may affect the occurrence of off-target effects. Reviewing all available literature for plants, to date no systematical analyzes, which factors actually affect the occurrence of off-target effects due to the application of CRISPR/Cas in plants were found.

### **Topic identification**

Risk assessors and decision makers are depending on the provision of a reliable body of evidence to support conclusions about potential risks being associated with the application of genome editing. In this context, the (potential) off-target effects caused by genome editing in contrast to its broadly claimed precision are a point of lasting criticism as they might lead to adverse alterations in plants (Eckerstorfer et al., 2019a). Additionally, the detailed analysis of the occurrence of off-target effects can support further enhancement of the specificity of CRISPR/Cas (Martin et al., 2016). This systematic review builds on the recently published systematic map on genome editing applications in plants (Modrzejewski et al., 2019). The systematic map and the a priori published systematic map protocol were conducted based on the guideline of Collaboration for Environmental Evidence (CEE) aiming to take the reader through the key stages of the review (Collaboration for Environmental Evidence, 2018). One topic of this map (secondary question 2) was the identification of the available evidence for the occurrence of such off-target effects. The map identified a knowledge cluster of publications considering the evaluation of off-target effects caused by CRISPR/Cas in plants (other genome editing tools were much less represented), supporting the conduct of an in depth analysis by a systematic review on this specific section of the map. In the reviewed studies, factors potentially modifying the occurrence of off-target effects such as the plant species, the nuclease-variant (e.g. Cas9, Cpf1(Cas12a)), the number of mismatches and the chosen sgRNA were mapped (Modrzejewski et al., 2019). Based on these data,

this systematic review identifies and systematically analyzes factors that may affect the occurrence of off-target effects caused by CRISPR/Cas in plants.

### 6.3 Objectives of the review

The primary objective of this systematic review is to collect and synthesize the available evidence about factors that do affect the occurrence of off-target effects caused by the application of CRISPR/Cas in plants.

Primary question: “Which factors affect the occurrence of off-target effects caused by the use of CRISPR/Cas in plants”?

#### Components of the primary question

Population: Any model plant or crop produced for agricultural production.

Intervention: The CRISPR/Cas technique was used to induce any on-target mutation.

Outcome: The occurrence of off-target effects was assessed. Either biased or unbiased detection methods were used to check whether off-target effects occurred in the plant genome.

### 6.4 Methods

The methods used to conduct this systematic review are based on the recently published systematic map (Modrzejewski et al., 2019), but adapted to the specific requirements of the systematic review question. The methods specifically used to conduct this systematic review as well as the deviations compared to the systematic map protocol are described below.

#### Search for articles

The CRISPR/Cas articles analyzed to answer secondary question 2 of the systematic map about the occurrence of off-target effects were included in this systematic review, covering the time period between 1996 and May 2018 (a list is provided in additional file 6 of the published systematic map (Modrzejewski et al., 2019)). In addition, the literature search was updated in March 2019, to identify all CRISPR/Cas articles published between May 2018 and March 2019 (Figure 6.1). The search string was similar to the search string of the systematic map but focused on CRISPR/Cas (the question addressed in the systematic map was broader in scope and included besides to CRISPR/Cas also other genome editing techniques). It comprises two parts: The first part defined the population of interest comprising less specific terms like crop, plant or seed as well as specific model plants and crops including their English and Latin names. The second part defined the intervention, comprising different CRISPR/Cas variants. The search terms within each class were combined using the Boolean operator “OR”; the two classes were combined using the Boolean operator “AND”. The final search string for Web of Science is provided in supplementary materials- table 1. The search string was adapted to the specific needs of each database.

The following online publication databases and platforms were searched for academic literature. Access was ensured by institutional subscriptions:

- (i) Web of Science (WoS), including Web of Science Core Collection, BIOSIS Citation Index, BIOSIS Preview, CABI:CAB Abstracts and Global Health, Data Citation Index, Derwent Innovations Index, KCI-Korean Journal Database, Russian Science Citation Index, SciELO Citation index, Zoological Records
- (ii) Scopus
- (iii) PubMed
- (iv) Science direct
- (v) AGRIS

Furthermore, Google Scholar (<https://scholar.google.com/>) was searched using 30 different combinations of the most relevant (model) plants and CRISPR/Cas terms. The first 20 search results, organized by relevance, of each combined search string were assessed at the title/abstract level. Deviating from the systematic map protocol, no search for grey literature on company webpages was conducted, because we noticed that they do not provided any data about the occurrence of off-target effects (Modrzejewski et al., 2019). All hits from each database were imported into an EndNote X8.0.1 library file. Duplicates were removed using the appropriate function within the EndNote software. Afterwards, the remaining records were imported into the open-access and non-profit database CADIMA to increase transparency and traceability during the review process (Kohl et al., 2018).

## **Article screening and study inclusion criteria**

### **Screening process**

When conducting the systematic map a consistency check was performed aiming to assure for a good inter-reviewer agreement. As the review team stayed the same and the eligibility criteria were nearly identical compared to the systematic map (except that only CRISPR/Cas articles were an eligible intervention and not of all genome editing techniques) we did not carry out a further consistency check/ kappa test within this review. Two reviewers conducted title/abstract screening independently. Two reviewers then checked the potentially relevant articles at full text level. Lists of unobtainable articles and articles excluded at full text level with the reason for exclusion are provided (supplementary materials- table 2). All members of the review team are authors of a few articles retrieved by the literature search. However, as none of their articles provides primary data, these articles were excluded at title/abstract level.

### **Eligibility criteria**

In order to be included in the systematic review each article had to meet the following inclusion criteria:

Eligible population: Any model plant or crop produced for agricultural production.

Eligible intervention: The CRISPR/Cas technique was used to induce any on-target mutation.

Eligible outcome: The occurrence of off-target effects was assessed using either biased or unbiased detection methods. In case no potential off-target sequences were identified using prediction tools

and thus no potential off-target sequences were tested for the occurrence of off-target effects, the study was excluded.

Eligible type of data: Only those references were included which comprise primary data referring to the use of a CRISPR/Cas and the detection of potential off-target effects.

Eligible languages: References in German and English language were included. Articles in other languages were included when besides title and abstract, further parts of the article, like figures or tables, were in English or German and the provided information allowed for a definite judgement of their relevance.

### Data coding strategy

One article can contain several studies e.g. when different plant species were investigated or different sgRNA, nucleases or delivery methods were applied. While articles were screened for relevance at title/abstract and full text level, relevant data were finally extracted at study level. Due to limited available resources, one reviewer extracted the relevant data. Unclear cases were marked by the first reviewer and cross-checked by another one. Building up on the data extracted within the systematic map, further details were extracted for the systematic review as shown in Table 6.1. The following data of each included study were extracted into one row in an excel sheet (each in a cell):

**Table 6.1** Overview about the extracted data within the systematic map and further data that were additionally extracted for the systematic review.

|                                  | Data extracted within the systematic map  | Further data additionally extracted for the systematic review                           |
|----------------------------------|---|---|
| <b>Bibliographic information</b> | - Abstract,<br>- Authors,<br>- DOI/ISBN<br>- Issue number,<br>- Keywords,<br>- Page range,<br>- Periodical,<br>- Title,<br>- Volume,<br>- Year, | None  |
| <b>Population</b>                | Plant species   | None  |
| <b>Intervention</b>              | Genome editing technique (e.g. CRISPR/Cas)  | - Specification of the nuclease variant (e.g. SpCas9, SaCas9, Cpf1)                     |
| <b>Transformation</b>            | None  | - Delivery method (DNA-stable, DNA-transient, RNA, Ribonucleoproteins (RNP))            |
| <b>Protospacer</b>               | None  | Guanin-cytosine (GC)-content of the target sequence                                     |
| <b>Target sequence</b>           | Sequence identifier (Name of the gene)  | - Different sgRNAs targeting one gene were extracted separately (each sgRNA in one row) |
| <b>Off-target prediction</b>     | - Number of identified potential off-target sequences<br>- Prediction tool  | None  |
| <b>Off-target detection</b>      | - Detection method<br>- Number of analyzed off-target sequences   | None<br>- Number of mismatches<br>- Position of mismatches in relation to the PAM       |

|                                      | Data extracted within the systematic map | Further data additionally extracted for the systematic review  |
|--------------------------------------|--|--|
| <b>Identified off-target effects</b> | Off-targets identified (Yes/ No)         | - Number of identified off-targets,<br>- Number of mismatches<br>- Position of mismatches in relation to the PAM |

In case relevant data were missing or not accessible, the corresponding author of the article was asked to provide the missing information within a period of two weeks. The extracted data are available in supplementary materials- table 3 (summary) and supplementary materials- table 4 (details of all potential off-target sequences).

### Study validity assessment

Articles that passed the study inclusion criteria were subjected to critical appraisal. The critical appraisal was conducted on a study-by-study basis. Criteria were defined within the review team in order to assess internal validity (risk of bias within a specific study) and external validity (generalizability). These criteria reflect critical variables that affect the reliability of study outcomes. The critical appraisal criteria are listed in Table 6.2. Each study was assessed by considering its compliance with each of the developed criteria (yes, partly, no, unclear). Based on these judgments, studies were categorized as having *high*, *medium*, *low* or *unclear validity*. In case that, due to the fact that more than one potential off-target sequences with diverging quality were included in one study, a study was considered as being of *medium validity* in a first step. In a next step the medium category was further subdivided into medium/high or medium/low depending on the characteristics of the individual sequence (for more detail also see Table 6.3). This may occur in case within one study (i) several potential off-target sequences were analyzed but only some of them were followed by a PAM, (ii) information were solely provided for a subset of analyzed potential off-target sequences or (iii) a subset of analyzed potential off-target sequences were identical to the on-target site. All potential off-target sequences within one medium-study that fulfilled the required criteria were moved to the medium/high category, while the remaining ones were rated as *medium/low validity* according to Table 6.2 and Table 6.3. All studies were included in the narrative report, but based on the overall ranking of the critical appraisal, the *high* and *medium/high validity* categories were handled separately from the *low* and *medium/low* as well as the *unclear validity* studies. Due to limited available resources, solely one reviewer assessed the validity of all studies. Doubtful cases were discussed within the whole review team. The results of the critical appraisal are detailed in supplementary materials- table 5.

**Table 6.2** Study validity assessment criteria.

| Question/ criterion   | Yes                                | Partly | No   | Unclear                                 |
|---|------------------------------------|--------|--|---|
| <b>Study design/ methods used</b>                           |                                    |        |  |   |
| Did the CRISPR/Cas technique induce any on-target mutation? | Any on-target mutation was induced | -      | No on-target mutation was induced  | Lacking sufficient information to judge |
| Is a reference genome of the edited plant available?        | Reference genome is available      | -      | Reference genome has not been fully sequenced/ Sequencing error when analyzing | Lacking sufficient information to judge |

|  |   |  | potential off-target sequences                           |   |
|--|---|--|--|---|
| <b>Potential off-target sequences</b>  |   |  |  |   |
| Did a PAM follow the potential off-target sequences?*  | All potential off-target sequences were followed by a PAM     | Some potential off-target sequences were followed by a PAM     | No potential off-target sequences were followed by a PAM | Lacking sufficient information to judge |
| Has the potential off-target sequence at least one mismatch to the on-target site  | All potential off-target sequences have at least one mismatch | Some potential off-target sequences have at least one mismatch | All potential off-target sequences have no mismatch      | Lacking sufficient information to judge |
| Are information about the potential off-target sequences provided? (number of mismatches, the position of mismatches to the PAM) | All information are provided                                  | Information are partly provided                                | No information are provided                              | -                                       |

One reviewer answered the questions with „Yes“, „Partly“, „No“ or „unclear“. The questions were answered based on the specification in the table. Questions answered with “Yes” means that the study has a *high validity* (low risk of bias) regarding this criteria, while “No” means that the study has a *low validity* (high risk of bias). Questions answered with “Partly” means that a subset of analyzed potential off-target sequences within this study has a *high validity* and a subset has a *low validity* regarding this criteria. \* CRISPR/SpCas9 and engineered SpCas9 nuclease variants: NGG-PAM or NAG-PAM or NGA-PAM (VQR-variants): CRISPR/Cpf1: TTN-PAM or TTTN-PAM; CRISPR/SaCas9 and CRISPR/eSaCas9: NNGRRT-PAM;

**Table 6.3** Overall assessment of study validity.

| <b>Studies were assigned <i>low validity</i> or <i>unclear validity</i> if any of the following factors applied</b>                               |
|---|
| <b>Any of these criteria answered with “No” or “Unclear”</b>  |
| Did the CRISPR/Cas technique induce any on-target mutation?   |
| Did a PAM follow the potential off-target sequences?  |
| Has the potential off-target sequence at least one mismatch to the on-target site?  |
| <b>Any of these criteria answered with “No”</b>   |
| Is a reference genome of the edited plant available?  |
| Did sequencing errors occur?  |
| Are information about the potential off-target sequences provided? (number of mismatches, the position of mismatches to the PAM)                  |
| <b>Studies that were not assigned <i>low validity</i> were considered to have <i>medium validity</i> if any of the following factors applied.</b> |
| <b>Any of the criteria answered with “Partly”</b>   |
| Did a PAM follow the potential off-target sequences?  |
| Has the potential off-target sequence at least one mismatch to the on-target site   |
| Are information about the potential off-target sequences provided? (number of mismatches, the position of mismatches to the PAM)                  |

---

**Potential off-target sequences within *medium validity* studies were handled as *medium/low validity* if any of the following factors was not applied.**

---

**Any of the criteria answered with “No”**

Did a PAM follow the potential off-target sequences?

Has the potential off-target sequence at least one mismatch to the on-target site

Are information about the potential off-target sequences provided? (number of mismatches, the position of mismatches to the PAM)

---

**Potential off-target sequences within *medium validity* studies were handled as *medium/high validity* if all the following factors applied.**

---

**All of the criteria answered with “Yes”**

Did a PAM follow the potential off-target sequences?

Has the potential off-target sequence at least one mismatch to the on-target site

Are information about the potential off-target sequences provided? (number of mismatches, the position of mismatches to the PAM)

---

If none of the above categories applied, the study was considered as *high validity*

### **Potential effect modifiers/ reasons for heterogeneity**

The following potential effect modifiers were considered for studies included in this systematic review:

- Number of mismatches: It is estimated that the number of mismatches between the target sequence and the potential off-target sequence significantly affects the occurrence of off-target effects.
- Position of mismatches proximal to the PAM: Mismatches occurring within the seed sequence (8 up to 12 nucleotides proximal to the PAM) may decrease off-target effects compared to when the mismatch is located more distant to the PAM.
- GC-content of the targeting sequence: The GC-content of a chosen target sequence may affect off-target effects, as high GC-content stabilize the sgRNA/genomic DNA hybridization (Fu et al., 2013). Therefore, a low GC-content may result in decreased off-target activity.
- Altered nuclease variants: Next to the most widely used nuclease Cas9 derived from *Streptococcus pyogenes* (SpCas9) several further Cas9 proteins from different bacterial or archaea species have been adapted (e.g. SaCas9, SpCas9-HF, Cpf1, Nickase).
- Time of incubation: Time during which genomic DNA is exposed to CRISPR/Cas (Stable or transient transformation with respective DNA sequences, RNA sequences or RNP).

Further potential effect modifiers were identified but not systematically assessed within this review:

- Different plant species (due to genome size and ploidy level)
- Prediction tools used to identify predicted potential off-target sequences
- Detection method used to identify off-target effects

### **Data synthesis and presentation**

All studies included in the systematic review were synthesized in a narrative report. No studies were excluded, but studies rated as *high* and *medium/high validity* were reported separately from studies rated as *low* and *medium/low* or *unclear validity*. All analyzed potential off-target sequences identified within the *high* and the *medium/high validity* studies were included to address five different hypothesis:

Hypothesis 1: *An increasing number of mismatches between the on-target site and the potential off-target sequence decreases the likelihood that off-target effects occur.*

Hypothesis 2: *Mismatches within the first eight nucleotides proximal to the PAM sequence (seed sequence) decrease the occurrence of off-target effects.*

Hypothesis 3: *A lower guanine-cytosine (GC)-content of the target sequence decrease the occurrence of off-target effects.*

Hypothesis 4: *Improved nuclease variants decrease the occurrence of off-target effects.*

Hypothesis 5: *The period in which the CRISPR/Cas system is active in the plant cell affects the occurrence of off-target effects.*

For all hypotheses, a descriptive overview of the available literature was given considering all *high* and *medium/high validity* studies. In addition, a quantitative synthesis was performed. Results of studies rated as *low* or *medium/low validity* as well as *unclear validity* studies were narratively reported but no specific validation regarding the formulated hypotheses was performed. Instead, based on the reasons for study validation as *low* and *medium/low* or *unclear* it was described whether off-target effects were detected or not.

