Degree project

UNIVERSITY OF SKÖVDE

Quantifying Off-Target Effects in Plant CRISPR/Cas9 Studies: A Systematic Review with Exploratory Quantitative Synthesis

Bachelor Degree Project in Bioscience G2E

BV515G, 30 credits

Spring semester 2026

Student: Glafira Surkova

Supervisor: Patric Nilsson

Abstract

CRISPR/Cas9 is widely used in plant biology because it enables targeted modification of genomic DNA in both model plants and crop species. Despite its precision, unintended off-target mutations remain an important concern because they can complicate the interpretation of edited phenotypes and influence how genome-edited plants are evaluated. This thesis examined how often experimentally tested candidate off-target sites were confirmed as mutated in published plant CRISPR/Cas9 studies. The work was designed as a systematic literature review and exploratory quantitative synthesis of primary research articles published from 2013 onward. No laboratory work was performed. Six study records provided directly extractable numerical data for quantitative synthesis. Across these studies, 7,921 candidate off-target sites were tested and 18 were confirmed as mutated, corresponding to a crude aggregate proportion of 0.23% with an exact 95% confidence interval of 0.13% to 0.36%. An exploratory random-effects meta-analysis produced a higher pooled estimate of 1.58%, but heterogeneity was very high. Off-target mutations were uncommon in routine specificity screens and in broad whole-genome studies using carefully designed guide RNAs, but they were clearly detectable in high-risk sequence contexts with highly similar genomic near-matches. The results indicate that plant CRISPR/Cas9 editing is often highly specific, but that specificity is conditional rather than absolute and must be interpreted together with guide design, genomic context, detection method, and background variation introduced during plant transformation and regeneration.

Popular Scientific Summary

CRISPR/Cas9 is a method that lets researchers change DNA at chosen places in a genome. In plant science, this has become a very useful tool. It can help researchers study what specific genes do, and it may also make it easier to develop crops with useful traits, such as better disease resistance or improved tolerance to environmental stress. Because the method is often described as precise, an important question is whether it always edits only the intended DNA site.

This thesis focused on that question by looking at off-target effects in plants. An off-target effect means that CRISPR/Cas9 changes DNA at an unintended location that looks similar to the real target. If this happens often, it can make research results harder to interpret. For example, if a plant shows a new trait after editing, it may be difficult to know whether the trait was caused by the intended edit or by another unintended mutation somewhere else in the genome.

The thesis did not include laboratory experiments. Instead, it was based on a structured review of published scientific studies. The main task was to find plant CRISPR/Cas9 papers that had actually tested possible off-target sites and reported numerical data showing how many sites were examined and how many were confirmed as mutated. These numbers were then compared across studies.

The results showed that confirmed off-target mutations were generally uncommon in the studies that could be pooled quantitatively. When all tested candidate sites from the included studies were combined, only a very small proportion were confirmed as true off-targets. At the same time, the results were not identical across all studies. Some experiments found no off-target mutations at all, while others found them more clearly when the guide RNA had very close matches elsewhere in the genome. This means that CRISPR/Cas9 specificity in plants is not simply “good” or “bad.” It depends strongly on how the guide RNA is designed and how similar other DNA sequences are to the intended target.

Another important finding was that not all mutations in edited plants are caused by CRISPR itself. Plant tissue culture, transformation, and regeneration can also introduce genetic changes. This means that researchers need to separate background mutations from true off-target edits when evaluating edited plants.

Overall, the thesis shows that CRISPR/Cas9 is often highly specific in plants, but careful guide design and careful analysis remain essential.

Table of Contents

1 Introduction **1**
1.1 Aim and objectives 3
2 Methods **3**
2.1 Literature search and study selection 4
2.1.1 Inclusion and exclusion criteria 4
2.2 Data extraction and analysis 6
3 Results **7**
4 Discussion **12**
4.1 Interpretation 12
4.2 Ethical aspects, gender equality perspective, and societal relevance 14
4.3 Future perspectives 15
4.4 Methodological limitations and conclusion 16
5 References **17**

List of abbreviations

Cas9 - CRISPR-associated protein 9

CI - confidence interval

CRISPR - clustered regularly interspaced short palindromic repeats

DNA - deoxyribonucleic acid

PAM - protospacer-adjacent motif

PCR - polymerase chain reaction

PRISMA - Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Q - Cochran's Q heterogeneity statistic

RNA - ribonucleic acid

sgRNA - single-guide RNA

1 Introduction

CRISPR/Cas9 has become an important tool in plant bioscience because it allows researchers to target specific genomic loci by redesigning the guide RNA rather than rebuilding the entire nuclease system. This flexibility has made CRISPR/Cas9 useful in both basic and applied plant research, including functional genomics, crop improvement, and trait development in model plants and crops. Chen et al. (2019), Gao (2021), and Jaganathan et al. (2018) describe CRISPR-based editing as a central technology in plant biotechnology because it can accelerate the analysis and modification of genes linked to disease resistance, abiotic stress tolerance, and agronomically important traits. Zhan et al. (2021) likewise identify genome editing as an important platform for both plant research and crop improvement.