### **Quantitative synthesis strategy**

363 out of 370 studies rated as *high* or *medium/high validity* were used for quantitative synthesis. Seven studies analyzing 5021 potential off-target sequences were not included for quantitative synthesis as unbiased detection methods were used to detect off-target effects. Due to the huge amount of analyzed potential sequences it was not possible to extract all mismatches including their position proximal to the PAM in detail. In none of these sequences, off-target effects were detected. Four potential off-target sequences contained three mismatches, while the remaining ones had at least four mismatches to the on-target sequence. Detailed information about the sequences included in this meta-analysis as well as a list of excluded sequences is provided in supplementary materials-table 6. We ran a binary logistic regression analysis to verify the five hypothesis to answer the question, which factors affect the occurrence of off-target effects by using CRISPR/Cas in plants. This type of analyses aims to investigate the relationship between a set of independent variables and the depended response variable and provides information about the probability of the response of interest, the independent variables which mostly affect the response of interest and the odd ratios (Walsh, 2016). As indicated in Table 6.4, number of mismatches ( $x_1 = 1/ 2/ 3/ \geq 4$ ), position of mismatch/es proximal to the PAM ( $x_2 = \text{Yes/ No}$ ), GC-content ( $x_3 = \text{high (>50\%)/ low (<50\%)}$ ), delivery method ( $x_4 = \text{DNA-stable/ others}$ ) and nuclease variant ( $x_5 = \text{SpCas9/ further nucleases}$ ) were considered as independent variables ( $x = \text{classification}$ ). Moreover, the occurrence of an off-target effect ( $Y = \text{Yes/ No}$ ) were applied as depended variable. The Fit Model platform implemented in the software package JMP (JMP® 2019. JMP®. Version 14. SAS Institute Inc. Cary, NC.) was used to fit a binary logistic regression model by using the binary logistic

personality(SAS Institute. 2018. JMP® 14 Fitting Linear Models). Therefore, the Yes level of the dependent variable was defined as response level i.e. the model estimates the probability of the Yes level of the off-target-effect variable (SAS Institute. 2018. JMP® 14 Fitting Linear Models). First, a model was fitted including all independent variables described in Table 6.4 (“full model”). All variables that appear insignificant based on effect likelihood ratio chi-square test were excluded. Then, a second model (“reduced model”) was fitted which includes all significant variables of the full model. Again, all variables that appear insignificant were excluded resulting in a final model (“final model”) obtaining the significant variables. The Fit Model platform provides three tests to evaluate the model fit: (i) whole model test, (ii) lack of fit test and (iii) likelihood ratio effect test (SAS Institute. 2018. JMP® 14 Fitting Linear Models). Additionally, to evaluate the accuracy and predictive ability of the final model, a confusion matrix was created (SAS Institute. 2018. JMP® 14 Fitting Linear Models) and sensitivity (true positive rate) and specificity (false positive rate) were calculated (Walsh, 2016). Furthermore, a receiver operating characteristic (ROC) curve was created and the area under the curve (AUC) was calculated. A ROC curve greatly exceeds the diagonal and a high AUC value indicates a model with good accuracy and a high predictive ability (Walsh, 2016). Furthermore, odds ratios were estimated based on the final model.

**Table 6.4:** Description of independent Variables.

| #  | Variable                                 | Variable description   | Level      | Number of analyzed potential off-target sequences | Total (%) |
|----|--|--|------------|---|-----------|
| x1 | Number of mismatches                     | One mismatch   | 1          | 154   | 11        |
|    |  | Two mismatches   | 2          | 218   | 16        |
|    |  | Three mismatches   | 3          | 352   | 25        |
|    |  | Four or more mismatches  | ≥4         | 671   | 48        |
| x2 | Position mismatch/es proximal to the PAM | (Any) mismatch within the seed sequence (Position 1-8 proximal to the PAM) | Yes        | 862   | 62        |
|    |  | Any mismatch/es within the seed sequence                                   | No         | 533   | 38        |
| x3 | GC-content                               | < 50%  | Low        | 505   | 36        |
|    |  | ≥ 50%  | High       | 890   | 64        |
| x4 | Delivery method                          | DNA-stable   | DNA-stable | 1229  | 88        |
|    |  | Others (DNA-transient, RNP, RNA)   | Others     | 166   | 12        |
| x5 | Nuclease variant                         | CRISPR/SpCas9  | SpCas9     | 1218  | 87        |
|    |  | Further Nucleases  | Others     | 177   | 13        |

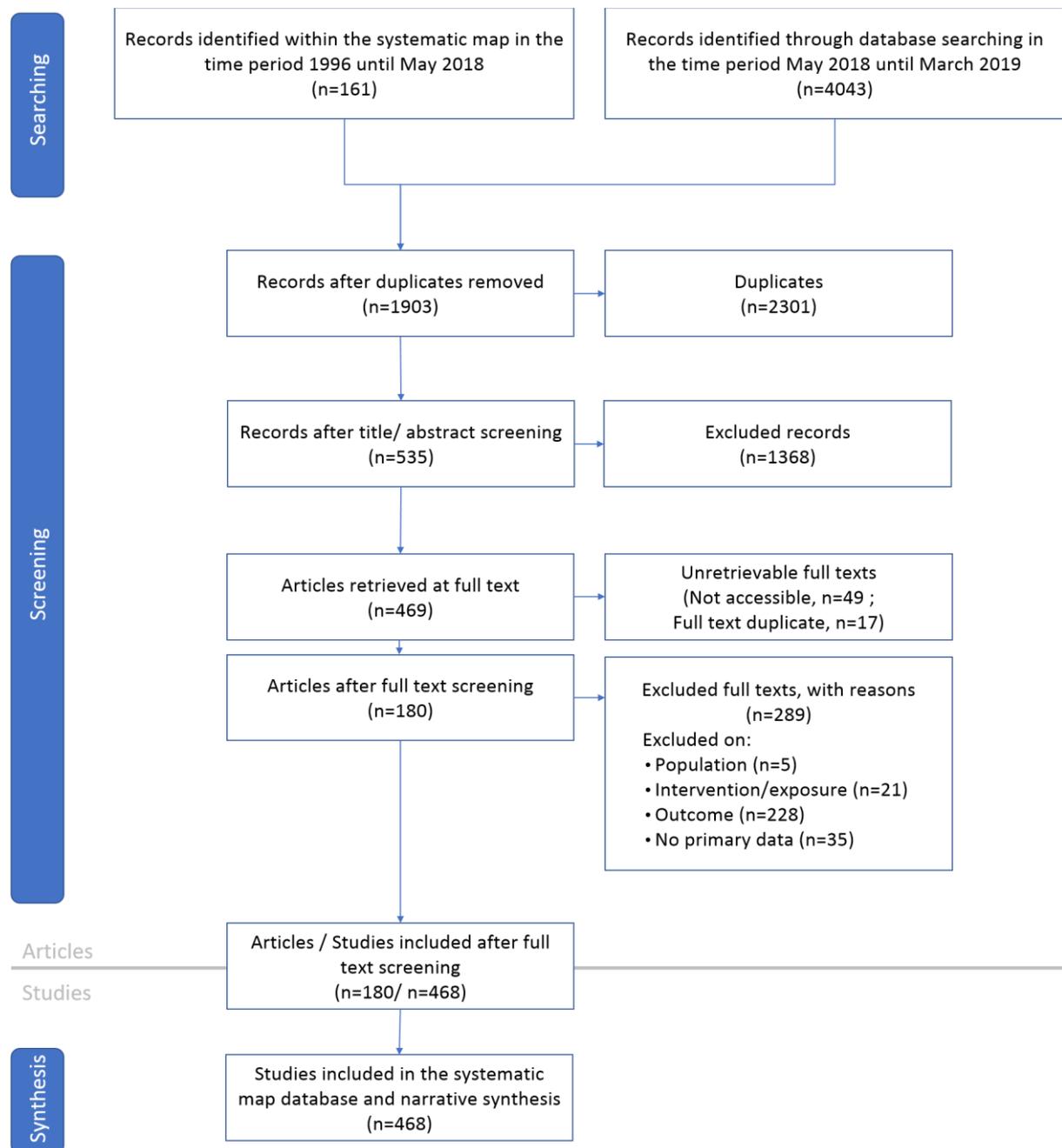
## 6.5 Results

### 6.5.1 Review descriptive statistics

#### The evidence base

Our systematic review included in total 468 studies from 180 articles. Figure 6.1 presents a flow diagram of the systematic review process with the number of articles and studies included and excluded at each stage. 133 articles (306 studies) were derived from the systematic map (secondary question 2) that preceded this review (Modrzejewski et al., 2019). 28 articles from the systematic map were not rated eligible for inclusion in this review because of ineligible intervention (n=19)

or ineligible outcome (n=9). The literature update identified further 47 articles (162 studies) for the period May 2018 until March 2019. A list of all articles and studies comprised in this systematic review is provided in supplementary materials- table 3 and includes the extracted data.



**Figure 6.1** Flow diagram of the systematic mapping process explaining the selection of relevant articles and studies. This diagram follows ROSES guidance (Haddaway et al., 2018).

### Characteristics of articles included in narrative synthesis

All articles included in this systematic review were published in peer-reviewed journals in English language.

### Articles per year

In 2013, the first four articles were identified that assessed off-target effects in plants after applying CRISPR/Cas. Since that time, the number of such articles on CRISPR/Cas increased continuously to 62 in 2018. In the first quarter in 2019, the number of relevant articles reached 21.

### Study design

We identified five different possibilities how potential off-target sequences have been identified and analyzed:

- (i) Potential off-target sequences that are similar to the target sequence were identified using bioinformatics approaches (mainly CRISPR-P, n=104; BLAST, n=61 and Cas-OFFinder, n=55). Then a subset of potential off-target sequences with the considered highest likelihood for occurrence of off-target effects were analyzed using biased detection tools (mainly PCR + Sequencing).
- (ii) Potential off-target sequences that are similar to the target sequence were identified using bioinformatics approaches. All potential off-targets sites with up to a certain number of mismatches were analyzed for the occurrence of off-target effects.
- (iii) Potential off-target sequences for the chosen target sequence were already known (for example due to polyploidy of plants or due to a priori research). Only these sequences were analyzed for the occurrence of off-target effects.
- (iv) To analyze the binding specificity of the CRISPR/Cas system, variations were made in the protospacer binding site. The target site was then assessed for the occurrence of mutations.
- (v) The whole genome was sequenced for the occurrence of off-target effects.

### Results critical appraisal

326 out of 468 studies were rated *high validity*, 44 to have *medium/high* and *medium/low validity*, 72 to have *low validity* and 26 to have *unclear validity*. One study potentially rated as *high validity* was downgraded to *low validity* because the sgRNA/Cas9 complex was constitutively overexpressed but it is not traceable how this overexpression was actually achieved (Ji et al., 2018). Detailed information about the validity assessment of all studies including the reason for their classification is provided in Table 6.5 and supplementary materials- table 5. Within the studies rated as *high validity*, 2267 potential off-target sequences were assessed for the occurrence of off-target effects. In addition, 4149 potential off-target sequences identified from 44 studies rated as *medium validity* fulfilled the required criteria and were rated as *medium/high validity* ((i) sequences were followed by a PAM (ii) All information were given (iii) sequences had at least one mismatch to the on-target sequence). Therefore, the database of *high* and *medium/high validity* studies comprises 370 on-target sequences and in total 6416 potential off-target sequences were analyzed to identify off-target effects. Within the studies rated as *low validity*, 197 potential off-target sequences were assessed for the occurrence of off-target effects. In addition, 154 potential off-target sequences identified from 44 studies were rated as *medium/low validity* as they didn't fulfill the required criteria ((i) sequences were not followed by a PAM (ii) Information are incomplete (iii) on-target and off-target site were identical). Adding these sites in total, 351 potential off-target sequences were rated as *low* or *medium/low validity*. 170 potential off-target sequences lack sufficient information and were rated as *unclear validity*.

**Table 6.5** Overview on the validity assessment of all studies and the reason for validation.

| Validity assessment     | Number of studies | Reason for validity assessment outcomes  | Number of analyzed potential off-target sites                                |
|-------------------------|-------------------|--|--|
| <i>High validity</i>    | 326               | All study validity assessment criteria were rated as “Yes”   | <i>High validity</i> : 2267  |
| <i>Medium validity</i>  | 44                | Some potential off-target sequences were followed by a PAM (n=28)<br>Information are considerably incomplete (n=15)<br>Only some potential off-target sequences show at least one mismatch (n=7)   | <i>Medium/ High validity</i> : 4149<br><br><i>Medium/ Low validity</i> : 154 |
| <i>Low validity</i>     | 72                | No on-target mutation was induced (n=27)<br>Reference genome has not been fully sequenced/sequencing error when analyzing potential off-target sequences (n=7)<br>Potential off-target sequences were not followed by a PAM (n=15)<br>All potential off-target sequences have no mismatch (n=5)<br>No information is provided (n=23)<br>Overexpression of the sgRNA/Cas9 complex (n=1) | <i>Low validity</i> : 197  |
| <i>Unclear validity</i> | 26                | Lacking sufficient information to judge (n=26)   | <i>Unclear validity</i> : 170  |

### Characteristics of *high* and *medium/high* validity studies

In total, 370 studies were rated as having *high* or *medium/high* validity. These studies analyzed a total number of 6416 potential off-target sequences for the occurrence of off-target effects. Table 6.6 summarizes key characteristics of these analyzed sequences, including the applied nuclease variant, delivery method and the number of analyzed potential off-target sequences. 13 different nuclease variants were identified that were applied and assessed for off-target effects in plants. Mostly the nuclease SpCas9 was used (97%) followed by Cpf1 (1%), eSpCas9 (1.0) (0.5%), eSpCas9 (1.1) (0.5%) and SaCas9 (0.5%).

In the large majority, the CRISPR/Cas system was delivered into the cells via stable transformation of DNA (321 studies). 39 studies transiently supplied the CRISPR/Cas system as plasmid DNA (*DNA-transient*). Eight studies used a DNA-free approach by supplying the CRISPR/Cas system as ribonucleoproteins (*RNP*) and two studies delivered the CRISPR/Cas system via RNA sequence. More than 93% of the analyzed potential off-target sequences were assessed when using the standard nuclease SpCas9 and the system was supplied into the cell as stably integrated DNA approach (*DNA-stable*).

**Table 6.6** Overview about the characteristic items and frequency of studies rated as *high* and *medium/high* validity.

| Study validity assessment                            | n=  | Nuclease variant | n=  | Delivery method | n=  | Analyzed potential off-target sequences |
|--|-----|------------------|-----|-----------------|-----|---|
| <i>High validity</i> and <i>medium/high</i> validity | 370 | SpCas9           | 311 | DNA-stable      | 283 | 5999                                    |
|  |     |                  |     | DNA-transient   | 18  | 78                                      |
|  |     |                  |     | RNPs            | 8   | 133                                     |
|  |     |                  |     | RNA             | 2   | 9                                       |

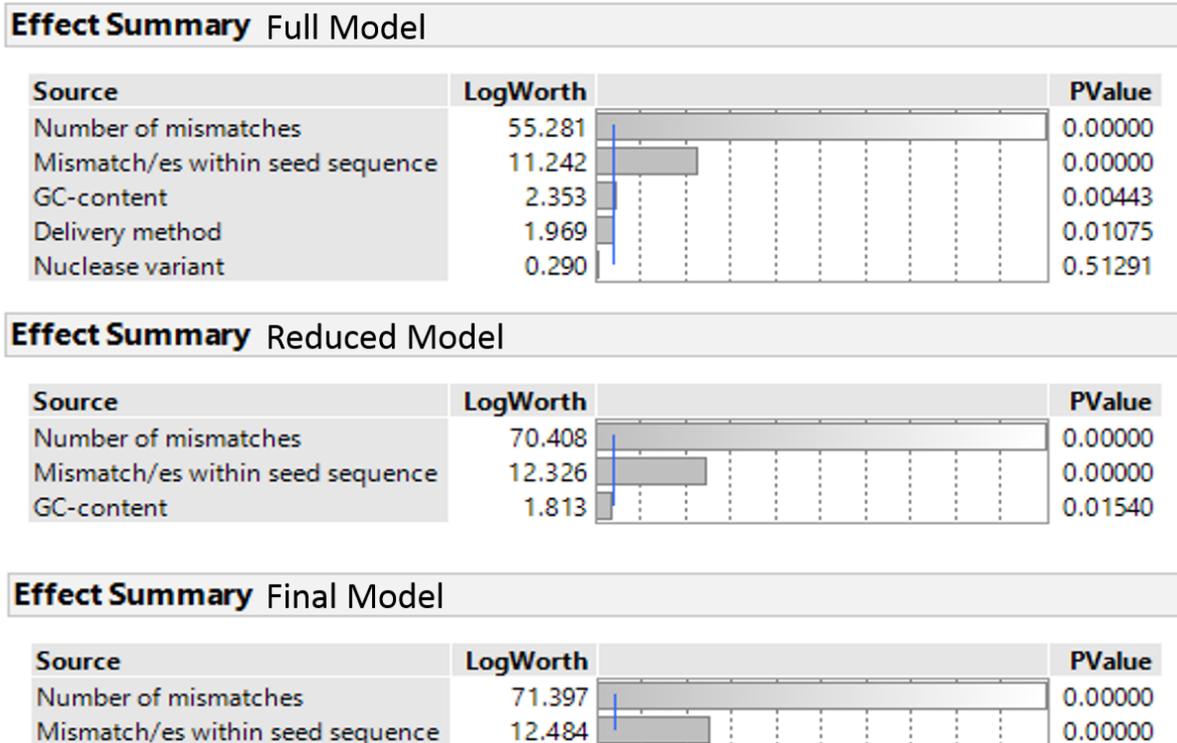
|               |    |               |    |    |
|---------------|----|---------------|----|----|
| Cpf1          | 20 | DNA-stable    | 20 | 62 |
| eSpCas9 (1.1) | 8  | DNA-stable    | 4  | 26 |
|               |    | DNA-transient | 4  | 7  |
| eSpCas9 (1.0) | 8  | DNA-stable    | 3  | 24 |
|               |    | DNA-transient | 5  | 8  |
| SaCas9        | 6  | DNA-stable    | 2  | 5  |
|               |    | DNA-transient | 4  | 4  |
| Others        | 17 | DNA-stable    | 9  | 47 |
|               |    | DNA-transient | 8  | 14 |

RNP=Ribonucleoproteins; RNA= Ribonucleic acid.

Based on the results gained within the *high* and *medium/high validity* studies, we conducted a meta-analysis and assessed five different hypothesis about the occurrence of off-target effects in CRISPR/Cas studies in plants. In addition, a descriptive overview based on the identified evidence base is provided for each hypothesis.

### 6.5.2 Meta-analysis of *high* and *medium/high validity* studies

Binary regression analysis was conducted to evaluate if the independent variables “number of mismatches”, “position of mismatches proximal to the PAM”, “GC-content”, “nuclease variant” and “delivery method” affect the occurrence of off-target effects. In a first step, a full model was fitted including all variables as indicated in Table 6.4. Then, the full model was reduced by removing the nuclease variant and the delivery methods, as these variables appeared to be insignificant (reduced model) (Figure 6.2). In a next step, a reduced model was fitted including the remaining variables. Due to insignificance, the reduced model was further reduced by removing the variable GC-content as it appeared to be insignificant. Therewith we gained the final model including the variables number of mismatches and position of mismatches proximal to the PAM (Figure 6.2). The final binary regression analysis revealed that the number of mismatches and the position of mismatches proximal to the PAM significantly affect the occurrence of off-target effects, whereby the number of mismatches has the dominant impact (Figure 6.2).



**Figure 6.2** Effect summary of the binary regression analysis for the full model, the reduced model and the final model. The LogWorth transformation adjusts PValue to provide an appropriate scale for graphing. A value that exceeds 2 is significant.

The whole model test, lack of fit test and likelihood ratio effect test indicate that the final model fitted the data in an appropriate and sufficient way (details are provided in supplementary materials-table 6). Furthermore, the observed accuracy rate indicated that the final model predicted 92% of the recorded off-target effects correctly i.e. correct assignment to the “Yes” or “No” level of the dependent variable. In detail, the amount of true negatives was higher than the amount of true positives, indicated by a specificity of 0.98 and a sensitivity of 0.46. Due to the bias in specificity and sensitivity, which can be caused by unbalanced ratio between the Yes and No level of the dependent variable, a ROC curve was additionally used to evaluate the model fit. The ROC curve exceeded the diagonal and the corresponding AUC value (0.92) indicates a model with good accuracy and a high predictive ability (supplementary materials- table 6). To verify the five hypotheses odds ratios were calculated for the variables with the final model.

*Hypothesis 1: An increasing number of mismatches between the on-target and the potential off-target sequence decreases the probability that off-target effects occur.*

The probability of the occurrence of an off-target effect if the number of mismatches is one is 3.8, 36.7 or 160.6 times higher than if the number of mismatches is 2, 3 or  $\geq 4$ , respectively (Table 6.7a). Therefore, the probability of the occurrence of an off-target effect is highest for number of mismatches = 1 and is lowest for the number of mismatches =  $\geq 4$ . To conclude based on the odds ratios calculated for the final model hypothesis 1 can be confirmed.

Hypothesis 2: *Mismatches within the first eight nucleotides proximal to the PAM decrease the occurrence of the off-target effect.*

The probability of the occurrence of an off-target effect if the mismatches are not within the first eight nucleotides proximal to the PAM is 4.9 times higher than if the mismatches are within the first eight nucleotides proximal to the PAM, (Table 6.7b). Therefore, based on the odds ratios calculated for the final model hypothesis 2 can be confirmed.

Hypothesis 3: *A lower GC-content of the target sequence decrease the occurrence of off-target effects.*

The variable GC-content did not significantly contribute to the fitted model and were not included as independent variable in the binary logistic regression analysis. Therefore, it is not possible to confirm this hypothesis.

Hypothesis 4: *Improved nuclease variants decrease the occurrence of off-target effects*

The variable nuclease variant did not significantly contribute to the fitted model and were not included as independent variable in the binary logistic regression analysis. Therefore, it is not possible to confirm this hypothesis.

Hypothesis 5: *The period in which the CRISP/Cas construct is active in the plant cell affects the occurrence of off-target effects.*

The delivery method did not significantly contribute to the fitted model and were not included as dependent variable in the binary logistic regression analysis. Therefore, it is not possible to confirm this hypothesis.

**Table 6.7** Odds ratios for number of mismatches and mismatch/es within seed sequence..

a) Number of mismatches.

| Level 1 | Level 2 | Odds Ratio | Prob>Chisq | Lower 95% | Upper 95% |
|---------|---------|------------|------------|-----------|-----------|
| 2       | 1       | 0.2661273  | <.0001     | 0.1661508 | 0.4262619 |
| 3       | 1       | 0.0272239  | <.0001     | 0.0135418 | 0.05473   |
| 3       | 2       | 0.1022967  | <.0001     | 0.0515393 | 0.2030414 |
| ≥ 4     | 1       | 0.0062283  | <.0001     | 0.0024145 | 0.016066  |
| ≥ 4     | 2       | 0.0234035  | <.0001     | 0.009159  | 0.0598016 |
| ≥ 4     | 3       | 0.2287805  | 0.0069     | 0.0784442 | 0.6672328 |
| 1       | 2       | 3.7576002  | <.0001     | 2.3459754 | 6.0186306 |
| 1       | 3       | 36.732378  | <.0001     | 18.271515 | 73.845412 |
| 2       | 3       | 9.7754885  | <.0001     | 4.9251049 | 19.402668 |
| 1       | ≥ 4     | 160.55728  | <.0001     | 62.243085 | 414.16069 |
| 2       | ≥ 4     | 42.728674  | <.0001     | 16.72197  | 109.18209 |
| 3       | ≥ 4     | 4.3710014  | 0.0069     | 1.4987273 | 12.747919 |

b) Mismatch/es within seed sequence.

| Level 1 | Level 2 | Odds Ratio | Prob>Chisq | Lower 95% | Upper 95% |
|---------|---------|------------|------------|-----------|-----------|
| No      | Yes     | 0.48958322 | <.0001     | 3.1263158 | 7.6669071 |
| Yes     | No      | 0.2042554  | <.0001     | 0.1304307 | 0.3198653 |

### 6.5.3 Descriptive synthesis of *high* and *medium/high validity* studies

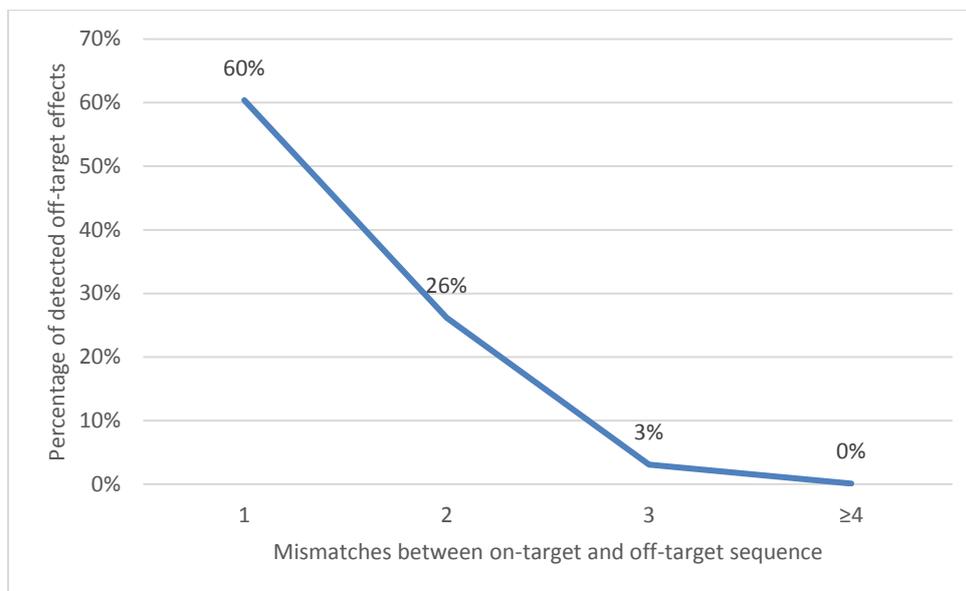
The 370 studies rated as *high* or *medium/high validity* assessed a total amount of 6416 potential off-target sequences. 154 analyzed potential off-target sequences showed one mismatch (2.4% of the total analyzed sequences), 218 had two mismatches (3.4%), 356 had three mismatches (5.6%), while the remaining 5688 sequences (88.7%) referred to four or more mismatches.

#### 6.5.3.1 Hypothesis 1: Number of mismatches

**Hypothesis 1:** *An increasing number of mismatches between the on-target site and the potential off-target sequence decreases the likelihood that off-target effects occur.*

It is purported that the main factor to control off-target effects is a careful selection of the target sequence (Hsu et al., 2013, Zhu et al., 2017, Zischewski et al., 2017). This hypothesis aims to assess the occurrence of off-target effects depending on the number of mismatches between the on-target and potential off-target sequences.

As shown in Figure 6.3, the number of identified off-target effects steeply decreases the more mismatches occur between the on-target and potential off-target sequences. One mismatch in respect to the on-target sequence lead to the detection of off-target effects in 93 out of 154 analyzed potential off-target sequences (59%). Two mismatches decreased the occurrence of off-target effects to 26% (57 identified off-target effects out of 218 analyzed potential off-target sequences). Three mismatches between the on-target and potential off-target sequence decreased the off-target rate to 3% (11 identified off-target sequences out of 356 analyzed potential off-target sequences). Four or more mismatches further decreased the off-target rate to 0.09% (five off-target effects out of 5688 analyzed potential off-target sequences). The five identified off-target effects had two times four and five mismatches each and once six mismatches.



**Figure 6.3** Percentage of detected off-target effects for one (n=154), two (n=218), three (n=356) and four or more (n=5688) mismatches. Calculation: Number of detected off-target effects/number of analyzed potential off-target sequences.

### Conclusions for hypothesis 1

The data clearly indicate that the likelihood of off-target effects can be minimized drastically by selecting a target sequence that shows more than three mismatches to any similar sequences in the genome.

### Consequences of hypothesis 1 for the evaluation of the hypotheses 2-5

The results for the first hypothesis show that the number of mismatches to the target sequence has a very strong influence on the occurrence of off-target effects. In addition, it was shown that off-target effects occur at considerable low likelihood when the potential off-target sequences bear at least four mismatches. Therefore, the hypotheses 2-5 will be examined separately for one and two mismatches to elucidate a consideration independently from the interfering number of mismatches. In addition, detected off-target effects containing three or more mismatches to the target sequence will be reviewed though the likelihood of an effect is generally low.

#### 6.5.3.2 Hypothesis 2: Position of mismatches

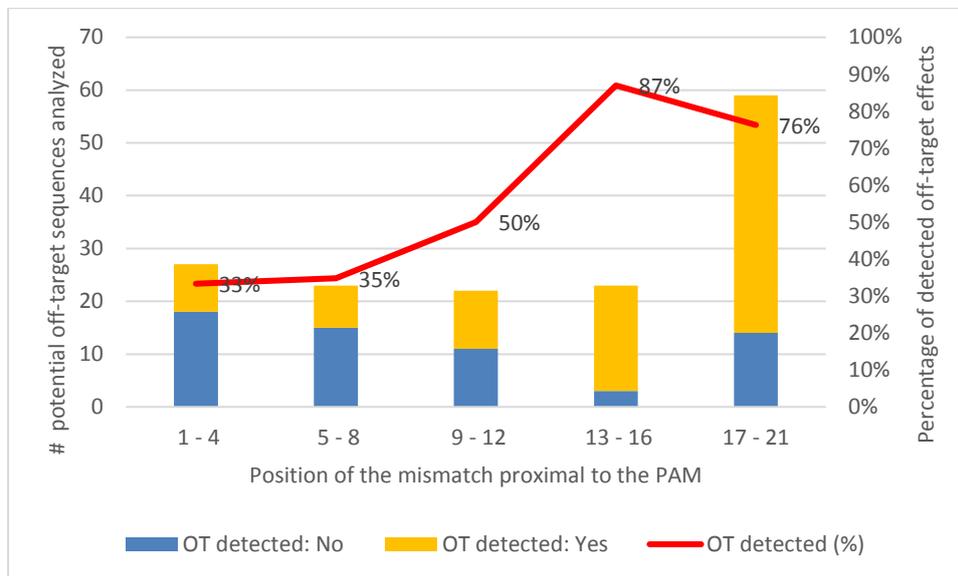
**Hypothesis 2:** *Mismatches within the first eight nucleotides proximal to the PAM sequence decrease the occurrence of off-target effects.*

It was considered that mismatches occurring within the first nucleotides proximal to the PAM determine the occurrence of off-target effects, leading to a decreased or even complete abolition of off-target effects (Endo et al., 2015, Hahn, Nekrasov, 2018, Hsu et al., 2013). This hypothesis aims to assess the occurrence of off-target effects considering the position of the mismatch/es in potential off-target sequences.

#### One mismatch between the on-target and the potential off-target sequence

In total, 154 potential off-target sequences with one mismatch were assessed for off-target effects. Because of the limited set of data, we analyzed the incidences of off-target effects based on the position of the mismatch proximal to the PAM in five intervals of four nucleotides each covering

the whole guide sequence (Figure 6.4). If the mismatch occurred at position one to four proximal to the PAM, off-target effects were detected in 33% of the analyzed sites (nine off-target effects identified out of 27 potential off-target sequences (9/27)). A mismatch at position five to eight proximal to the PAM showed a similar off-target rate of 35% (8/23). The off-target rate considerably increased to 50% when the mismatch was located at position 9-12 (11/22) and up to 87% in case the mismatch was located at position 13-16 (20/23). A high off-target rate of 76% (45/59) was also observed when the mismatch was located at position 17-21 to the PAM.



**Figure 6.4** Percentage of detected off-target effects for one mismatch based on the position the mismatch is located proximal to the PAM. The guide sequence is divided in five intervals of four nucleotides each (Position 1-4 proximal to the PAM, position 5-8, 9-12, 13-16, 17-21). Yellow: Number of studies with detected off-target effects; blue: Number of studies without detected off-target effects; Red: Percentage of detected off-target effects for each interval.

### Two mismatches between the on-target and the potential off-target sequences

218 potential off-target sequences with two mismatches were assessed for off-target effects. We draw a cross table to display the off-target rate based on the position of the two mismatches proximal to the PAM within the intervals as before (Figure 6.5). The results indicate that the further the mismatches are distant from the PAM the higher the off-target rate. The results also indicate that the off-target rate increases when the two mismatches are located close to each other. When both mismatches are located within the seed sequence at positions five to eight, off-target effects occurred in 35% of the analyzed potential off-target sequences (six off-target effects were detected out of 17 potential off-target sequences). This rate is in the same range as if the potential off-target sequence has only one mismatch located at the position five to eight (35%). In case both mismatches are located at position 9-12, 13-16 or 17-21, off-target effects were detected in a range up to 52% of the analyzed potential off-target sequences. In case the two mismatches were more distant from each other, off-target effects occurred less frequent. Independently of the position of the mismatches to the PAM, in case the two mismatches were directly next to each other, the off target rate was 38%. In case the mismatches were not located next to each other, the off-target rate was only 15%.

|                                  |       | Position to PAM: First mismatch |               |                |               |                |
|----------------------------------|-------|---------------------------------|---------------|----------------|---------------|----------------|
|                                  |       | 1-4                             | 5-8           | 9-12           | 13-16         | 17-21          |
| Position to PAM: Second mismatch | 1-4   | 4%<br>(1/23)                    | 18%<br>(2/11) | 6%<br>(1/18)   | 0%<br>(0/3)   | 6%<br>(1/16)   |
|                                  | 5-8   | 18%<br>(2/11)                   | 35%<br>(6/17) | 0%<br>(0/12)   | 11%<br>(1/9)  | 17%<br>(1/6)   |
|                                  | 9-12  | 6%<br>(1/18)                    | 0%<br>(0/12)  | 52%<br>(11/21) | 35%<br>(6/17) | 50%<br>(5/10)  |
|                                  | 13-16 | 0%<br>(0/3)                     | 11%<br>(1/9)  | 35%<br>(6/17)  | 44%<br>(7/16) | 24%<br>(4/17)  |
|                                  | 17-21 | 6%<br>(1/16)                    | 17%<br>(1/6)  | 50%<br>(5/10)  | 24%<br>(4/17) | 50%<br>(11/22) |

| Off-target rate |
|-----------------|
| 0-15%           |
| 16-30%          |
| 31-45%          |
| 46-60%          |

**Figure 6.5** Incidence of off-target effects when two mismatches occur proximal to the PAM. (x/y): (“x” off-target effects identified / “y” potential off-target sequences analyzed).

### Three and more mismatches between the target sequence and the potential off-target sequence

Only 0.27% off-target effects (16 out of the 6028) were detected when there were three or more mismatches. Seven of these sequences had one mismatch between position one to eight proximal to the PAM and the remaining ones beyond position eight. In nine potential off-target sequences, all mismatches were located beyond position eight.

### Conclusion for hypothesis 2

The data base (i.e. number of cases per class) for one and two mismatches is relatively limited, but based on the available evidence, there is a tendency that off-target effects are reduced when the mismatch/es are located within the first eight nucleotides proximal to the PAM. If the mismatch/es are located within this region, off-target effects appear reduced compared to the case when the mismatch/es are located outside the seed sequence. Nevertheless, the database regarding one mismatch indicates, that even when the mismatch is located inside this region off-target effects still occurred in around one third of the analyzed potential off-target sequences (compared to 59% observed for the whole guide sequence in Figure 6.3). It is further indicated that adjoined mismatches increase the likelihood of off-target effects. Hence, care should be taken when selecting a target sequence that the mismatches of potential off-target sequences are not located next to each other.