The CRISPR/Cas9 system achieves targeted editing through the interaction between the Cas9 nuclease and a single-guide RNA. The guide RNA contains a sequence that is designed to complement the intended genomic target, while Cas9 recognizes a nearby protospacer-adjacent motif and introduces a double-strand break. Cellular DNA-repair pathways then repair the break, which can generate small insertions or deletions or, when a repair template is available, a specified sequence change. Přibylová and Fischer (2024) explain that editing efficiency and specificity are influenced by guide design, target accessibility, the protospacer-adjacent motif, and the repair processes operating in plant cells. This mechanism makes the system programmable, but it also creates the possibility that Cas9 will bind and cleave a genomic sequence that resembles the intended target. Tsai and Joung (2016) describe these unintended cleavage events as off-target activity and emphasize that mismatch tolerance is affected by both the number and the position of sequence differences between the guide RNA and a potential genomic site.

In plant research, specificity matters because the value of a genome-editing experiment depends on how confidently an observed phenotype can be linked to the intended mutation. If edits occur at unintended genomic sites, then genotype–phenotype interpretation becomes less secure. This issue is especially important in plants because transformation, tissue culture, and regeneration can also introduce background variation that is unrelated to Cas9 cleavage. Tang et al. (2018), Li et al. (2019), and Wang et al. (2021) showed that variants detected in regenerated edited plants cannot automatically be attributed to CRISPR/Cas9. Sturme et al. (2022) likewise emphasized the need to distinguish confirmed nuclease-associated modifications from other sequence variation. As a result, off-target effects in plants are not only a question of whether unintended edits occur, but also of how confidently true off-target mutations can be separated from other sources of sequence variation.

Off-target detection strategies also differ in what they are able to observe. Targeted approaches usually begin with computationally predicted candidate sites and then test those sites by PCR, amplicon sequencing, or another focused

method. These approaches can detect low-frequency edits at selected loci, but they cannot identify mutations outside the nominated candidate set. Whole-genome sequencing provides broader coverage and can reveal variants across the genome, although causal interpretation remains difficult when the edited plant differs from the reference genome or has accumulated mutations during tissue culture and regeneration. Tang et al. (2018) used whole-genome data and appropriate controls to distinguish nuclease-associated events from background variation in rice, while Li et al. (2019) demonstrated the importance of inherent and somaclonal variation in cotton. Wang et al. (2021) further showed that genotype–reference differences can alter which sites are recognized as possible off-targets in grapevine. Therefore, estimates of off-target frequency are influenced not only by Cas9 activity but also by guide selection, candidate-site definition, sequencing strategy, validation criteria, and the controls used to assign causality.

The plant literature does not give one simple answer to the question of CRISPR/Cas9 specificity. Several studies reported very low or undetectable off-target activity. Peterson et al. (2016) found no detectable off-target mutations among 178 predicted sites in Arabidopsis. Zhang et al. (2020) found no off-target mutations among the tomato candidate sites they examined. Li et al. (2019) and Wang et al. (2021), both using whole-genome sequencing-based approaches in crops, likewise reported only rare bona fide off-target mutations despite very large candidate-site sets. Earlier studies by Feng et al. (2014) and Zhang et al. (2014) also reported high specificity in Arabidopsis and rice, respectively. These findings support the view that well-designed plant CRISPR/Cas9 systems can be highly specific.

At the same time, other studies showed clearly that off-target mutations can occur under particular conditions. Zhang et al. (2018) reported high-frequency off-target mutagenesis in Arabidopsis when a guide had highly similar genomic near-matches. Tang et al. (2018) found a cluster of bona fide off-target sites in rice for one problematic sgRNA, whereas Young et al. (2019) showed in maize that a promiscuous guide could generate off-target edits even though bioinformatically specific guides did not. These findings suggest that the key question is not simply whether off-target effects occur, but under which conditions they occur and how often they are reported in studies where the data can be compared numerically.

That interpretation is also supported by broader evidence synthesis. In their plant-specific systematic review, Modrzejewski et al. (2020) found that off-target occurrence decreased sharply as mismatch number increased and that mismatches close to the PAM-proximal seed region affected off-target likelihood. Sturme et al. (2022) similarly concluded that most reported plant off-target modifications were small and that the current evidence base remains methodologically heterogeneous. Together, these reviews suggest that a structured quantitative synthesis is useful when it focuses on studies with

directly comparable numerical outcomes rather than simply pooling all descriptive claims about specificity.

A further difficulty is that the phrase “off-target frequency” can refer to different denominators. Some studies report the proportion of edited plants carrying an unintended mutation, some report the proportion of sequencing reads containing an edit at a particular locus, and others report the proportion of tested candidate sites that were confirmed as mutated. These outcomes answer different biological questions and should not be combined without a clearly defined unit of analysis. The present thesis focuses on the last of these outcomes: confirmed off-target sites divided by experimentally tested candidate sites. This definition makes it possible to extract a transparent numerator and denominator from each eligible study record. It does not represent the probability that an individual plant will carry an off-target mutation, but it provides a consistent basis for comparing studies that used substantially different plant species and detection methods. The remaining methodological differences still require cautious interpretation, which is why both crude and random-effects summaries are considered.

1.1 Aim and objectives

The aim of this thesis was therefore to estimate, from published plant CRISPR/Cas9 studies with extractable numerical data, the proportion of experimentally tested candidate off-target sites that were confirmed as mutated and to examine how this proportion varied across guide-risk contexts and detection methods. The objectives were to identify eligible plant studies, extract directly comparable numerator–denominator data, summarize study-level proportions, perform an exploratory quantitative synthesis, and interpret the findings in relation to sequence context and methodological design. The decision to frame the study as a systematic review with exploratory quantitative synthesis follows the transparent-reporting principles described by Page et al. (2021) and reflects the methodological heterogeneity identified by Modrzejewski et al. (2020). The analysis was not intended to estimate a universal plant CRISPR/Cas9 off-target rate.