### 6.5.3.3 Hypothesis 3: GC-content

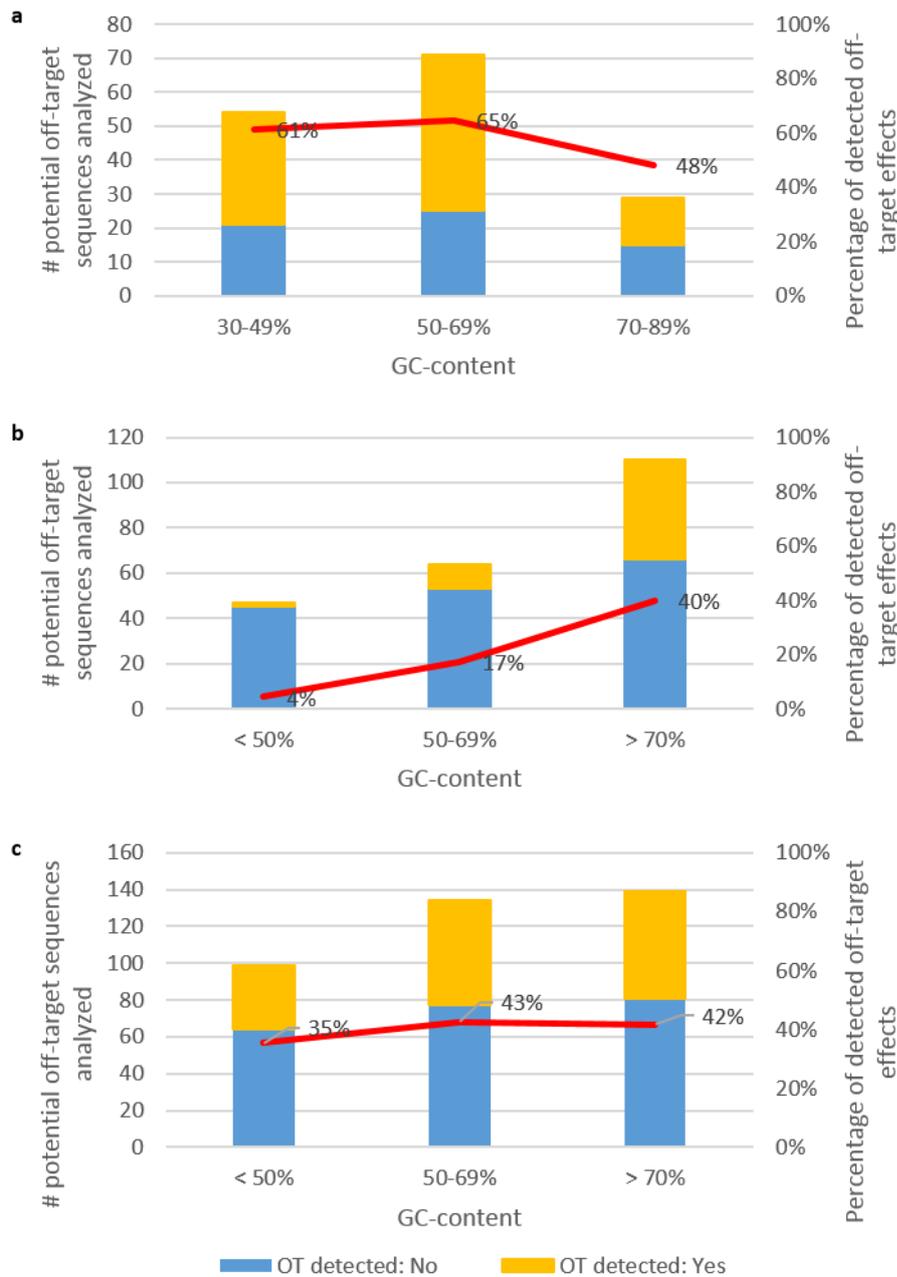
**Hypothesis 3:** *A lower guanine-cytosine (GC)-content of the target sequence decreases the occurrence of off-target effects.*

Some scientists suggested that the GC-content of a chosen target sequence may influence off-target effects, as a high GC-content stabilize the sRNA/genomic DNA hybridization (Fu et al., 2013).

Therefore, it is estimated that a low GC-content decreases the occurrence of off-target effects (Yu et al., 2017). This hypothesis aims to assess whether the GC-content affects the occurrence of off-target effects. The GC-content was calculated as follows:  $GC = \#\{G \text{ or } C \text{ in guide}\} / \text{guide\_length}$ .

### **One mismatch between the target sequence and the potential off-target sequence**

Figure 6.6a considers the number of analyzed potential off-target sequences having one mismatch to the target sequence in relation to the GC-contents of the target sequence. The target sequences had a GC-content in a range of 30% and 90%. Due to the limited database, we subdivided the potential off-target sequences in three groups based on the GC-content: 30-49%, 50-69% and 70-89%. In case the GC-content was between 30 to 49% (n=54 analyzed potential off-target sequences), 60% of the analyzed off-target sequences showed off-target effects. A similar off-target rate was detected when the GC-content was between 50 to 69%, as 65% of the analyzed sequences showed off-target effects (n=71). If the GC-content was between 70 to 89% (n=29), off-target effects were detected at 48% of the analyzed sequences.



**Figure 6.6** Rate of identified off-target effects assessing potential off-target sequences with a) one mismatch b) two mismatches c) one and two mismatches to the target sequence considering the GC-content of the protospacer.

### Two mismatches between the target sequence and the potential off-target sequences

Regarding two mismatches, the GC-content of the target sequences varied in a range of 25 to 95%, but only two sequences had a GC-content below 30% and two higher than 90%. Based on these findings we subdivided the potential off-target sequences in three groups based on the GC-content: <49%, 50-69% and >70%. As shown in Figure 6.6b the number of identified off-target effects increased from 4% identified off-target effects for a GC-content lower than 49% (n=45 analyzed sequences), over 17% when the GC-content was between 50 and 69% (n=53), to 40% in case the GC-content was higher than 70% (n=110).

Regarding one mismatch, the results indicate that the occurrence of off-target effects is decreased when the GC-content is  $\geq 70\%$ , whereas when considering two mismatches, it seems that the off-target rate is increasing when the GC-content is  $\geq 70\%$ . These results may indicate an interaction

with the number and location of mismatches, though it cannot be resolved from the given data set. In Figure 6.6c, we assessed off-target effects for different GC-contents of the target sequence for one and two mismatches combined.

### **Three and more mismatches between the target sequence and the potential off-target sequence**

Regarding three mismatches, in total 11 off-target effects were identified. In four cases the targeted sequence had a GC-content lower than 50%. Four off-target effects were detected when the target sequence had a GC-content between 50 to 69% and in three cases the target sequence had a GC-content higher than 70%. Regarding at least four mismatches five off-target effects were identified and three of these sequences had a GC-content lower than 50%, while two ones had a GC-content between 50 to 69% (details are provided in supplementary materials- table 4).

### **Conclusion for hypothesis 3**

Based on the available evidence the results do not indicate that the occurrence of off-target effects shows an obvious trend in relation to the GC-content of the protospacer sequence. Due to the limited database it was not possible to analyze in detail the GC-content independently of further factors potentially affecting off-target effects.

### **6.5.3.4 Hypothesis 4: Nuclease variant**

**Hypothesis 4:** *Improved nuclease variants decrease the occurrence of off-target effects.*

Much effort is spent into the research of new nucleases and the further development of existing ones. Off-target effects may be reduced due to longer recognition sequences (e.g. SaCas9), different PAMs (e.g. Cpf1) or protein-engineered approaches resulting in altered SpCas9 variants (e.g. SpCas9-HF). Hypothesis 4 aims to investigate off-target effects in different nuclease variants.

Table 6.8 provides an overview about the analyzed potential off-target sequences for different nuclease variants. In total, 13 different nucleases were investigated. 6219 potential off-target sequences were analyzed using the standard SpCas9 nuclease (97% of the total). Further 197 potential off-target sequences were analyzed with regards to the nuclease variants Cpf1 (1%), eSpCas9 (1.1) (0.5%), eSpCas9 (1.0) (0.5%) and SaCas9 (0.1%). Another 61 potential off-target sequences (1% of the total) were analyzed for nucleases like eSaCas9, Cas9-HF1, VQR variants, xCas9 3.7, SpCas9-DE, SpCas9-KA or truncated SpCas9. Since none of these nuclease variants were analyzed in more than three studies their data were summarized among “others”. As the data pool is quite limited except for SpCas9, the direct comparison of the general off-target activity for different nucleases is only rarely possible. The investigated potential off-target sequences of the different nuclease variants differ strongly in the number of mismatches. Regarding the Cpf1 nuclease, only seven potential off-target sequences were examined with one or two mismatches to the target sequences (11%), while the remaining 55 sequences had three or more mismatches. In contrast, the potential off-target sequences of the mutated SpCas9 variants eSpCas9 (1.1) and eSpCas9 (1.0) had one or two mismatches in 91-97% of the investigated sequences, while no sequence had four or more mismatches. Nevertheless, in all nuclease variants off-target effects have been identified if the potential off-target sequence differs from the target one in just one mismatch. Since the data basis is limited for all nucleases with except for the standard nuclease

SpCas9, no reliable conclusions can be made about differences in the specificity of different nuclease variants.

**Table 6.8** Overview about the characteristics of analyzed potential off-target sequences using different nuclease variants.

| Nuclease variant | Number of studies | Analyzed potential off-target sequences | Number of mismatches | Analyzed potential off-target sequences | Off-target effects identified |
|------------------|-------------------|---|----------------------|---|-------------------------------|
| SpCas9           | 311               | 6219                                    | 1 mismatch           | 95                                      | 54                            |
|                  |                   |   | 2 mismatches         | 147                                     | 35                            |
|                  |                   |   | 3 mismatches         | 344                                     | 10                            |
|                  |                   |   | ≥ 4 mismatches       | 5628                                    | 5                             |
| Cpf1             | 20                | 62                                      | 1 mismatch           | 4                                       | 3                             |
|                  |                   |   | 2 mismatches         | 3                                       | 0                             |
|                  |                   |   | 3 mismatches         | 4                                       | 0                             |
|                  |                   |   | ≥ 4 mismatches       | 51                                      | 0                             |
| eSpCas9 (1.1)    | 8                 | 33                                      | 1 mismatch           | 9                                       | 6                             |
|                  |                   |   | 2 mismatches         | 21                                      | 9                             |
|                  |                   |   | 3 mismatches         | 3                                       | 1                             |
|                  |                   |   | ≥ 4 mismatches       | 0                                       | -                             |
| eSpCas9 (1.0)    | 8                 | 32                                      | 1 mismatch           | 10                                      | 8                             |
|                  |                   |   | 2 mismatches         | 21                                      | 8                             |
|                  |                   |   | 3 mismatches         | 1                                       | 0                             |
|                  |                   |   | ≥ 4 mismatches       | 0                                       | -                             |
| SaCas9           | 6                 | 9                                       | 1 mismatch           | 4                                       | 4                             |
|                  |                   |   | 2 mismatches         | 1                                       | 0                             |
|                  |                   |   | 3 mismatches         | 0                                       | -                             |
|                  |                   |   | ≥4 mismatches        | 4                                       | 0                             |
| Others           | 17                | 61                                      | 1 mismatch           | 32                                      | 18                            |
|                  |                   |   | 2 mismatches         | 25                                      | 5                             |
|                  |                   |   | 3 mismatches         | 4                                       | 0                             |
|                  |                   |   | ≥ 4 mismatches       | 0                                       | -                             |

However, in three articles, targeted experiments were conducted by applying the same study design differing just by the nuclease variant to directly compare their off-target effects. In the first article four different nuclease variants (CRISPR/SpCas9, CRISPR/eSpCas9 (1.0), CRISPR/eSpCas9 (1.1) and CRISPR/SpCas9-HF1) were compared for off-target effects (Zhang et al., 2017). The off-target analysis showed that the three altered SpCas9 variants CRISPR/eSpCas9 (1.0), CRISPR/eSpCas9 (1.1) and CRISPR/SpCas9-HF1 had substantially less off-target effects compared to the wild type SpCas9 variant and SpCas9-HF1 consistently exhibited the lowest off-target activity at the five examined potential off-target sequences. In addition, the on-target: off-target indel frequency ratios of the three altered SpCas9 variants were, on average, 273-fold higher compared to wild type SpCas9. However, no nuclease variant was completely free of off-target effects when there was just one mismatch between the on-target and potential off-target sequence. Raitskin et al. (2019)

conducted three experiments. In the first one they directly compared five variants of SpCas9 (CRISPR/SpCas9, CRISPR/SpCas9-DE, CRISPR/SpCas9-KA, CRISPR/eSpCas9 (1.0), CRISPR/eSpCas9 (1.1)) by designing a set of five sgRNAs each with a mutation in a different base of the spacer. Results indicate that while the number of mutations induced by standard SpCas9 was significantly reduced when the spacer contained a mutation in the region close to the PAM, the presence of a mismatch between the spacer and target in the distal region had minimal effects. In contrast, the frequency of mutations induced by the variants eCas9 1.0 and 1.1 was significantly reduced by a mismatch in any region of the sgRNA (Raitskin et al., 2019). In the second experiment, four nuclease variants (CRISPR/SpCas9, CRISPR/eSpCas9 (1.0), CRISPR/eSpCas9 (1.1), CRISPR/xCas9 3.7) were compared and eight targets were addressed, all having one mismatch to a potential off-target sequence. Wild type SpCas9 induced eight on-target mutations and mutations were detected in all corresponding off-target sequences. Compared to this the SpCas9 variants were only able to induce mutations at three (CRISPR/eSpCas9 (1.0), CRISPR/xCas9 3.7) or four (CRISPR/eSpCas9 (1.1)) on-targets and these variants did not always reduce mutagenesis at the off-target sequences. In the third experiment the efficiency and specificity of SaCas9 was compared with eSaCas9 by addressing four sequences that have a potential off-target sequence with one mismatch. SaCas9 induced mutations at all four on-targets as well as the corresponding potential off-target sequences, while eSaCas9 were only able to induce three on-target mutations and only one potential off-target sequence was mutated (Raitskin et al., 2019). In the third study, Endo et al. (2019) compared off-target mutations induced by SpCas9 and SpCas9-NGv1 in rice (Endo et al., 2019). The results showed that off-target effects were detected for wild type SpCas9 when there was one mismatch at position 19 proximal to the PAM, while no mutations were detected when the mismatch was located at the 6th nucleotide from the PAM. Compared to this, no off-target effects were detected for SpCas9-NGv1 independently on the position of the mismatch.

#### **Conclusion for hypothesis 4**

The data pool for the occurrence of off-target effects when employing the standard nuclease SpCas9 is large. For all other nuclease variants, we identified only a very limited number of analyzed potential off-target sequences. Therefore, no conclusions can be made regarding this hypothesis providing a general tendency. However, a few investigations directly compared different nuclease variants. In these experiments the likelihood of off-target effects was reduced by using alternative nucleases, but off-target effects were not completely excluded when sequences were similar to the target sequence.

#### **6.5.3.5 Hypothesis 5: Delivery method**

**Hypothesis 5:** *The period in which the CRISPR/Cas system is active in the plant cell affects the occurrence of off-target effects.*

A further aspect that may have an impact on the occurrence of off-target effects is the delivery of the CRISPR/Cas system into the plant cell. It can be delivered to the (plant) cell either as DNA (transiently or stably expressed), RNA or directly as Ribonucleoproteins (RNP). Stable transformation leads to a permanent expression of the CRISPR/Cas system compared to the other approaches in which the CRISPR/Cas system is available only for a limited time. Therefore, it is assumed that a permanent expression leads to an increased on-target as well as an increased off-

target impact (Jansing et al., 2019, Metje-Sprink et al., 2018, Zischewski et al., 2017). Additionally, DNA-free approaches using RNP or RNA may further reduce off-target effects, as the CRISPR/Cas system is degraded rapidly within the range of few days and the mode of action is only present in the edited cells but not in the regenerated plants (Metje-Sprink et al., 2018, Woo et al., 2015). Hypothesis 5 aims to investigate off-target effects considering the period of time in which CRISPR/Cas9 is active provoked by different delivery methods.

Table 6.9 provides an overview about the number of analyzed potential off-target sequences for different delivery methods. 321 studies delivered the CRISPR/Cas system as DNA with stable integration into the genome and subsequent expression (referred as DNA-stable in the table). Within these studies, 6163 potential off-target sequences were analyzed (96% of the total). 39 studies supplied the CRISPR/Cas system as DNA but the system was not integrated into the genome and hence only transiently expressed; in total 111 potential off-target sequences were analyzed (referred as DNA-transient in the table). In eight studies, RNP were used to deliver directly the nuclease and the RNA and 133 potential off-target sequences were analyzed. In two studies, RNA was used to deliver the CRISPR/Cas system into the plant cell and nine potential off-target sequences were analyzed.

**Table 6.9** Overview about the characteristics of analyzed potential off-target sequences using delivery methods to deliver the CRISPR system into the plant cells.

| Nuclease variant | Number of studies | Analyzed potential off-target sequences | Number of mismatches | Analyzed potential off-target sequences | Off-target effects identified |
|------------------|-------------------|---|----------------------|---|-------------------------------|
| DNA-stable       | 321               | 6163                                    | 1 mismatch           | 101                                     | 52                            |
|                  |                   |   | 2 mismatches         | 207                                     | 55                            |
|                  |                   |   | 3 mismatches         | 323                                     | 9                             |
|                  |                   |   | ≥ 4 mismatches       | 5532                                    | 4                             |
| DNA-transient    | 39                | 111                                     | 1 mismatch           | 51                                      | 39                            |
|                  |                   |   | 2 mismatches         | 7                                       | 1                             |
|                  |                   |   | 3 mismatches         | 18                                      | 2                             |
|                  |                   |   | ≥ 4 mismatches       | 35                                      | 1                             |
| RNP              | 8                 | 133                                     | 1 mismatch           | 1                                       | 1                             |
|                  |                   |   | 2 mismatches         | 4                                       | 1                             |
|                  |                   |   | 3 mismatches         | 13                                      | 0                             |
|                  |                   |   | ≥ 4 mismatches       | 115                                     | 0                             |
| RNA              | 2                 | 9                                       | 1 mismatch           | 1                                       | 1                             |
|                  |                   |   | 2 mismatches         | 0                                       | -                             |
|                  |                   |   | 3 mismatches         | 2                                       | 0                             |
|                  |                   |   | ≥ 4 mismatches       | 6                                       | 0                             |

RNP: Ribonucleoproteins; DNA: deoxyribonucleic acid; RNA: ribonucleic acid

The off-target rate for the DNA-stable approach and RNP appear around 2%, which is much lower than for the DNA-transient approach with 39%. But, examining the data in detail also demonstrate that the data for the different delivery methods represent considerable systematic differences in the likelihood that off-target effects can be observed: following from the results derived for hypothesis

1 the uneven distribution of low and high number of mismatches between the delivery methods counteract a balanced evaluation. Applying the DNA-stable approach, only 5% of the investigated potential off-target sequences represent one or two mismatches to the target sequence (with higher likelihood of off target effects) while for the DNA-transient approach 52% of the off-target sequences investigated referred to one or two mismatches. This unequal distribution of the numbers of mismatches per sequence in the analyzed samples is as well pronounced for the investigation conducted with RNP. Only 4% off-target sequences differed from the target sequence in one or two mismatches, while 96% had at least three mismatches. Due to the low comparability of included studies, no sound conclusions can be drawn based on the general overview of all studies.