2 Methods

This thesis was designed as a systematic literature review and exploratory meta-analysis of published studies investigating off-target effects in plant CRISPR/Cas9 genome editing. No laboratory work was performed. The analysis was based on numerical data reported in primary research articles. The main outcome was the proportion of experimentally tested candidate off-target sites that were confirmed to contain off-target mutations. The review followed the transparent-reporting principles described by Page et al. (2021), while the quantitative synthesis was conducted according to the meta-analytical procedures described by Harrer et al. (2021) and Balduzzi et al. (2019).

2.1 Literature search and study selection

The final literature searches were conducted on 12 May 2026 in Scopus, Web of Science Core Collection, PubMed, ScienceDirect, Google Scholar, and the University of Skövde library discovery tool. Searches combined terms relating to CRISPR/Cas9, off-target effects, and plants and were limited to English-language records published from 2013 onward. These sources were selected to cover plant biology, genetics, biotechnology, and genome-editing literature. The complete platform-specific search strings, filters, search dates, and numbers of records retrieved are provided in Appendix 1 (Search Documentation Table).

Scopus, Web of Science, PubMed, and the University of Skövde library discovery tool formed the formal screening set and yielded 4,108 records. ScienceDirect and Google Scholar were used only as supplementary search tools because their broad result sets could not be narrowed reproducibly enough for formal record accounting; their results were therefore excluded from the PRISMA totals. Duplicate records were identified using title, DOI, author, and publication-year metadata. Removal of 589 duplicates left 3,519 records for title and abstract screening.

2.1.1 Inclusion and exclusion criteria

Studies were eligible if they were peer-reviewed primary research articles published in English from 2013 to 12 May 2026; investigated CRISPR/Cas9 editing in a plant species; experimentally assessed candidate off-target sites; and reported, or allowed direct extraction of, both the number of candidate sites tested and the number confirmed as mutated. Studies that investigated several editing systems were eligible only when the CRISPR/Cas9 results were reported separately. Reviews, systematic maps, commentaries, protocols, and other non-primary publications were excluded from the quantitative synthesis. Studies were also excluded if they concerned non-plant systems, assessed off-target risk only computationally without experimental validation, did not provide an extractable numerator–denominator pair, or reported results for other editing systems without separate CRISPR/Cas9 data. Records for which the full text could not be accessed were excluded because eligibility and numerical data could not be verified.

The remaining records were screened first by title and abstract. Records were excluded at this stage if they were clearly unrelated to plant CRISPR/Cas9 genome editing, did not contain off-target analysis, were non-plant studies, or were obviously non-primary formats such as reviews, commentaries, or protocol papers. After title and abstract screening, 107 full-text articles remained and were sought for retrieval. Of these, 20 could not be retrieved, 87 full-text reports were assessed, and 81 were excluded. The main reasons for exclusion at full-text stage were that the article was not a peer-reviewed primary research study, did not experimentally assess off-target mutations, reported only computational predictions, focused on non-plant systems, did not provide extractable

numerator–denominator data, or used other genome-editing systems without separately reported CRISPR/Cas9 results. Six study records were included in the final quantitative synthesis. The PRISMA flow diagram correlating to the actions taken to reach these six records can be found in Figure 1.