Three targeted experiments were conducted by applying the same study design, differing just by the delivery method to directly compare off-target effects. Two of these articles were published by one institution and delivery methods (DNA-stable, DNA-transient, RNP, RNA) were applied and compared in hexaploid bread wheat (Liang et al., 2017, Zhang et al., 2016). The authors edited a target sequence of the TaGW2 gene, which shows identical sequences on the chromosomes B and D, whereas on chromosome A it has one mismatch at position nine proximal to the PAM. Independently of the applied delivery methods, an off-target effect occurred at this potential off-target sequence. However, Liang et al. reported that RNP showed a decreased on-target and off-target mutation frequency compared to the DNA-transient approach (RNP: Mutagenesis frequencies for TaGW2-B1 and -D1 (on-target): 33.4% and 21.8%, for TaGW2-A1 (off-target): 5.7%; DNA-transient: Mutagenesis frequencies for TaGW2-B1 and -D1 (on-target): 42.2% and 35.6%, for TaGW2-A1 (off-target): 30.8%) (Liang et al., 2017). Zhang et al. reported that delivering the system as RNA, the on- and off-target mutation frequency decreased compared to DNA-stable and DNA-transient delivery approach (RNA: Mutagenesis frequencies for TaGW2-B1 and -D1 (on-target): 1.1% and 1.1%, for TaGW2-A1 (off-target): 0.4%; DNA-stable: Mutagenesis frequencies for TaGW2-B1 and -D1 (on-target): 2.9% and 2.6%, for TaGW2-A1 (off-target): 2.0%; DNA-transient: Mutagenesis frequencies for TaGW2-B1 and -D1 (on-target): 2.9% and 3.0%, for TaGW2-A1 (off-target): 2.3%) (Zhang et al., 2016). In addition, in this study, further potential off-target sequences were predicted and analyzed. Eight potential off-target sequences with three to four mismatches to the target sequence were analyzed for the DNA-stable, DNA-transient and RNA approach but none of these showed the occurrence of off-target effects (Zhang et al., 2016). In another article, Svitashv et al (2016) conducted a DNA-stable and RNP delivery experiment by addressing the MS45 gene in rice (Svitashv et al., 2016). This site had one potential off-target sequence with two mismatches at position 12 and 20 proximal to the PAM. In both experiments off-target effects were detected, but the RNP off-target mutation rate was decreased compared to the DNA-stable delivery approach (DNA-stable: Mutagenesis frequency for MS45 (on-target): 0.34%, MS45 (off-target sequence): 0.18%; RNP: Mutagenesis frequency for MS45 (on-target): 0.69%, MS45 (off-target sequence): 0.01%).

### **Conclusion for hypothesis 5**

The data pool to assess this hypothesis for the occurrence of off-target effects when applying the DNA-stable approach is large, while for DNA-transient, RNP and RNA only a limited number of analyzed potential off-target sequences exist. Therefore, based on the available evidence it is not possible to conclude on this hypothesis. However, direct comparison of different delivery methods

indicate that the off-target rate can be reduced (but not totally avoided) by delivering the CRISPR/Cas system to the plant cell either as DNA-transient, RNA or RNP.

#### **6.5.4 Descriptive synthesis of *low* and *medium/low* as well as *unclear validity* studies**

##### **Studies of *low* and *medium/low* validity**

In total, 72 studies were rated as having *low validity* and 44 as having *medium/low validity*. The reason for rating as *low* or *medium/low validity* is summarized in Table 6.5. Detailed information about the validity assessment of all studies including the reason for the classification is provided in supplementary materials- table 5. 27 studies were rated as *low validity* as the CRISPR/Cas technique did not induce an on-target mutation. Within these studies, three off-target effects were detected in three different studies. Two of these studies had additional differences between the genotype and the reference genome that may explain off-target effects. In one study no on-target mutation was detected but a potential off-target sequence with one mismatch at position 20 proximal to the PAM showed an off-target effect. In seven studies, no reference genome has been fully sequenced or the authors remark that sequencing errors may be responsible for off-target effects. Within these studies, four off-target effects were detected in three different studies. In 15 studies, no PAM followed the analyzed potential off-target sequences. No mutations were detected when assessing potential off-target effects in these studies. In four studies, the identified so called “off-target sequences” were actually identical to the on-target sequence. In three of these studies, mutations were detected. 23 studies didn't provide any information about the analyzed potential off-target sequences, and in none of these studies off-target effects were reported. In one study, the sgRNA/Cas9 complex was extremely overexpressed but no information were provided how this was achieved. Within this study, ten potential off-target sequences were analyzed and eight off-target effects have been identified. All of these sequences had three or four mismatches to the target sequence.

Regarding studies rated as having *medium/low validity*, in 28 studies potential off-target sequences were assessed, but not all of the analyzed potential off-target sequences were followed by a PAM. Potential off-target sequences without a PAM were selected as *medium/low validity* and no off-target effects had been detected in any of these sequences. 15 studies used Whole Genome Sequencing but information about potential off-target sequences were incomplete. None of these potential off-target sequences with incomplete information were documented showing off-target effects. In seven studies, some named off-target sequences were actually identical to the on-target sequence and one off-target mutation was detected.

##### **Studies of *unclear validity***

Corresponding authors of studies that lack sufficient information on study conduct were asked to provide missing information but for 26 studies we did not get any response. In 18 studies, potential off-target sequences were analyzed but no information was provided whether a PAM followed the potential off-target sequences. Within these studies, two off-target effects were detected. One off-target effect had one mismatch at position nine proximal to the PAM and the other one had two mismatches at position eight and twenty proximal to the PAM. Eight studies were rated as *unclear validity* as the provided information about potential off-target sequences were convoluted. In five

studies, the potential off-target sequence could not be assigned to any target sequence. Another three studies focused on the first 12 nucleotides proximal to the PAM only, while no information about the nucleotides 13 to 20 have been provided. For details see supplementary materials- table 5.

## 6.6 Discussion

### 6.6.1 Factors affecting off-target effects caused by the use of CRISPR/Cas in plants

This systematic review aimed to collect and synthesize the available evidence about factors that may affect the occurrence of off-target effects caused by the application of CRISPR/Cas-system in plants. The 370 studies rated as *high* or *medium/high validity* assessed a total amount of 6416 potential off-target sequences indicating a robust evidence base. However, only 154 analyzed potential off-target sequences showed one mismatch (2.4% of the total analyzed sequences) and 218 had two mismatches (3.4% of the total analyzed sequences), while the remaining 6044 sequences (94.2%) referred to three or more mismatches. The reason for this may be that it is frequently recommended to select target sequences with at least three mismatches to other similar sequences and in many articles this criterion has been taken into account when choosing a target sequence e.g. (Collonnier et al., 2017, Martín-Pizarro et al., 2019). The fact that there are several studies in which potential off-target sequences with one or two mismatches have been investigated may have two reasons:

- i. A target sequence was deliberately chosen with other very similar sequences in the genome in order to further investigate the occurrence of off-target effects.
- ii. Polyploidic plant species were edited that have highly similar homoeoalleles with only one nucleotide mismatch. In these cases off-target effects can be a benefit by increasing the multiplexing capacity (Jacobs et al., 2017).

The evidence base was well suited to confirm hypothesis 1 that an increase number of mismatches between the on-target sequence and the potential off-target sequence steeply decreases the likelihood that off-target effects occur. Descriptive and quantitative analysis clearly indicate that there is a significant difference in the occurrence of off-target effects for one, two, three and at least four mismatches. Within this pool of data, the off-target rate (Number of detected off-target effects/number of analyzed potential off-target sequences) for one mismatch appeared to be nearly 60% of the analyzed sequences, while nearly no off-target effect occurs for a sequence with at least four mismatches to further similar sequences in the genome. These findings are in line with several articles published so far (e.g. (Doll et al., 2019, Fu et al., 2013, Hahn, Nekrasov, 2018)). The observation that a single mismatch between on- and off-target sequences often leads to off-target effects while designing a highly specific sgRNA reduces off-target effects to a minimum allows a flexible choice of sgRNA design depending on the research question. This flexibility can be used as benefit either to mutate several similar alleles in a single experiment or to design a highly specific sgRNA which possesses four or more mismatched positions to all other sequences in the genome.

Potential off-target sequences with three or more mismatches were not suitable to investigate hypotheses 2-5 in detail, as independently on the position of mismatches proximal to the PAM, the

GC-content of the protospacer, the used nuclease variants and the applied delivery methods, nearly no off-target effects occurred. Therefore, hypotheses 2-5 were descriptively investigated for one and two mismatches separately comprising a total evidence base of only 352 analyzed potential off-target sequences. Although, the data base for one and two mismatches is limited, based on the available evidence there is a tendency that off-target effects are reduced when the mismatch/es are located within the first eight nucleotides proximal to the PAM. Statistical meta-analysis also indicates that the position of the mismatch/es significantly affects the occurrence of off-target effects but less intense compared to the number of mismatches of the on-target and off-target sequences. These results are consistent with previous findings (e.g. (Endo et al., 2015, Jiang et al., 2015, Mali et al., 2013)). However, the data base is fairly limited to define properly the number of nucleotides that form the “seed sequence”; but based on the reviewed data we suggest to define it to be eight nucleotides proximal to the PAM. If the mismatch/es are located within this region off-target effects are reduced compared to when the mismatch/es are located outside the seed sequence. The data pool for one mismatch indicates that even if the mismatch is located inside the defined seed sequence, off-target effects may still occur in about one third of the potential off-target sequences which, nevertheless, is half the rate of cases compared with a mismatch at any position of the target sequence. The data pool for two mismatches further indicates that care should be taken when selecting a target sequence for genome editing with at least two potential mismatches that these are not located next to each other; but further research is needed to address this aspect more thoroughly. There was no clear indication found that the GC-content of the target sequence significantly affects the occurrence of off-target effects. In one article it was argued that the low GC-content of the target sequence (45%) is the reason that no off-target effects occurred (Yu et al., 2017), but actually the potential off-target sequences showed four and seven mismatch to the target sequence which could be sufficient to reduce the likelihood to zero according to the conclusion drawn for hypothesis 1. Regarding the nuclease variants and the delivery methods the given data base is considerably poor as the large majority of studies applied the standard nuclease SpCas9 and a cassette for the CRISPR/Cas system were stably integrated to the genome. The available evidence does not allow to conclude that altered nuclease variants or delivery methods significantly reduce the occurrence of off-target effects in general. However, a limited number of articles directly compare different delivery methods or nuclease variants indicate that such different approaches might decrease the occurrence of off-target effects but cannot completely preclude it.

In the given set of data it is likely that various factors have overlapping impacts. Either different intensities and/or insufficient data may have limited their quantitative assessment. This review highlighted the number of mismatches and the occurrence of mismatches in the seed sequence as significant impact factors determining off-target effects of the CRISPR/Cas-System in plants in a general overview of published scientific studies.

### **6.6.2 Reasons for heterogeneity**

Besides the limited evidence base to answer hypotheses 2-5 in more detail, the individual studies differ widely in design and conduct. Considering all analyzed potential off-target sequences with one or two mismatches, we identified heterogeneity between studies regarding:

- (i) The position of mismatches proximal to the PAM: Nucleotide mismatches occurred in 21 different positions proximal to the PAM.

- (ii) The GC-content of the protospacer: GC-content varied widely between 25 and 95%.
- (iii) The used nuclease variants: In total, 13 different nuclease variants were used.
- (iv) The applied delivery method: The CRISPR/Cas system was supplied to the plant cell either as DNA (stably or transiently expressed), RNA or directly as RNP, consequently leading to different amounts of protein and varying exposure times of the protein to the target DNA.

Beside to these factors that were assessed within this systematic review, additional characteristics further increased heterogeneity: In total, 43 different plant species were used for the application of CRISPR/Cas. In principle, plants with bigger genomes and higher ploidy levels have statistically more similar sequences to the target sequence in the genome than those with small genomes and a diploid set of chromosomes. In most studies a biased detection method approach was applied to identify potential off-target sequences, meaning that only one or a few similar sequences were identified a priori and only these sequences were assessed for the detection of off-target effects. However, this approach is not suitable for assessing the influence of the genome size of different plant species on off-target effects. The ploidy level may have a stronger influence on the occurrence of off-target effects as polyploidic plants may have very similar homoeoalleles with only a single nucleotide mismatch. Editing one of such sequences may increase the finding of off-target effects. However, as it is often intended to edit all homoeoalleles scientists may look for desired off-target cutting (Liang et al., 2017, Wang et al., 2014). In addition, many different bioinformatics tools and detection methods have been used to predict potential off-target sequences and to investigate them for off-target effects.

### 6.6.3 Review limitations

#### Limitations of the review methodology

The search for relevant literature was limited to German and English language, whereby all identified relevant literature was in English. Therefore, we might have missed studies published in other languages. More than half of the relevant studies were conducted in an Asian country and literature was also identified in Asian languages but excluded on title/abstract level due to language barriers. It can be assumed that a literature search in these languages would identify further relevant studies. In addition, no studies have been identified with corresponding authors from South America, although it is known that genome editing is already broadly used there. A search in Spanish or Portuguese would possibly also identify additional relevant literature. Nevertheless, a detailed evaluation of each author's role and locations which would provide more details about any international cooperation was not in the scope of this review. The full text of 49 articles that have been rated as relevant on title/abstract level were not accessible within the course of the review project and therefore they were not included in this systematic review (see supplementary materials- table 2).

#### Limitations of the evidence base

Altogether the evidence base is comprehensive with over 6400 analyzed potential off-target sequences. This was sufficient to evaluate and identify some general key factors which determine the occurrence of off-target effects. However, more than 94% of the analyzed sequences represented cases with three or more mismatches to the target sequence. Since off-target effects rarely occurred in these sequences, the evidence base was not adequate to allow clear conclusions

about all but two factors potentially affecting the occurrence of off-target effects. In addition, the available studies were very heterogeneously designed. Therefore, the assessment of the hypotheses regarding the roles of the GC-content, the nuclease variant and the delivery method were restricted and no concluding evaluation can be made for those factors. This identified evidence gap needs to be filled by systematic studies that apply the same study design varying just the specific form of one impact factor to be tested (e.g. the delivery method or the nuclease variant). For these experiments, one should use an almost identical reference sequence in several different plant species to allow a more generalizing evaluation of their potential to affect the occurrence of off-target effects. In addition, more experiments should be done designing a set of sgRNAs for a specific target each with a mutation in a different base of the whole sgRNA (similar ones have been done in e.g. (Raitskin et al., 2019, Zhang et al., 2017)). Doing so, it is for example possible to address the weight of each sgRNA position for on- and off-target activity and in parallel to evaluate the importance of proper folding of the HNH domain as well. It might come out that for a given target a less perfect sgRNA can achieve even higher on-target rates due to better HNH folding or tighter DNA:RNA binding.

## 6.7 Conclusions

### 6.7.1 Implication for policy/management

The risk of off-target effects in plants caused by genome editing approaches as by the broadly employed CRISPR/Cas system is not as critical as in clinical research. As for the latter side effects have to be excluded to protect future patients, for plants the identified off-target mutations can be segregated by outcrossing or mutants without off-target effects can be selected for further variety development (Zhang et al., 2016). Since plant breeding is the immediate context in this review, it is worth to recall the occurrence of off-target effects in natural mutations and routinely used breeding techniques such as regular crossing or undirected mutagenesis using tissue cultures, chemical mutagens or irradiation (Scientific Advice Mechanism, 2017). The natural mutation rate of *Arabidopsis thaliana* is approximately one mutation per 150,000 kilobase pairs (kbp) which means that around one mutation occurs per generation (Bartsch et al., 2018, Ossowski et al., 2010). Using chemicals like EMS or irradiation to induce mutations in plants the mutation rate increases dramatically, e.g. Jander et al. identified at least 700 mutations in EMS-mutagenized *Arabidopsis* lines (Jander et al., 2003). Another example is the regeneration of plants from cell culture (somaclonal variation). Experiments showed that somaclonal variation increases the mutation rate by a factor up to 250 compared to spontaneous mutations (Miyao et al., 2012). Compared to these techniques, off-target effects through genome-editing occur by orders of magnitude less frequently. Risk assessors and decision makers should take this aspect into account when drawing conclusions about general risks being associated with the application of genome editing in plants.

### 6.7.2 Implication for research

Around 10% of the initially identified potentially relevant studies could not be considered in this review due to lacking information about the study design and/or the off-target sequences. To enable broader analysis and evaluation the following information should be provided when analyzing off-target effects:

- (i) The reference on-target sequence.
- (ii) The prediction tools and detection methods used.
- (iii) and for the potential off-target sequences:
  - a. The number of identified potential off-target sequences.
  - b. The sequences of the potential off-target sites plus the PAM (and therewith the number and the position of the mismatches proximal to the PAM).

The results of the review show that the occurrence of off-target effects prominently depends on the number of mismatches to other similar sequences and the position of the mismatch/es proximal to the PAM. So far, only a few studies applied an identical study design by systematically varying the modifications of one impact factor, in order to examine the impacts on the occurrence of off-target effects. More comparative studies are necessary to provide oversight of a general impact pattern to guide further application in research and development.

In order to minimize off-target effects a priori, it is recommended from the analysis of hypothesis 1 and 2 that a target sequence is chosen which differs in at least four disjunct positions from similar genomic sequences. This can reduce further crossing or selection efforts to support thorough investigations of e.g. gene functions or variety development.

#### **Additional files**

<https://www.frontiersin.org/articles/10.3389/fpls.2020.574959/full#supplementary-material>

## **6.8 References**

Altschul, S. Basic Local Alignment Search Tool. *Journal of molecular biology* 1990, 215 (3), 403–410. DOI: 10.1006/jmbi.1990.9999.

Bartsch, D., Bendiek, J., Braeuning, A., Dagand, E., Duensing, N., Fladung, M., C Franz, E Groeneveld, L Grohmann, D Habermann, G Leggewie, A Matthies, U Middelhoff, H Niemann, B Petersen, A Scheepers, W Schenkel, T Sprink, A Stolz, C Tebbe, R Wilhelm. Wissenschaftlicher Bericht zu den neuen Techniken in der Pflanzenzüchtung und der Tierzucht und ihren Verwendungen im Bereich der Ernährung und Landwirtschaft. [https://www.bmel.de/SharedDocs/Downloads/DE/\\_Landwirtschaft/Gruene-Gentechnik/Bericht\\_Neue\\_Zuechtungstechniken.html](https://www.bmel.de/SharedDocs/Downloads/DE/_Landwirtschaft/Gruene-Gentechnik/Bericht_Neue_Zuechtungstechniken.html) (accessed February 23, 2018).