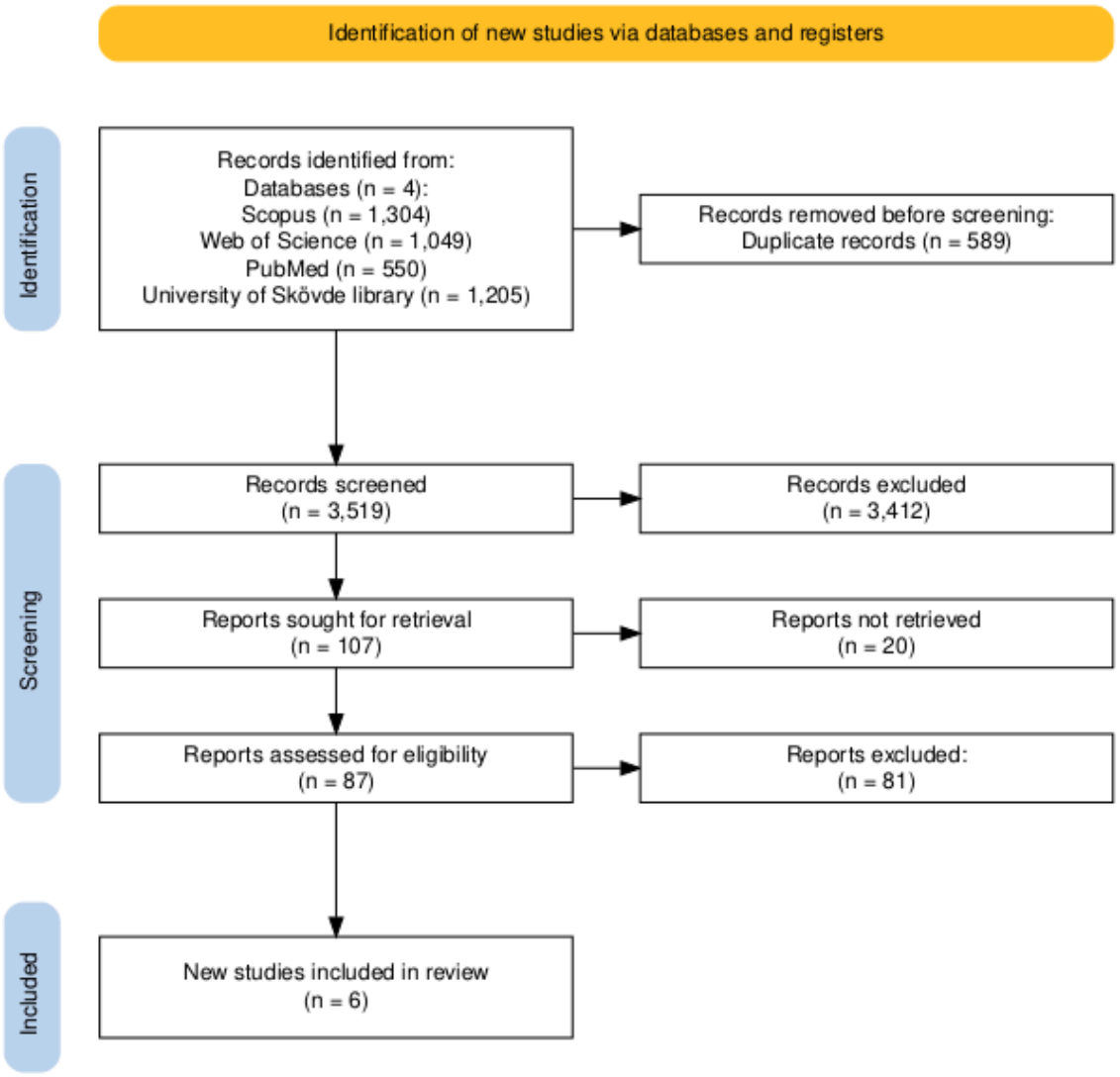

Figure 1. PRISMA flow diagram showing record identification, screening, full-text assessment, and inclusion in the quantitative synthesis..

The six study records included in the quantitative synthesis were Peterson et al. (2016), Zhang et al. (2018), Tang et al. (2018), Li et al. (2019), Zhang et al. (2020), and Wang et al. (2021). These records were included because they met the eligibility criteria and provided directly extractable numerical data on the number of candidate off-target sites tested and the number of confirmed off-target mutations.

2.2 Data extraction and analysis

For each included study, the following information was extracted: author and year, plant species, target gene, CRISPR/Cas system, number of guide RNAs when available, off-target detection method, number of candidate off-target sites tested, and number of confirmed off-target mutations. The unit of analysis was the study record. In practice, this meant either one paper or one separately extractable off-target dataset within a paper. In the present synthesis, the six included papers contributed six study records. The effect size was the proportion of confirmed off-target mutations among experimentally tested candidate off-target sites. The numerator was the number of confirmed off-target mutations reported in a study record, and the denominator was the number of candidate off-target sites tested experimentally in that same study record. Zero-event studies were retained because their exclusion would have inflated the apparent frequency of off-target effects. Peterson et al. (2016), Zhang et al. (2018), Tang et al. (2018), Li et al. (2019), Zhang et al. (2020), and Wang et al. (2021) provided the six numerator–denominator pairs.

The quantitative synthesis consisted of a crude aggregate summary and exploratory random-effects meta-analyses. The crude aggregate proportion was calculated by dividing the total number of confirmed off-target sites by the total number of tested candidate sites across all six records. The primary model was an inverse-variance random-effects meta-analysis of logit-transformed proportions, fitted in R with the meta package and the metaprop() function. Restricted maximum likelihood estimation was used for the between-study variance. Harrer et al. (2021) and Balduzzi et al. (2019) describe the use and interpretation of random-effects models in R. The Freeman–Tukey double-arcsine transformation had been considered as a sensitivity approach during planning, but it was not included because the thesis reports only the logit-transformed model and does not present a complete set of parallel sensitivity analyses.

For studies with zero confirmed events, a continuity correction of 0.5 was applied only when required to calculate the logit-transformed estimate. Individual study confidence intervals were calculated as exact binomial intervals, while pooled confidence intervals were calculated on the logit scale and then back-transformed to proportions. Heterogeneity was evaluated using Cochran's Q, tau-squared, and I-squared.

Two exploratory subgroup classifications were analyzed using the same logit-transformed random-effects procedure. The first compared high-risk homology or problematic-guide contexts with routine specificity screens. The second compared small targeted assays with whole-genome sequencing-based studies. Separate between-study variance estimates were calculated within each subgroup. Random-effects pooled proportions, 95% confidence intervals, tau-squared, and I-squared were reported for each subgroup. No leave-one-out analysis was performed, and it was therefore not included among the reported methods. Because each subgroup contained only two to four study records, these

analyses were treated as exploratory and were not used to establish definitive differences between groups.

Possible publication bias was considered qualitatively. Formal funnel-plot asymmetry tests were not performed because the synthesis contained only six highly heterogeneous study records, making such tests uninformative. Publication bias could nevertheless affect the evidence base if studies with null off-target findings were less likely to be published or if studies without easily extractable numerical results were less likely to enter the quantitative synthesis. This possibility was considered when interpreting the pooled estimates.

A simple quality assessment was also performed. Studies were considered more reliable if they clearly described the plant material, target gene, guide RNA design, off-target detection method, number of tested candidate sites, and number of confirmed mutations. This assessment was not used to exclude studies that met the inclusion criteria, but it was used when interpreting the strength of the results. Tang et al. (2018), Li et al. (2019), and Wang et al. (2021) illustrate why controls for background variation and genotype-aware analysis are important, while Sturme et al. (2022) describe the broader methodological heterogeneity of the plant off-target literature. Because only six records provided directly extractable numerator–denominator data, the synthesis was interpreted as exploratory rather than as a definitive estimate of a universal plant CRISPR/Cas9 off-target rate.

3 Results

Six study records met the requirement of reporting directly extractable numerator–denominator data and were therefore entered into the quantitative synthesis. These studies covered Arabidopsis, rice, cotton, tomato, and grapevine, and they represented both small targeted off-target assays and broader whole-genome sequencing studies. Table 1 summarizes the study-level data used in the synthesis. The first point that stands out is that several studies reported zero confirmed off-target events. Peterson et al. (2016) found none among 178 candidate sites in Arabidopsis, and Zhang et al. (2020) likewise found none among 18 candidate sites in tomato. In contrast, Zhang et al. (2018) and Tang et al. (2018) reported much higher proportions of confirmed off-target mutations. Li et al. (2019) and Wang et al. (2021) are different again: both had very large denominators but very low event rates, which means they contributed heavily to the overall crude proportion. Table 1 therefore shows immediately that the included studies were not measuring exactly the same experimental context.

Table 1. Study-level data included in the quantitative synthesis

Citation	Plant species	Detection method	Tested candidate sites	Confirmed off-target sites	Proportion

Peterson et al. (2016)	*Arabidopsis thaliana*	Deep sequencing of predicted sites	178	0	0.00%
Zhang et al. (2018)	*Arabidopsis thaliana*	Targeted PCR and sequencing of predicted sites	3	2	66.67%
Tang et al. (2018)	Rice	Whole-genome sequencing with computational nomination	37	11	29.73%
Li et al. (2019)	Cotton	Whole-genome sequencing	4,413	4	0.09%
Zhang et al. (2020)	Tomato	PCR and sequencing of predicted sites	18	0	0.00%
Wang et al. (2021)	Grapevine	Whole-genome sequencing	3,272	1	0.03%

For Tang et al. (2018), one PAM-different locus discussed in the article was treated as technically on-target and therefore was not counted as a bona fide off-target in the study-level table. Study-level values were extracted from the original primary studies.

The studies in Table 1 illustrate why off-target frequency cannot be summarized adequately by a single average alone. Peterson et al. (2016) examined 178 predicted Arabidopsis sites and found no evidence of Cas9-induced indels. Zhang et al. (2018), in contrast, tested only three Arabidopsis candidate sites but detected off-target mutations in two of them, specifically in TRY and CPC. Tang et al. (2018) predicted 37 candidate off-target sites associated with their rice system and validated 11 bona fide off-target mutations, all linked to one problematic sgRNA with highly homologous genomic matches. Li et al. (2019) and Wang et al. (2021) found only rare bona fide off-target indels despite large whole-genome sequencing datasets, and Zhang et al. (2020) detected no off-target mutations in tomato among the loci examined.

All six included records met the minimum reporting requirements because they described the plant material, target or guide context, detection method, and extractable numbers of tested and confirmed off-target sites. Reporting was strongest in Tang et al. (2018), Li et al. (2019), and Wang et al. (2021), which used genome-wide data with controls or genotype-aware analyses to distinguish likely nuclease-associated variants from background variation. The targeted studies described the examined candidate loci clearly, but their smaller, selectively nominated site sets limited broader inference. No record was excluded on quality grounds; differences in reporting and control design were considered when interpreting the heterogeneous results.

When the six studies in Table 1 were combined, 7,921 candidate off-target sites had been experimentally tested and 18 were confirmed as mutated. This gave a crude aggregate proportion of 0.23%, with an exact 95% confidence interval of 0.13% to 0.36%. An exploratory random-effects meta-analysis using the logit-transformed model produced a higher pooled estimate of 1.58%, with a very wide confidence interval and very high heterogeneity. Table 2 presents these summary results. The main point the reader should notice is that the crude aggregate proportion and the random-effects pooled estimate are not answering exactly the same question. The crude aggregate proportion is dominated by the large studies with low event rates, especially Li et al. (2019) and Wang et al. (2021), because those studies contribute most of the tested candidate sites. By contrast, the random-effects model gives more weight to between-study variation, which is why it produces a higher point estimate and a much wider confidence interval. As Harrer et al. (2021) explain, a random-effects estimate must be interpreted together with heterogeneity rather than as a universal common rate.

Table 2. Synthesis summary

Summary measure	Value
Number of pooled study records	6
Total candidate off-target sites tested	7,921
Total confirmed off-target sites	18
Crude aggregate proportion	0.23%
Exact 95% CI for crude aggregate proportion	0.13% to 0.36%
Exploratory random-effects pooled proportion	1.58%
Approximate 95% CI for random-effects proportion	0.09% to 21.53%
Cochran’s Q	137.81
tau^2	11.47
I^2	96.4%

The pooled estimate in Table 2 is based on the logit-transformed random-effects model described in the Methods section. Because the dataset was small and highly heterogeneous, the pooled estimate was treated as exploratory rather than definitive.

Figure 2 presents the study-level proportions and exact 95% confidence intervals together with the exploratory random-effects estimate. The forest plot makes the scale of between-study variation visible: the two small high-risk records have much higher proportions than the large low-event records, while the pooled confidence interval remains wide.