Bertier, L. D.; Ron, M.; Huo, H.; Bradford, K. J.; Britt, A. B.; Michelmore, R. W. High-Resolution Analysis of the Efficiency, Heritability, and Editing Outcomes of CRISPR/Cas9-Induced Modifications of NCED4 in Lettuce (*Lactuca sativa*). *G3* (Bethesda, Md.) 2018, 8 (5), 1513–1521. DOI: 10.1534/g3.117.300396.

Chen, J. S.; Dagdas, Y. S.; Kleinstiver, B. P.; Welch, M. M.; Harrington, L. B.; Sternberg, S. H.; Joung, J. K.; Yildiz, A.; Doudna, J. A. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy, 2017.

Collaboration for Environmental Evidence. Guidelines and Standards for Evidence synthesis in Environmental Management. Version 5.0 (AS Pullin, GK Frampton, B Livoreil & G Petrokofsky,

Eds). <http://www.environmentalevidence.org/guidelines/table-of-contents> (accessed November 3, 2019).

Collonnier, C.; Epert, A.; Mara, K.; Maclot, F.; Guyon-Debast, A.; Charlot, F.; White, C.; Schaefer, D. G.; Nogu e, F. CRISPR-Cas9-mediated efficient directed mutagenesis and RAD51-dependent and RAD51-independent gene targeting in the moss *Physcomitrella patens*. *Plant biotechnology journal* 2017, 15 (1), 122–131. DOI: 10.1111/pbi.12596.

Dagd as, Y. S.; Chen, J. S.; Sternberg, S. H.; Doudna, J. A.; Yildiz, A. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Science Advance* [Online] 2017.

Doll, N. M.; Gilles, L. M.; G erentes, M. F.; Richard, C.; Just, J.; Fierlej, Y.; Borrelli, V. M. G.; Gendrot, G.; Ingram, G. C.; Rogowsky, P. M.; Widiez, T. Single and multiple gene knockouts by CRISPR–Cas9 in maize. *Plant cell reports* 2019. DOI: 10.1007/s00299-019-02378-1.

D onmez, D.;  ımsek  ; Aka Kacar, Y. Genetic Engineering Techniques in Fruit Science. *IJOEAR* 2016, 2 (12), 115–128.

Eckerstorfer, M. F.; Dolezel, M.; Heissenberger, A.; Miklau, M.; Reichenbecher, W.; Steinbrecher, R. A.; Wa mann, F. An EU Perspective on Biosafety Considerations for Plants Developed by Genome Editing and Other New Genetic Modification Techniques (nGMs). *Frontiers in bioengineering and biotechnology* 2019a, 7, 31. DOI: 10.3389/fbioe.2019.00031.

Eckerstorfer, M. F.; Engelhard, M.; Heissenberger, A.; Simon, S.; Teichmann, H. Plants Developed by New Genetic Modification Techniques-Comparison of Existing Regulatory Frameworks in the EU and Non-EU Countries. *Frontiers in bioengineering and biotechnology* 2019b, 7, 26. DOI: 10.3389/fbioe.2019.00026.

Endo, M.; Mikami, M.; Endo, A.; Kaya, H.; Itoh, T.; Nishimasu, H.; Nureki, O.; Toki, S. Genome editing in plants by engineered CRISPR–Cas9 recognizing NG PAM. *NPLANTS* [Online] 2019, 5 (1), 14–17. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85058157282&doi=10.1038%2fs41477-018-0321-8&partnerID=40&md5=c7a8b99d252449123a682dd5b526828a>.

Endo, M.; Mikami, M.; Toki, S. Multigene knockout utilizing off-target mutations of the CRISPR/Cas9 system in rice. *Plant & cell physiology* 2015, 56 (1), 41–47. DOI: 10.1093/pcp/pcu154.

Fu, Y.; Foden, J. A.; Khayter, C.; Maeder, M. L.; Reyon, D.; Joung, J. K.; Sander, J. D. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology* 2013, 31 (9), 822–826. DOI: 10.1038/nbt.2623.

Haddaway, N. R.; Macura, B.; Whaley, P.; Pullin, A. S. ROSES RepOrting standards for Systematic Evidence Syntheses: pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. *Environ Evid* 2018, 7 (1), 409. DOI: 10.1186/s13750-018-0121-7.

Hahn, F.; Nekrasov, V. CRISPR/Cas precision: do we need to worry about off-targeting in plants? *Plant cell reports* [Online] 2018.

Hinz, J. M.; Laughery, M. F.; Wyrick, J. J. Nucleosomes Inhibit Cas9 Endonuclease Activity in Vitro. *Biochemistry* 2015, 54 (48), 7063–7066. DOI: 10.1021/acs.biochem.5b01108.

- Hsu, P. D.; Scott, D. A.; Weinstein, J. A.; Ran, F. A.; Konermann, S.; Agarwala, V.; Li, Y.; Fine, E. J.; Wu, X.; Shalem, O.; Cradick, T. J.; Marraffini, L. A.; Bao, G.; Zhang, F. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology* 2013, 31 (9), 827–832. DOI: 10.1038/nbt.2647.
- Jacobs, T. B.; LaFayette, P. R.; Schmitz, R. J.; Parrott, W. A. Targeted genome modifications in soybean with CRISPR/Cas9. *BMC biotechnology* 2015, 15, 1–10. DOI: 10.1186/s12896-015-0131-2.
- Jacobs, T. B.; Zhang, N.; Patel, D.; Martin, G. B. Generation of a Collection of Mutant Tomato Lines Using Pooled CRISPR Libraries. *Plant physiology* 2017, 174 (4), 2023–2037. DOI: 10.1104/pp.17.00489.
- Jander, G.; Baerson, S. R.; Hudak, J. A.; Gonzalez, K. A.; Gruys, K. J.; Last, R. L. Ethylmethanesulfonate saturation mutagenesis in Arabidopsis to determine frequency of herbicide resistance. *Plant physiology* 2003, 131 (1), 139–146. DOI: 10.1104/pp.102.010397.
- Jansing, J.; Schiermeyer, A.; Schillberg, S.; Fischer, R.; Bortesi, L. Genome Editing in Agriculture: Technical and Practical Considerations. *International journal of molecular sciences* 2019, 20 (12). DOI: 10.3390/ijms20122888.
- Ji, X.; Si, X.; Zhang, Y.; Zhang, H.; Zhang, F.; Gao, C. Conferring DNA virus resistance with high specificity in plants using virus-inducible genome-editing system. *Genome biology* [Online] 2018, 19 (1), 197. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85056645050&doi=10.1186%2fs13059-018-1580-4&partnerID=40&md5=5533e4be8bb078c691db4fef82df05bc>.
- Jiang, F.; Zhou, K.; Ma, L.; Gressel, S.; Doudna, J. A. STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science (New York, N.Y.)* 2015, 348 (6242), 1477–1481. DOI: 10.1126/science.aab1452.
- Jiang, W. Z.; Henry, I. M.; Lynagh, P. G.; Comai, L.; Cahoon, E. B.; Weeks, D. P. Significant enhancement of fatty acid composition in seeds of the allohexaploid, *Camelina sativa*, using CRISPR/Cas9 gene editing. *Plant biotechnology journal* 2017, 15 (5), 648–657. DOI: 10.1111/pbi.12663.
- JMP® 2019. JMP®. Version 14. SAS Institute Inc. Cary, NC.
- Kaya, H.; Mikami, M.; Endo, A.; Endo, M.; Toki, S. Highly specific targeted mutagenesis in plants using *Staphylococcus aureus* Cas9. *Scientific reports* 2016, 6, 1–9. DOI: 10.1038/srep26871.
- Kohl, C.; McIntosh, E. J.; Unger, S.; Haddaway, N. R.; Kecke, S.; Schiemann, J.; Wilhelm, R. Online tools supporting the conduct and reporting of systematic reviews and systematic maps: a case study on CADIMA and review of existing tools. *Environ Evid* 2018, 7 (1), 2420. DOI: 10.1186/s13750-018-0115-5.
- LeBlanc, C.; Zhang, F.; Mendez, J.; Lozano, Y.; Chatpar, K.; Irish, V. F.; Jacob, Y. Increased efficiency of targeted mutagenesis by CRISPR/Cas9 in plants using heat stress. *The Plant Journal* 2018, 93 (2), 377–386. DOI: 10.1111/tbj.13782.
- Lee, J.; Chung, J.-H.; Kim, H. M.; Kim, D.-W.; Kim, H. Designed nucleases for targeted genome editing. *Plant biotechnology journal* 2016, 14 (2), 448–462. DOI: 10.1111/pbi.12465.

- Lei, Y.; Lu, L.; Liu, H.-Y.; Li, S.; Xing, F.; Chen, L.-L. CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. *Molecular plant* 2014, 7 (9), 1494–1496. DOI: 10.1093/mp/ssu044.
- Liang, Z.; Chen, K.; Li, T.; Zhang, Y.; Wang, Y.; Zhao, Q.; Liu, J.; Zhang, H.; Liu, C.; Ran, Y.; Gao, C. Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. *Nature communications* 2017, 8, 14261. DOI: 10.1038/ncomms14261.
- Liu, G.; Yin, K.; Zhang, Q.; Gao, C.; Qiu, J.-L. Modulating chromatin accessibility by transactivation and targeting proximal dsgRNAs enhances Cas9 editing efficiency in vivo. *Genome biology* 2019, 20 (1), 145. DOI: 10.1186/s13059-019-1762-8.
- Liu, H.; Ding, Y.; Zhou, Y.; Jin, W.; Xie, K.; Chen, L.-L. CRISPR-P 2.0: An Improved CRISPR-Cas9 Tool for Genome Editing in Plants. *Molecular plant* 2017, 10 (3), 530–532. DOI: 10.1016/j.molp.2017.01.003.
- Mali, P.; Aach, J.; Stranges, P. B.; Esvelt, K. M.; Moosburner, M.; Kosuri, S.; Yang, L.; Church, G. M. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature biotechnology* 2013, 31 (9), 833–838. DOI: 10.1038/nbt.2675.
- Martin, F.; Sánchez-Hernández, S.; Gutiérrez-Guerrero, A.; Pinedo-Gomez, J.; Benabdellah, K. Biased and Unbiased Methods for the Detection of Off-Target Cleavage by CRISPR/Cas9: An Overview. *International journal of molecular sciences* 2016, 17 (9). DOI: 10.3390/ijms17091507.
- Martín-Pizarro, C.; Triviño, J. C.; Posé, D. Functional analysis of the TM6 MADS-box gene in the octoploid strawberry by CRISPR/Cas9-directed mutagenesis. *Journal of experimental botany* [Online] 2019, 70 (3), 949–961. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85061140063&doi=10.1093%2fjxb%2fery400&partnerID=40&md5=eb9025f2ba4300e25db04f450b2ad133>.
- Metje-Sprink, J.; Menz, J.; Modrzejewski, D.; Sprink, T. DNA-Free Genome Editing: Past, Present and Future. *Frontiers in plant science* 2018, 9, 2–9. DOI: 10.3389/fpls.2018.01957.
- Mitchell, B. P.; Hsu, R. V.; Medrano, M. A.; Zewde, N. T.; Narkhede, Y. B.; Palermo, G. Spontaneous Embedding of DNA Mismatches Within the RNA:DNA Hybrid of CRISPR-Cas9. *Frontiers in molecular biosciences* 2020, 7, 39. DOI: 10.3389/fmolb.2020.00039.
- Miyao, A.; Nakagome, M.; Ohnuma, T.; Yamagata, H.; Kanamori, H.; Katayose, Y.; Takahashi, A.; Matsumoto, T.; Hirochika, H. Molecular spectrum of somaclonal variation in regenerated rice revealed by whole-genome sequencing. *Plant & cell physiology* 2012, 53 (1), 256–264. DOI: 10.1093/pcp/pcr172.
- Modrzejewski, D.; Hartung, F.; Sprink, T.; Krause, D.; Kohl, C.; Wilhelm, R. What is the available evidence for the range of applications of genome-editing as a new tool for plant trait modification and the potential occurrence of associated off-target effects: a systematic map. *Environ Evid* 2019, 8 (1), 11. DOI: 10.1186/s13750-019-0171-5.
- Montague, T. G.; Cruz, J. M.; Gagnon, J. A.; Church, G. M.; Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic acids research* 2014, 42 (Web Server issue), W401-7. DOI: 10.1093/nar/gku410.

- Ossowski, S.; Schneeberger, K.; Lucas-Lledó, J. I.; Warthmann, N.; Clark, R. M.; Shaw, R. G.; Weigel, D.; Lynch, M. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* (New York, N.Y.) 2010, 327 (5961), 92–94. DOI: 10.1126/science.1180677.
- Puchta, H.; Fauser, F. Synthetic nucleases for genome engineering in plants: Prospects for a bright future. *The Plant journal : for cell and molecular biology* 2014, 78 (5), 727–741. DOI: 10.1111/tpj.12338.
- Raitskin, O.; Schudoma, C.; West, A.; Patron, N. J. Comparison of efficiency and specificity of CRISPR-associated (Cas) nucleases in plants: An expanded toolkit for precision genome engineering. *PloS one* [Online] 2019, 14 (2). <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85062174701&doi=10.1371%2fjournal.pone.0211598&partnerID=40&md5=668a86fcbd1fa8d5cd735fbee01ecbe9>.
- Ricci, C. G.; Chen, J. S.; Miao, Y.; Jinek, M.; Doudna, J. A.; McCammon, J. A.; Palermo, G. Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics. *ACS central science* 2019, 5 (4), 651–662. DOI: 10.1021/acscentsci.9b00020.
- Russo, M. T.; Aiese Cigliano, R.; Sanseverino, W.; Ferrante, M. I. Assessment of genomic changes in a CRISPR/Cas9 *Phaeodactylum tricornutum* mutant through whole genome resequencing. *PeerJ* [Online] 2018, 6. <Go to ISI>://WOS:000447208100001.
- SAS Institute. 2018. JMP® 14 Fitting Linear Models, Chu ban: Cary, NC: SAS Institute Inc.
- Scientific Advice Mechanism (SAM). New techniques in Agricultural Biotechnology. High Level Group of Scientific Advisors. <https://ec.europa.eu/research/sam/index.cfm?pg=agribiotechnology#> (accessed 18.07.2017).
- Sprink, T.; Eriksson, D.; Schiemann, J.; Hartung, F. Regulatory hurdles for genome editing: Process- vs. product-based approaches in different regulatory contexts. *Plant cell reports* 2016, 35 (7), 1493–1506. DOI: 10.1007/s00299-016-1990-2.
- Stemmer, M.; Thumberger, T.; Del Sol Keyer, M.; Wittbrodt, J.; Mateo, J. L. CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PloS one* 2015, 10 (4), e0124633. DOI: 10.1371/journal.pone.0124633.
- Svitashev, S.; Schwartz, C.; Lenderts, B.; Young, J. K.; Mark Cigan, A. Genome editing in maize directed by CRISPR-Cas9 ribonucleoprotein complexes. *Nature communications* 2016, 7, 1–7. DOI: 10.1038/ncomms13274.
- Tycko, J.; Myer, V. E.; Hsu, P. D. Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity. *Molecular cell* 2016, 63 (3), 355–370. DOI: 10.1016/j.molcel.2016.07.004.
- Walsh, Susan. Binary Logistic Regression – What, When, and How.
- Wang, Y.; Cheng, X.; Shan, Q.; Zhang, Y.; Liu, J.; Gao, C.; Qiu, J.-L. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nature biotechnology* 2014, 32 (9), 947–951. DOI: 10.1038/nbt.2969.
- Woo, J. W.; Kim, J.; Kwon, S. I.; Corvalán, C.; Cho, S. W.; Kim, H.; Kim, S.-G.; Kim, S.-T.; Choe, S.; Kim, J.-S. DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nature biotechnology* 2015, 33 (11), 1162–1164. DOI: 10.1038/nbt.3389.

- Wu, X.; Kriz, A. J.; Sharp, P. A. Target specificity of the CRISPR-Cas9 system. *Quantitative biology (Beijing, China)* 2014, 2 (2), 59–70. DOI: 10.1007/s40484-014-0030-x.
- Yu, Q.-h.; Wang, B.; Li, N.; Tang, Y.; Yang, S.; Yang, T.; Xu, J.; Guo, C.; Yan, P.; Wang, Q.; Asmutola, P. CRISPR/Cas9-induced Targeted Mutagenesis and Gene Replacement to Generate Long-shelf Life Tomato Lines. *Scientific reports* 2017, 7 (1), 818. DOI: 10.1038/s41598-017-12262-1.
- Zhang, D.; Zhang, H.; Li, T.; Chen, K.; Qiu, J.-L.; Gao, C. Perfectly matched 20-nucleotide guide RNA sequences enable robust genome editing using high-fidelity SpCas9 nucleases. *Genome biology* 2017, 18 (1), 191. DOI: 10.1186/s13059-017-1325-9.
- Zhang, Y.; Liang, Z.; Zong, Y.; Wang, Y.; Liu, J.; Chen, K.; Qiu, J.-L.; Gao, C. Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nature communications* 2016, 7, 1–8. DOI: 10.1038/ncomms12617.
- Zhao, H.; Wolt, J. D. Risk associated with off-target plant genome editing and methods for its limitation. *Emerg. Top. Life Sci.* 2017, 1 (2), 231–240. DOI: 10.1042/ETLS20170037.
- Zhu, C.; Bortesi, L.; Baysal, C.; Twyman, R. M.; Fischer, R.; Capell, T.; Schillberg, S.; Christou, P. Characteristics of Genome Editing Mutations in Cereal Crops. *Trends in plant science* 2017, 22 (1), 38–52. DOI: 10.1016/j.tplants.2016.08.009.
- Zischewski, J.; Fischer, R.; Bortesi, L. Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnology advances* 2017, 35 (1), 95–104. DOI: 10.1016/j.biotechadv.2016.12.003.

## 7 Summary and outlook

Genome editing techniques enable a targeted modification of a DNA sequence in a site-directed manner. They are a promising tool to speed up the breeding process and to develop new plant varieties with beneficial traits for the agricultural value chain. Nevertheless, there is still an ongoing and controversial debate about the benefits of genome editing for agriculture and the extent to which off-target effects occur and matter. The overall objective of this study was to systematically collect and evaluate the available evidence on the application and impact of genome editing in plants regarding the traits modified and the potential occurrence of associated off-target effects. To achieve this, a multi-stage approach including a systematic map protocol - that developed a research strategy for the exploration of literature data, a systematic map with two secondary review questions - that provided an systematic overview on the existing evidence on genome editing applications in plants and a systematic review that build up on the results of the map was applied - and that provides an in depth analysis of the state of knowledge about factors governing off-target effects caused by CRISPR/Cas applications in plants.