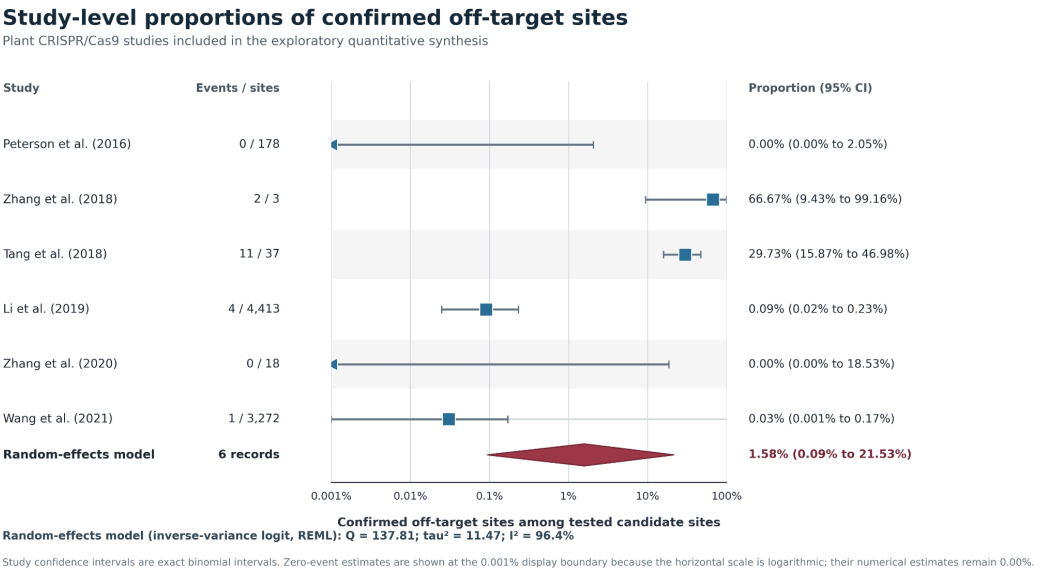

Figure 2. Forest plot of the proportions of confirmed off-target sites in the six included study records. Squares represent study-level proportions, with their sizes reflecting random-effects weights, and horizontal lines represent exact 95% binomial confidence intervals. The diamond represents the inverse-variance, logit-transformed random-effects pooled estimate calculated using restricted maximum likelihood estimation. Zero-event estimates are positioned at the 0.001% display boundary because the horizontal axis is logarithmic; their numerical estimates remain 0.00%.

The descriptive subgroup comparison in Table 3 makes the context dependence even more obvious. The two clearly high-risk homology contexts, represented by Zhang et al. (2018) and Tang et al. (2018), accounted for 13 confirmed off-target mutations among only 40 tested candidate sites, corresponding to 32.50%. By contrast, the routine specificity screens accounted for only five confirmed off-target mutations among 7,881 tested sites, corresponding to 0.06%. A second descriptive grouping showed that the whole-genome sequencing-based studies had a lower crude proportion than the small targeted assays once all tested sites were combined. The important point here is that these subgroup comparisons are descriptive, not inferential. With only six study records, the dataset is too small to support strong statistical subgroup conclusions. Table 3 should therefore be read as a way of organizing the pattern of results rather than as evidence of formally established subgroup effects.

Table 3. Descriptive subgroup comparisons

Analytical context	Included study basis	Tested candidate sites	Confirmed off-target sites	Crude proportion
High-risk homology or problematic-guide	Zhang et al. (2018); Tang et al. (2018)	40	13	32.50%

contexts				
Routine specificity screens with rare or no detectable off-targets	Peterson et al. (2016); Li et al. (2019); Zhang et al. (2020); Wang et al. (2021)	7,881	5	0.06%
Small targeted assays	Peterson et al. (2016); Zhang et al. (2018); Zhang et al. (2020)	199	2	1.01%
Whole-genome sequencing-based studies	Tang et al. (2018); Li et al. (2019); Wang et al. (2021)	7,722	16	0.21%

Table 3 is descriptive and is intended to clarify the pattern of results across different experimental contexts. It should not be interpreted as a formal inferential subgroup meta-analysis.

To complement these descriptive comparisons, exploratory random-effects meta-analyses were performed separately within each subgroup. The same inverse-variance, logit-transformed model and restricted maximum likelihood estimator used for the overall synthesis were applied. The subgroup estimates and heterogeneity statistics are presented in Table 4. Because each subgroup contained only two to four study records, the results were interpreted as exploratory and were not used to establish definitive subgroup differences.

Table 4. Exploratory random-effects subgroup meta-analyses

Subgroup	**Study records**	**Random-effects pooled proportion (95% CI)**	**tau^2**	**I^2**
High-risk homology or problematic-guide contexts	2	37.18% (14.46%–67.44%)	0.39	32.5%
Routine specificity screens with rare or no detectable off-targets	4	0.16% (0.03%–0.84%)	1.75	58.4%
Small targeted assays	3	5.33% (0.12%–72.07%)	9.57	84.7%

Whole-genome sequencing-based studies	3	0.52% (0.006%–30.54%)	14.95	98.4%

Table 4 presents the results of the exploratory random-effects meta-analyses conducted within the four subgroups shown descriptively in Table 3. The table reports the pooled proportion, 95% confidence interval, and heterogeneity statistics for each subgroup. Because each analysis included only two to four study records, the results were interpreted cautiously as exploratory summaries rather than as evidence of definitive subgroup differences.

The two subgroup classifications represent different ways of organizing the same six study records. Each record therefore belongs to one guide-risk subgroup and one detection-method subgroup, and the four rows should not be treated as independent datasets. The wide confidence intervals and high heterogeneity in several subgroups reflect the small number of records and substantial methodological differences between studies. Consequently, the subgroup estimates describe patterns within the included evidence but do not demonstrate causal differences between guide-risk contexts or detection methods.

The pooled dataset was restricted to studies with directly extractable numerator–denominator data and was not intended to capture every plant CRISPR/Cas9 specificity paper. Qualitative evidence outside the pooled tables nevertheless supported the same interpretation. Young et al. (2019) reported that computationally unique maize guides showed no detectable off-target cutting in plants, whereas a deliberately promiscuous guide produced expected off-target editing. Yang et al. (2017) likewise reported no mutations in the putative off-target sites examined in Brassica napus. These studies reinforce the conclusion that plant off-target outcomes depend strongly on guide choice and sequence context.

Taken together, the Results section supports three main observations. First, the overall proportion of confirmed off-target sites among experimentally tested candidates was low when all six records were combined. Second, that low crude proportion concealed large between-study differences, as shown by the higher random-effects estimate, the extreme overall heterogeneity, and Figure 2. Third, both the descriptive counts and the exploratory subgroup models showed the highest proportions in high-risk sequence contexts, while routine screens and large crop studies contributed much lower event rates. The width of several subgroup confidence intervals also demonstrates that the available evidence remains too limited for a precise universal estimate.

4 Discussion

4.1 Interpretation

The main finding of this thesis is that off-target mutations in plant CRISPR/Cas9 studies were uncommon overall, but clearly not absent under all conditions. The

crude aggregate proportion of 0.23% indicates that confirmed off-target mutations were rare when all tested candidate sites from the pooled studies were combined. At the same time, the exploratory random-effects estimate was higher and heterogeneity was extreme, which means that the included studies were not repeatedly measuring one common biological rate. Instead, the synthesis suggests that off-target risk in plants is strongly context dependent. This conclusion agrees closely with Modrzejewski et al. (2020), who showed across a broader plant evidence base that off-target occurrence drops sharply as mismatch number increases and that PAM-proximal mismatch patterns matter. The present thesis arrives at the same broad interpretation, but it does so from a much narrower evidence base because only studies with directly extractable numerator–denominator data were included in the quantitative synthesis. Compared with broader systematic reviews, this makes the numerical comparison more transparent, but it also makes the pooled dataset much smaller and more sensitive to study-level variation.

The subgroup models strengthen this interpretation while also demonstrating its uncertainty. The pooled proportion for high-risk homology or problematic-guide contexts was 37.18%, compared with 0.16% for routine specificity screens. This large contrast is consistent with the biological expectation that highly similar genomic near-matches increase off-target risk. However, the comparison was based on only two high-risk and four routine records, and the subgroups differed in more than guide-risk status. The detection-method subgroups were also highly heterogeneous: targeted assays had a pooled estimate of 5.33% and whole-genome sequencing-based studies had an estimate of 0.52%, but both confidence intervals were very wide. These values should therefore be understood as exploratory summaries of the included records rather than evidence that one detection method produces a fixed level of off-target activity. The subgroup results mainly show that guide context, study design, and denominator structure cannot be separated reliably in such a small evidence base.

The comparison between the included primary studies helps explain why the pooled results were so heterogeneous. Peterson et al. (2016) found no off-target mutations in Arabidopsis among 178 candidate sites, whereas Zhang et al. (2018) found a very high proportion in a much smaller Arabidopsis dataset. These results are not truly contradictory once the experimental context is considered. Peterson et al. (2016) examined a much broader set of predicted sites in a system that appeared well controlled, while Zhang et al. (2018) focused on a guide RNA that had highly similar genomic near-matches and showed that off-target cleavage could still occur even with three mismatches if the mismatch pattern was permissive. In other words, Peterson et al. illustrate high specificity in a routine screening context, whereas Zhang et al. illustrate what can happen in a locally high-risk sequence environment.

A similar contrast appears when Tang et al. (2018) are compared with Li et al. (2019) and Wang et al. (2021). Tang et al. reported many off-targets in one rice

guide context, but those off-targets were concentrated in a single problematic sgRNA with highly homologous alternative sites. By contrast, Li et al. in cotton and Wang et al. in grapevine examined very large candidate-site sets using whole-genome sequencing and found only rare bona fide off-target mutations. The most plausible interpretation is not that rice is generally more error-prone than cotton or grapevine, but that guide-level sequence context mattered more than species identity alone. This again matches the broader pattern described by Modrzejewski et al. (2020), where mismatch number and mismatch location were major determinants of off-target occurrence.

Young et al. (2019) provide useful supporting evidence for this interpretation even though the paper was not included in the pooled denominator–numerator table. In maize, they observed no detectable off-target cleavage for bioinformatically specific guides but did observe off-target events for a deliberately promiscuous guide. That experiment supports the central interpretation of the present thesis: off-target activity in plant CRISPR/Cas9 systems is not best understood as either uniformly rare or uniformly common, but as highly dependent on guide design and genomic context. The present synthesis therefore aligns well with both the targeted experimental comparisons in Young et al. (2019) and the broader evidence mapping in Modrzejewski et al. (2020) and Sturme et al. (2022). What distinguishes this thesis is that it focuses specifically on the proportion of confirmed off-target sites among experimentally tested candidate sites, which makes the outcome numerically clear but also necessarily excludes many studies that were informative in other ways.