### 7.1 Evidence synthesis tools for evidence-based decision-making

This thesis aimed to explore the rapidly expanding evidence base of genome editing in plants by applying a multi-stage approach. As a first step, a systematic map protocol was published as an a priori guide to reduce reviewer bias in further steps of data acquisition and exploration. In addition, it aimed to make the review process more rigorous, well-defined and transparent in order to inform interested stakeholders and future users. Based on the systematic map protocol, the systematic map summarized the available evidence base of applications of genome editing in plants. It surveyed the status of research, the progress in genome editing in plants and the occurrence of associated off-target effects. This systematic map encourages an informed public debate and provides fact-based input for policy makers, interested stakeholders and a broader public:

- The systematic map protocol and the systematic map published have already been downloaded more than 9000 times and were cited as a reference in 22 peer-review articles.
- The results of the systematic map regarding market-oriented traits are of great interest for decision-makers, including the Federal Ministry for Food and Agriculture (BMEL). In addition to the peer-review published map, a summary of the results regarding market-oriented applications for the period until March 2017 was provided on the webpage of the BMEL. Furthermore, on behalf of the BMEL two updates for the period until May 2018 and June 2019 were performed based on the systematic map protocol to identify market-oriented applications published after the original publication [68–70].
- A repository was established on the project webpage ([www.dialog-gea.de](http://www.dialog-gea.de)) aiming to provide a searchable database on relevant articles from journals, company web pages and web pages of governmental agencies about applications of genome editing in plants [71]

- The results of the systematic map were used in several statements of prestigious national and international institutions, including the German National Academy of Sciences Leopoldina and the European Scientific Advice Mechanism (SAM), which supports the European Commission with high quality and independent scientific advice for its policy-making activities [72, 73].
- The impartial data analysis regarding applications of genome editing and risk assessment is acknowledged as both proponents of genome editing but also institutions with critical attitudes towards genome editing already used the published protocol and the systematic map as reference in their publications [74, 75].
- To inform interested stakeholder communities the results of the systematic map were presented on several events and conferences, including a side event of the World Trade Organization and an ELSA-GEA stakeholder workshop.
- Besides supporting decision makers and scientists, the systematic map aimed to support an informed public debate within society. To achieve this the results of the map were used to develop a CRISPR advent calendar and a genome editing library [76, 77]. In this calendar and library several market-oriented applications of genome editing were described by outlining challenges facing agriculture and the contributions of genome editing toward tackling them. The advent calendar was published in ten languages including German, English, French, Italian, Dutch, Greek, Spanish, Portuguese, Czech and Hungarian. So far, the calendar and the library have been called up more than 15.000 times all over the world.

Furthermore, this work demonstrates that a systematic map is a suitable tool to identify relevant topics for a more detailed analysis in the form of a systematic review. A Result of the systematic map was the identification of a dedicated knowledge cluster of publications considering the analysis of off-target effects caused by CRISPR/Cas applications in plants, while other genome editing tools were less well represented. In addition, it was shown that the data basis within pool of CRISPR/Cas studies varies broadly, as studies are very heterogeneous in their structure and design with regard to factors that potentially affect the occurrence of off-target effects. This essentially includes factors like the number of mismatches between the on-target and potential off-target sequence, the position of the mismatch/es proximal to the PAM, the applied nuclease variant, the delivery methods of the CRISPR/Cas-System to the plant cell and the GC-content of the target sequence. All these parameters have been described to potentially affect the occurrence of off-target effects. However, so far it was not systematically analyzed what is the external validity of such studies and reviews and which factors do evidently dominate the occurrence of off-target effects in general. Therefore, based on the findings of the systematic map the systematic review aimed to identify and systematically analyze these factors. A systematic review has advantages regarding transparency, impact, validity and causality compared to a traditional literature review:

- Detailed information about the methods used to perform this systematic review are presented in the methods section of the published systematic review. The high reporting standards ensure transparency and traceability in order to maximize objectivity and to minimize bias.

- The literature search was conducted based on the systematic map protocol, including a scoping search that was carried out in order to test the search strategy and to identify as many relevant articles as possible.
- In order to be included in the systematic review, each article had to meet pre-defined and documented eligibility criteria regarding the population, intervention and outcome. Excluded articles were documented with the reason for exclusion.
- A critical appraisal was performed to assess internal validity (risk of bias within a specific study) and external validity (generalizability) and hence to reflect critical variables that affect the reliability of study outcomes. Within this review, almost 20% of the studies were assessed as of *low* or *unclear validity* and were handled separately from studies rated as of *high validity*. In a traditional literature review, these studies could have been considered together with all others and would have distorted the results.
- Results of the systematic review were fully reported and care was taken to avoid reaching conclusions beyond the evidence base.

The results demonstrate the advantages of these evidence synthesis tools compared to traditional reviews. However, conducting this kind of synthesis is very time consuming and resource intensive. In addition, the number of publications about genome editing in plants has risen significantly in recent years. Therefore, the systematic review of this thesis slightly differs from the high level guidelines of the CEE [64] and has methodological restrictions:

- The systematic review protocol and the finalized systematic review were published in one publication. No peer-review process took place for the methodological part of the review before starting the systematic review.
- In general, the implementation of a Systematic Review is based on the 4-eyes principle to reduce human error bias. However, due to time and cost reasons, parts of the review such as data extraction and critical appraisal were performed by one reviewer only.
- Stakeholder involvement is of crucial importance when conducting a systematic review. In this thesis, the results of the systematic map were presented and discussed at several meetings and stakeholder conferences. Feedback gained within these events was taken into account when planning and conducting the Systematic Review. However, no further stakeholder exchange took place during the implementation of the Review.

## **7.2 Summary and outlook of the results of the systematic map and the systematic review**

One objective of this thesis was to provide a general overview of the application of genome editing in plants. Systematic mapping of the available evidence indicates that the publication rate of primary research was quite low between 1996 and 2012 and the number of studies using Zinc-finger nucleases, Meganucleases, Oligonucleotide-directed mutagenesis, TALENs and Base editing has remained on a low level until today. Since 2013, when the CRISPR/Cas technique was first used in plants, the number of studies has risen significantly. Today, it is the dominant genome editing technique and much effort is invested in examining other CRISPR nuclease variants and in

developing improved and modified ones. The map further demonstrates that the majority of recent studies describe the induction of point mutations, mainly achieved with SDN1, meaning that no template is added and the DSB was repaired by NHEJ. This sequence alteration is comparable to spontaneous mutations or undirected mutagenesis using chemicals or irradiation. Only in a subset of studies was a repair template added. This template was either similar to the target sequence except for one or a few nucleotides (SDN2) or harbored a recombinant DNA sequence (SDN3) and the DSB was repaired via HDR. The latter approach results in more complex alterations like the insertion of entire genes from both closely as well as distantly related organisms. This is comparable to traditional genetic engineering but the insertion is site-specifically targeted. In total, studies from 33 different countries were identified. About three quarter of these studies originate from either China (40%) or the USA (33%), while only 13% of the studies originate from all European countries. This indicates that Europe already lags behind the leading countries with regard to research and development efforts in genome editing in plants. Regarding the specific regulatory provisions throughout the world, it can be assumed that this gap will widen in future as there is no global harmonization of regulatory oversight [78]. Several countries, including Argentina, Brazil, Colombia, Chile, USA, Japan, Russia and Australia, have adapted legislations to the genome editing techniques or have released guidelines to enable the use of genome editing. Compared to this, the Court of Justice of the European Union (CJEU) decided in 2018 that organisms obtained by mutagenesis including genome editing are genetically modified organisms (GMO) within the meaning of the Directive 2001/18/EC. The consequence of this ruling is that genome edited plants have to pass through a comprehensive and time consuming approval process associated with cost- and labor-intensive pre-market evaluation regardless whether only point mutations were introduced (SDN1) or foreign genes were transferred (SDN3). It is feared that the current EU regulation will negatively affect investment, research and innovation efforts and make it much more difficult to commercialize genome edited products in Europe. Therefore, plant breeders, business associations from the agri food value chain and scientists are calling the EU Commission to alter the regulation of genome editing in a way that it is scientifically based and more differentiated [72, 73, 79].

### **7.2.1 Market-oriented applications of genome editing**

Another objective of this thesis was to provide an overview of the traits modified by genome editing in plants in order to inform policy makers and interested stakeholder communities. The large majority of studies were classified as proof of concept studies and basic research. Nevertheless, in the period until May 2018, almost 100 different applications aimed to produce beneficial traits in 28 different agricultural crops and were defined as “market-oriented”. Most of these applications were carried out in cash crops such as maize, rice, wheat and soybean, but economically less important crops such as cucumber, lettuce, peanut or grapefruit were also worked on. In June 2019, an update was performed based on the systematic map protocol on behalf of the BMEL. Results show that within 14 months the number of market-oriented applications had increased by around 50% and further plant species like barley, banana or kiwifruit had been genome edited. The wide range of market-oriented applications address broad demands in the agricultural value chain. Breeding objectives include yield improvement, improved growth characteristics, improved food and feed quality, tolerances to biotic and abiotic stress, herbicide tolerance and industrial utilization.

So far, most of the recent marked-oriented applications have addressed one specific DNA sequence in the genome resulting in a single point mutation. However, CRISPR/Cas is able to induce several specific mutations in polyploidic plant species that have highly similar homoeoalleles. For example, the simultaneous editing of three homoeoalleles in hexaploid wheat lines was successfully performed [35]. More recently, genome editing strategies have been developed aiming to combine several agronomically interesting traits for de novo domestication of wild varieties by applying multiplexing. For example, Zsögen et al. (2018) edited six loci in the genome of wild tomato *Solanum pimpinellifolium* resulting in compact plants with an increased fruit size, an increased fruit number and an improved lycopene accumulation [80]. It can be assumed that multiplexing and de novo domestication will be the next milestone in plant breeding.

### 7.2.2 Occurrence of off-target effects

Regarding the occurrence of off-target effects, the objectives of this thesis were to provide an overview of the available evidence about the occurrence of off-target effects in plants and to assess whether the available evidence base is suitable for an in-depth analysis by a systematic review. Based on the results of the systematic map, the systematic review aimed to identify factors that affect the occurrence of off-target effects caused by the application of CRISPR/Cas in plants. Systematic mapping of the available evidence identified a total of 252 relevant studies from 161 articles. The large majority of studies (90%) were performed using the CRISPR/Cas technique, whereas only a few studies used TALENs, BE, ZFN, MN or ODM. In most CRISPR/Cas studies, biased detection methods were used, meaning that potential off-target sequences were first predicted using bioinformatics programs followed by targeted sequencing of the off-target sequence to track the occurrence of off-target effects. Off-target effects were detected in several studies and it was shown that individual studies are very heterogeneous in their structure and design. Heterogeneity was identified regarding the plant species, the nuclease variant, the prediction tools and detection methods used, the amount of tolerated mismatches and the chosen sgRNA. Based on these results another objective of this thesis was to identify and systematically analyze factors that may affect the occurrence of off-target effects caused by the application of CRISPR/Cas in plants. The systematic review team focused on five factors which are described in literature as probably affecting the occurrence of off-target effects: (i) The number of mismatches between the on-target and potential off-target sequence (ii) The position of mismatches proximal to the PAM (iii) The GC-content of the targeting sequence (iv) Altered nuclease variants (v) The time of incubation of the CRISPR construct. The literature available clearly shows that an increase number of mismatches between the on-target and potential off-target sequence significantly decreases the likelihood that off-target effects occur. The observed off-target rate decreases from 59% when there is one mismatch between the on-target and potential off-target sequences towards 0% when four or more mismatches exist. This indicates that in future research projects one may use this flexibility as benefit to mutate several similar alleles in a single experiment (e.g. in polyploidic plant species) or to design a very specific sgRNA which possesses at least four mismatched positions to further sequences in order to significantly decrease the likelihood for the occurrence of off-target effects. In addition, the evidence base indicates that off-target effects can be reduced significantly when the mismatch/es are located within the first eight nucleotides proximal to the PAM and there is also a tendency that adjoined mismatches increase the likelihood of off-target effects. There is no evidence that the occurrence of off-target effects shows an obvious

trend in relation to the GC-content of the chosen protospacer sequence. The data base regarding the impact of the nuclease variant and the delivery method is very poor as the majority of studies applied the standard nuclease SpCas9 and the CRISPR/Cas system was stably delivered in the genome. Hence, a general significant impact of the nuclease variants and the delivery method on the occurrence of off-target effects cannot be proved. However, the evidence base is too limited to allow any quantitative conclusions. If various factors affect the occurrence of off-target effects, it is likely that factors overlap and thus less dominant factors may not show a significant impact. The results further demonstrate that there is a great heterogeneity between studies. Only a small number of studies applied an identical study design by just varying one factor step by step, in order to examine the influence of a single factor on off-target effects. Future research should focus on this aspect when exploring off-target effects. Finally, from a risk assessment point of view off-target effects occurring due to the use of CRISPR/Cas should be assessed in the context of naturally occurring mutations and breeding techniques routinely used including undirected mutagenesis techniques or somaclonal variation, in which the number of off-target mutations is far higher compared to CRISPR/Cas applications. Nevertheless, for safety reasons as well as to minimize additional selection efforts in breeding programs it is recommended to minimize the likelihood of off-target effects when planning by considering the major impact factors. Bioinformatics tools provide needed information cost efficiently.

### 7.3 References (Chapter 1, chapter 7)

1. Becker H. Pflanzenzüchtung. 2nd ed. Stuttgart: UTB; 2011.
2. FAO - Committee on Agriculture (17th Session). 06.03.2003. <http://www.fao.org/docrep/meeting/006/y8704e.htm>. Accessed 25 Sep 2018.
3. Intergovernmental Panel on Climate Change (IPCC). IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse gas fluxes in Terrestrial Ecosystems. 2019. <https://www.ipcc.ch/srccl/>. Accessed 4 Feb 2020.
4. Ronald P. Plant genetics, sustainable agriculture and global food security. *Genetics*. 2011;188:11–20. doi:10.1534/genetics.111.128553.
5. Sprink T, Eriksson D, Schiemann J, Hartung F. Regulatory hurdles for genome editing: Process- vs. product-based approaches in different regulatory contexts. *Plant Cell Rep*. 2016;35:1493–506. doi:10.1007/s00299-016-1990-2.
6. Jorasch P. How Much Plant Breeding Innovation is Politically Intended? 2016. <https://european-seed.com/2016/05/much-plant-breeding-innovation-politically-intended/>. Accessed 5 Feb 2020.
7. Songstad DD, Petolino JF, Voytas DF, Reichert NA. Genome Editing of Plants. *Critical Reviews in Plant Sciences*. 2017;36:1–23. doi:10.1080/07352689.2017.1281663.
8. Barton KA, Binns AN, Matzke AJM, Chilton M-D. Regeneration of intact tobacco plants containing full length copies of genetically engineered T-DNA, and transmission of T-DNA to R1 progeny. *Cell*. 1983;32:1033–43. doi:10.1016/0092-8674(83)90288-X.
9. Herrera-Estrella L, Depicker A, van Montagu M, Schell J. Expression of chimaeric genes transferred into plant cells using a Ti-plasmid-derived vector. *Nature*. 1983;303:209–13. doi:10.1038/303209a0.

10. Govindan G, Ramalingam S. Programmable Site-Specific Nucleases for Targeted Genome Engineering in Higher Eukaryotes. *J Cell Physiol.* 2016;231:2380–92. doi:10.1002/jcp.25367.
11. Kim Y-GG, Cha J, Chandrasegaran S. Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A.* 1996;93:1156–60.
12. Scientific Advice Mechanism (SAM). New techniques in Agricultural Biotechnology. 2017. <https://ec.europa.eu/research/sam/index.cfm?pg=agribiotecnologia>. Accessed 18.07.2017.
13. European Food Safety Authority. Scientific opinion addressing the safety assessment of plants developed using Zinc Finger Nuclease 3 and other Site-Directed Nucleases with similar function: EFSA Panel on Genetically Modified Organisms (GMO). *EFSA Journal.* 2012;10:2943. doi:10.2903/j.efsa.2012.2943.
14. van de Wiel CCM, Schaart JG, Lotz LAP, Smulders MJM. New traits in crops produced by genome editing techniques based on deletions. *Plant Biotechnol Rep.* 2017;11:1–8. doi:10.1007/s11816-017-0425-z.
15. Podevin N, Davies HV, Hartung F, Nogué F, Casacuberta JM. Site-directed nucleases: A paradigm shift in predictable, knowledge-based plant breeding. *Trends Biotechnol.* 2013;31:375–83. doi:10.1016/j.tibtech.2013.03.004.
16. Puchta H, Fauser F. Synthetic nucleases for genome engineering in plants: Prospects for a bright future. *Plant J.* 2014;78:727–41. doi:10.1111/tbj.12338.
17. Rinaldo AR, Ayliffe M. Gene targeting and editing in crop plants: A new era of precision opportunities. *Mol Breeding.* 2015;35:1126. doi:10.1007/s11032-015-0210-z.
18. Tereza Sovová, Gerard Kerins, Kateřina Demnerová, Jaroslava Ovesná. Genome Editing with Engineered Nucleases in Economically Important Animals and Plants: State of the Art in the Research Pipeline. *Current Issues in Molecular Biology.* 2017:41–62. doi:10.21775/cimb.021.041.
19. Perner H, Modrzejewski D, Hartung F, Sprink T. Mega, Crispr, Finger? Die Techniken des Genome Editings. <https://www.dialog-gea.de/de/themen/inhalte/2017-07-10-mega-crispr-finger-die-techniken-des-genome-editings>. Accessed 9 Feb 2020.
20. Boch J, Bonas U. Xanthomonas AvrBs3 family-type III effectors: Discovery and function. *Annual review of phytopathology.* 2010;48 (2010):419–36. doi:10.1146/annurev-phyto-080508-081936.
21. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* 2012;337:816–21. doi:10.1126/science.1225829.
22. Subburaj S, Tu L, Jin Y-T, Bae S, Seo PJ, Jung YJ, Lee G-J. Targeted genome editing, an alternative tool for trait improvement in horticultural crops. *Hortic. Environ. Biotechnol.* 2016;57:531–43. doi:10.1007/s13580-016-0281-8.
23. Swarts DC, Jinek M. Cas9 versus Cas12a/Cpf1: Structure-function comparisons and implications for genome editing. *Wiley Interdiscip Rev RNA.* 2018:e1481. doi:10.1002/wrna.1481.
24. Dönmez D, Şimsek Ö, Aka Kacar Y. Genetic Engineering Techniques in Fruit Science. *IJOEAR.* 2016;2:115–28.

25. Kanchiswamy CN, Maffei M, Malnoy M, Velasco R, Kim J-S. Fine-Tuning Next-Generation Genome Editing Tools. *Trends Biotechnol.* 2016;34:562–74. doi:10.1016/j.tibtech.2016.03.007.
26. Khandagale K, Nadaf A. Genome editing for targeted improvement of plants. *Plant Biotechnol Rep.* 2016;10:327–43. doi:10.1007/s11816-016-0417-4.
27. Lee J, Chung J-H, Kim HM, Kim D-W, Kim H. Designed nucleases for targeted genome editing. *Plant Biotechnol J.* 2016;14:448–62. doi:10.1111/pbi.12465.
28. Hilscher J, Bürstmayr H, Stoger E. Targeted modification of plant genomes for precision crop breeding. *Biotechnol J.* 2017;12:1–14. doi:10.1002/biot.201600173.
29. Lusser, Emilio MPCPDR-C. New plant breeding techniques. 2011. <https://publications.jrc.ec.europa.eu/repository/bitstream/JRC63971/jrc63971.pdf>. Accessed 20 Jan 2020.
30. Hartung F, Schiemann J. Precise plant breeding using new genome editing techniques: Opportunities, safety and regulation in the EU. *Plant J.* 2014;78:742–52. doi:10.1111/tpj.12413.
31. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature.* 2016;533:420–4. doi:10.1038/nature17946.
32. Bartsch D, Bendiek J, Braeuning A, Dagand E, Duensing N, Fladung M, et al. Wissenschaftlicher Bericht zu den neuen Techniken in der Pflanzenzüchtung und der Tierzucht und ihren Verwendungen im Bereich der Ernährung und Landwirtschaft. 2018. [https://www.bmel.de/SharedDocs/Downloads/DE/\\_Landwirtschaft/Gruene-Gentechnik/Bericht\\_Neue\\_Zuechtungstechniken.html](https://www.bmel.de/SharedDocs/Downloads/DE/_Landwirtschaft/Gruene-Gentechnik/Bericht_Neue_Zuechtungstechniken.html). Accessed 23 Feb 2018.
33. Cardi T, Varshney R. Cisgenesis and genome editing: Combining concepts and efforts for a smarter use of genetic resources in crop breeding. *Plant Breed.* 2016;135:139–47. doi:10.1111/pbr.12345.
34. Fritsch J, Steinicke H. The opportunities and limits of genome editing. 2015. <http://nbn-resolving.de/urn:nbn:de:gbv:3:2-73788>.
35. Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, Qiu J-L. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat Biotechnol.* 2014;32:947–51. doi:10.1038/nbt.2969.
36. Riccroch A. Global developments of genome editing in agriculture. *Transgenic Res.* 2019;28:45–52. doi:10.1007/s11248-019-00133-6.
37. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol.* 2013;31:822–6. doi:10.1038/nbt.2623.
38. Yu Q-h, Wang B, Li N, Tang Y, Yang S, Yang T, et al. CRISPR/Cas9-induced Targeted Mutagenesis and Gene Replacement to Generate Long-shelf Life Tomato Lines. *Sci Rep.* 2017;7:818. doi:10.1038/s41598-017-12262-1.
39. Zhu C, Bortesi L, Baysal C, Twyman RM, Fischer R, Capell T, et al. Characteristics of Genome Editing Mutations in Cereal Crops. *Trends Plant Sci.* 2017;22:38–52. doi:10.1016/j.tplants.2016.08.009.
40. Zischewski J, Fischer R, Bortesi L. Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnol Adv.* 2017;35:95–104. doi:10.1016/j.biotechadv.2016.12.003.

41. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol.* 2013;31:827–32. doi:10.1038/nbt.2647.
42. Hahn F, Nekrasov V. CRISPR/Cas precision: do we need to worry about off-targeting in plants? *Plant Cell Rep* 2018. doi:10.1007/s00299-018-2355-9.
43. Endo M, Mikami M, Toki S. Multigene knockout utilizing off-target mutations of the CRISPR/Cas9 system in rice. *Plant Cell Physiol.* 2015;56:41–7. doi:10.1093/pcp/pcu154.
44. Zhang D, Zhang H, Li T, Chen K, Qiu J-L, Gao C. Perfectly matched 20-nucleotide guide RNA sequences enable robust genome editing using high-fidelity SpCas9 nucleases. *Genome Biol.* 2017;18:191. doi:10.1186/s13059-017-1325-9.
45. Raitskin O, Schudoma C, West A, Patron NJ. Comparison of efficiency and specificity of CRISPR-associated (Cas) nucleases in plants: An expanded toolkit for precision genome engineering. *PLoS ONE* 2019. doi:10.1371/journal.pone.0211598.
46. Tycko J, Myer VE, Hsu PD. Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity. *Mol Cell.* 2016;63:355–70. doi:10.1016/j.molcel.2016.07.004.
47. Kaya H, Numa H, Nishizawa-Yokoi A, Toki S, Habu Y. DNA Methylation Affects the Efficiency of Transcription Activator-Like Effector Nucleases-Mediated Genome Editing in Rice. *Front Plant Sci.* 2017;8:302. doi:10.3389/fpls.2017.00302.
48. Russo MT, Aiese Cigliano R, Sanseverino W, Ferrante MI. Assessment of genomic changes in a CRISPR/Cas9 *Phaeodactylum tricornutum* mutant through whole genome resequencing. *Peerj* 2018. doi:10.7717/peerj.5507.
49. Jansing J, Schiermeyer A, Schillberg S, Fischer R, Bortesi L. Genome Editing in Agriculture: Technical and Practical Considerations. *Int J Mol Sci* 2019. doi:10.3390/ijms20122888.
50. Altschul S. Basic Local Alignment Search Tool. *J Mol Biol.* 1990;215:403–10. doi:10.1006/jmbi.1990.9999.
51. Bae S, Park J, Kim J-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics.* 2014;30:1473–5. doi:10.1093/bioinformatics/btu048.
52. Lei Y, Lu L, Liu H-Y, Li S, Xing F, Chen L-L. CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. *Mol Plant.* 2014;7:1494–6. doi:10.1093/mp/ssu044.
53. Martin F, Sánchez-Hernández S, Gutiérrez-Guerrero A, Pinedo-Gomez J, Benabdellah K. Biased and Unbiased Methods for the Detection of Off-Target Cleavage by CRISPR/Cas9: An Overview. *Int J Mol Sci* 2016. doi:10.3390/ijms17091507.
54. Wu X, Kriz AJ, Sharp PA. Target specificity of the CRISPR-Cas9 system. *Quant Biol.* 2014;2:59–70. doi:10.1007/s40484-014-0030-x.
55. Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG, et al. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science.* 2010;327:92–4. doi:10.1126/science.1180677.
56. Jander G, Baerson SR, Hudak JA, Gonzalez KA, Gruys KJ, Last RL. Ethylmethanesulfonate saturation mutagenesis in *Arabidopsis* to determine frequency of herbicide resistance. *Plant Physiol.* 2003;131:139–46. doi:10.1104/pp.102.010397.
57. Miyao A, Nakagome M, Ohnuma T, Yamagata H, Kanamori H, Katayose Y, et al. Molecular spectrum of somaclonal variation in regenerated rice revealed by whole-genome sequencing. *Plant Cell Physiol.* 2012;53:256–64. doi:10.1093/pcp/pcr172.

58. Zhao H, Wolt JD. Risk associated with off-target plant genome editing and methods for its limitation. *Emerg. Top. Life Sci.* 2017;1:231–40. doi:10.1042/ETLS20170037.
59. Eckerstorfer MF, Dolezel M, Heissenberger A, Miklau M, Reichenbecher W, Steinbrecher RA, Waßmann F. An EU Perspective on Biosafety Considerations for Plants Developed by Genome Editing and Other New Genetic Modification Techniques (nGMs). *Front Bioeng Biotechnol.* 2019;7:31. doi:10.3389/fbioe.2019.00031.
60. Johnson R, Watkinson A, Mabe M. The STM Report- An overview of scientific and scholarly publishing. [https://www.stm-assoc.org/2018\\_10\\_04\\_STM\\_Report\\_2018.pdf](https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf). Accessed 24 May 2020.
61. Jinha AE. Article 50 million: an estimate of the number of scholarly articles in existence. *Learn. Pub.* 2010;23:258–63. doi:10.1087/20100308.
62. Haddaway NR, Feierman A, Grainger MJ, Gray CT, Tanriver-Ayder E, Dhaubanjhar S, Westgate MJ. EviAtlas: a tool for visualising evidence synthesis databases. *Environ Evid.* 2019;8:258. doi:10.1186/s13750-019-0167-1.
63. Kohl C, McIntosh EJ, Unger S, Haddaway NR, Kecke S, Schiemann J, Wilhelm R. Online tools supporting the conduct and reporting of systematic reviews and systematic maps: a case study on CADIMA and review of existing tools. *Environ Evid.* 2018;7:2420. doi:10.1186/s13750-018-0115-5.
64. Collaboration for Environmental Evidence. Guidelines and Standards for Evidence synthesis in Environmental Management. Version 5.0 (AS Pullin, GK Frampton, B Livoreil & G Petrokofsky, Eds). 2018. <http://www.environmentalevidence.org/guidelines/table-of-contents>. Accessed 3 Nov 2019.
65. EFSA. Application of systematic review methodology to food and feed safety assessments to support decision making. *EFSA Journal.* 2010;8:1637. doi:10.2903/j.efsa.2010.1637.
66. HEARD Project. Rapid Review vs. Systematic Review: What are the differences? 2019. <https://www.heardproject.org/news/rapid-review-vs-systematic-review-what-are-the-differences/>. Accessed 2 Nov 2019.
67. Mallett R, Hagen-Zanker J, Slater R, Duvendack M. The benefits and challenges of using systematic reviews in international development research. *Journal of Development Effectiveness.* 2012;4:445–55. doi:10.1080/19439342.2012.711342.
68. Kohl C, Modrzejewski D, Kopertekh L, Dietz-Pfeilstetter A, Fischer M, Menz J, et al. Übersicht über Nutz- und Zierpflanzen, die mittels Gentechnik und neuer molekularbiologischer Techniken für die Bereiche Ernährung, Landwirtschaft, Gartenbau, Arzneimittelherstellung und -forschung entwickelt werden. 2017. [https://www.bmel.de/SharedDocs/Downloads/Landwirtschaft/Pflanze/GrueneGentechnik/NMT\\_Stand-Regulierung\\_Anlage4.pdf?\\_\\_blob=publicationFile](https://www.bmel.de/SharedDocs/Downloads/Landwirtschaft/Pflanze/GrueneGentechnik/NMT_Stand-Regulierung_Anlage4.pdf?__blob=publicationFile). Accessed 25 Apr 2020.
69. Modrzejewski D, Hartung F, Sprink T, Krause D, Kohl C, Wilhelm R. 1. Aktualisierung der Übersicht über Nutz- und Zierpflanzen, die mittels neuer molekularbiologischer Techniken für die Bereiche Ernährung, Landwirtschaft und Gartenbau erzeugt wurden. 2018. [https://www.bmel.de/SharedDocs/Downloads/Landwirtschaft/Pflanze/GrueneGentechnik/NMT\\_Stand-Regulierung\\_Anlage4-Aktualisierung.pdf%3F\\_\\_blob%3DpublicationFile](https://www.bmel.de/SharedDocs/Downloads/Landwirtschaft/Pflanze/GrueneGentechnik/NMT_Stand-Regulierung_Anlage4-Aktualisierung.pdf%3F__blob%3DpublicationFile). Accessed 25 Apr 2020.
70. Modrzejewski D, Hartung F, Sprink T, Menz J, Kohl C, Delventhal R, Wilhelm R. 2. Aktualisierung der Übersicht über Nutz- und Zierpflanzen, die mittels neuer molekularbiologischer Techniken für die Bereiche Ernährung, Landwirtschaft und Gartenbau erzeugt

- wurden – marktorientierte Anwendungen. 2020.  
[https://www.bmel.de/SharedDocs/Downloads/DE/\\_Landwirtschaft/Gruene-Gentechnik/NMT\\_Uebersicht-Zier-Nutzpflanzen.pdf;jsessionid=5ECF3F75CE9277E0FA110E0D1240B1A3.internet2842?\\_\\_blob=publicationFile&v=3](https://www.bmel.de/SharedDocs/Downloads/DE/_Landwirtschaft/Gruene-Gentechnik/NMT_Uebersicht-Zier-Nutzpflanzen.pdf;jsessionid=5ECF3F75CE9277E0FA110E0D1240B1A3.internet2842?__blob=publicationFile&v=3). Accessed 18 May 2020.
71. Modrzejewski D. Repository about applications of genome editing in plants. <https://www.dialog-gea.de/de/service/repositorium>. Accessed 20 May 2020.
  72. Fritsch J. Wege zu einer wissenschaftlich begründeten, differenzierten Regulierung genomeditierter Pflanzen in der EU. 2019. [https://www.leopoldina.org/uploads/tx\\_leopublication/2019\\_Stellungnahme\\_Genomeditierte\\_Pflanzen\\_web.pdf](https://www.leopoldina.org/uploads/tx_leopublication/2019_Stellungnahme_Genomeditierte_Pflanzen_web.pdf). Accessed 25 Apr 2020.
  73. Scientific Advice Mechanism (SAM). Statement by the Group of Chief Scientific Advisors. 2018. [https://ec.europa.eu/info/sites/info/files/2018\\_11\\_gcsa\\_statement\\_gene\\_editing\\_1.pdf](https://ec.europa.eu/info/sites/info/files/2018_11_gcsa_statement_gene_editing_1.pdf). Accessed 25 Apr 2020.
  74. Bär K, Mann J. Gentechnik 2.0. 2019. [http://www.umweltinstitut.org/fileadmin/Mediapool/Druckprodukte/Gentechnik/PDF/Argumentationshilfe\\_Gentechnik\\_Download.pdf](http://www.umweltinstitut.org/fileadmin/Mediapool/Druckprodukte/Gentechnik/PDF/Argumentationshilfe_Gentechnik_Download.pdf). Accessed 25 Apr 2020.
  75. Then C, Miyazaki J, Bauer-Panskus A, Kawall K. Overview of genome editing applications using SDN-1 and SDN-2 in regard to EU regulatory issues. <https://www.testbiotech.org/node/2569>. Accessed 25 Apr 2020.
  76. Modrzejewski D. CRISPR-Adventskalender. 2019. <https://progressive-agrarwende.org/crispr-adventskalender-blog/>. Accessed 25 Apr 2020.
  77. Modrzejewski D. CRISPR-Bibliothek. <https://progressive-agrarwende.org/crispr-bibliothek/>. Accessed 25 Apr 2020.
  78. Jorasch P. Will the EU stay out of step with science and the rest of the world on plant breeding innovation? *Plant Cell Rep.* 2020;39:163–7. doi:10.1007/s00299-019-02482-2.
  79. Bioökonomierat. Genome Editing: Europa benötigt ein neues Gentechnikrecht. 2019. [https://biooekonomierat.de/fileadmin/Publikationen/berichte/BOERMEMO\\_07\\_final.pdf](https://biooekonomierat.de/fileadmin/Publikationen/berichte/BOERMEMO_07_final.pdf). Accessed 25 Apr 2020.
  80. Zsögön A, Čermák T, Naves ER, Notini MM, Edel KH, Weigl S, et al. De novo domestication of wild tomato using genome editing. *Nat Biotechnol* 2018. doi:10.1038/nbt.4272.

# Danksagung

Mein besonderer Dank gilt Ralf Wilhelm und Joachim Schiemann, die es mir ermöglicht haben zu diesem spannenden Thema am Julius Kühn-Institut zu promovieren. Des Weiteren danke ich Frau Professor Mahlein und Herrn Professor Spiller, die mich in den letzten Jahren mit hilfreichen Ratschlägen und Anregungen tatkräftig unterstützt haben. Auch danke ich Herrn Professor Scholten, dass er sich kurzfristig bereit erklärt hat, in der Prüfungskommission mitzuwirken. Ganz herzlich möchte ich mich auch bei meinem gesamten Review Team, bestehend aus Ralf Wilhelm, Frank Hartung, Thorben Sprink, Christian Kohl, Dörthe Krause, Heike Lehnert und Jens Keilwagen bedanken. Unsere Interaktionen bei Meetings sowie die Möglichkeit, an zahlreichen Veranstaltungen und Konferenzen teilzunehmen und aktiv mitzuwirken, waren zur Entwicklung meines Verständnisses über Genome Editing von zentraler Bedeutung. Ein weiterer Dank geht an alle JKI-Mitarbeiterinnen und JKI-Mitarbeiter des Instituts für die Sicherheit biotechnologischer Verfahren bei Pflanzen, für das angenehme Arbeitsklima und die zahlreichen Aktivitäten innerhalb und außerhalb der Arbeit.

# Erklärung

1. Hiermit erkläre ich, dass diese Arbeit weder in gleicher noch in ähnlicher Form bereits anderen Prüfungsbehörden vorgelegt hat.

Weiter erkläre ich, dass ich mich an keiner anderen Hochschule um einen Doktorgrad beworben habe.

Gerlingen, den 18.11.2020

.....  
(Unterschrift)

2. Hiermit erkläre ich eidesstattlich, dass diese Dissertation selbstständig und ohne unerlaubte Hilfe angefertigt wurde

Gerlingen, den 18.11.2020

.....  
(Unterschrift)