A second major point is that background variation remained crucial in interpreting plant off-target results. Tang et al. (2018) concluded that many mutations in edited rice lines were more plausibly explained by tissue culture and transformation than by nuclease cleavage itself. Li et al. (2019) described substantial inherent and somaclonal variation in cotton, and Wang et al. (2021) showed that genotype–reference divergence in grapevine altered the off-target search landscape. This means that plant specificity cannot be evaluated simply by counting all variants in an edited genome. Instead, the most convincing studies are those that distinguish true off-target edits from the broader background of non-CRISPR variation. That point helps explain why the present thesis found very low crude proportions overall despite clear evidence that high-risk guide contexts can produce measurable unintended editing.

4.2 Ethical aspects, gender equality perspective, and societal relevance

Although this literature-based thesis did not involve laboratory experiments, human participants, or animals, it still raises ethical and societal questions. Responsible risk communication is important. It would be misleading to describe plant CRISPR/Cas9 as perfectly precise when off-target mutations can occur, but it would also be misleading to describe the technology as broadly unreliable when confirmed off-target events were uncommon in many plant studies. Tang et al. (2018), Modrzejewski et al. (2020), and Sturme et al. (2022) support a

balanced interpretation in which specificity depends on guide design, genomic context, and detection method.

Public trust in plant biotechnology depends partly on transparent reporting of what was tested, what was detected, and what remained uncertain. This is relevant because genome editing is discussed in relation to food security, climate adaptation, and sustainable agriculture. By clarifying the conditions under which unintended edits are more or less likely, this study may contribute to society by supporting safer experimental design, more reliable evaluation of genome-edited crops, and clearer public and regulatory communication. Such evidence can help researchers prioritize careful guide selection and appropriate validation without overstating either the benefits or the risks of the technology. Chen et al. (2019) and Gao (2021) describe the potential contribution of genome editing to crop improvement, making accurate specificity assessment important for responsible use.

A gender equality perspective is relevant even though plant genomes are not gendered. Access to agricultural biotechnology, participation in scientific research, and the benefits of improved crop varieties are shaped by broader social structures. Transparent communication can support inclusive public engagement and ensure that diverse groups, rather than only technical specialists, consider the potential benefits and risks of plant genome editing.

4.3 Future perspectives

The results of this thesis suggest several concrete directions for future work. First, primary studies should report off-target data in a more standardized way. Many plant papers discuss specificity, but fewer provide directly extractable numerator–denominator pairs for candidate sites tested and confirmed events. Page et al. (2021) emphasize transparent reporting in systematic reviews, while Modrzejewski et al. (2020) show how inconsistent reporting limits comparison across plant studies. Consistent reporting of candidate-site counts, validation counts, and guide-level results would strengthen future evidence synthesis and support formal subgroup meta-analysis across plant systems.

Second, future studies should distinguish clearly between true off-target mutations and background variation introduced by tissue culture, transformation, and regeneration. Tang et al. (2018), Li et al. (2019), and Wang et al. (2021) each demonstrated that detected genomic variants cannot automatically be attributed to CRISPR/Cas9. Genome-wide validation should therefore include appropriate controls, such as non-edited but similarly regenerated material. Wang et al. (2021) also showed that genotype-aware reference genomes are important because reference divergence can alter the apparent off-target landscape.

Third, guide RNA properties should be reported more consistently. Metadata such as mismatch number, PAM-proximal mismatch placement, guide specificity scores, target-copy context, and candidate-site nomination strategy would

improve comparisons between studies. Young et al. (2019) and Modrzejewski et al. (2020) both indicate that guide-level sequence context may explain more variation in observed off-target rates than plant species alone.

Finally, larger and more comparable datasets are needed. Sturme et al. (2022) described substantial methodological heterogeneity in the plant off-target literature, while Přibylová and Fischer (2024) emphasized the importance of careful target selection and validation. Future work should therefore develop coordinated datasets with standardized reporting across detection methods, particularly targeted sequencing and whole-genome sequencing. Such datasets would allow detection-method effects, guide-risk contexts, and possible publication bias to be evaluated more rigorously.

4.4 Methodological limitations and conclusion

The most important limitation of this thesis is that the pooled quantitative dataset was small. Although the plant CRISPR/Cas9 literature is broader than the six studies included in Table 1, many relevant papers did not report the number of tested candidate sites and confirmed off-target mutations in one directly extractable form. Haddaway et al. (2022) show how standardized PRISMA 2020 tools can strengthen transparent study-selection reporting, although such tools cannot resolve missing outcome data. The synthesis therefore favored transparency and comparability over maximal study count. Modrzejewski et al. (2019, 2020) included broader evidence bases, whereas the present dataset should be understood as a focused subset rather than a complete census of plant specificity studies.

A second limitation is methodological heterogeneity. The included studies differed in plant species, guide design, candidate-site nomination strategy, denominator size, and off-target detection method. A study testing three highly enriched near-matches is not directly comparable to a study evaluating thousands of sites after whole-genome sequencing. This explains the wide random-effects confidence interval and very high overall I^2. The subgroup analyses did not remove this limitation because each subgroup contained only two to four records and, in several cases, remained highly heterogeneous. In this situation, the descriptive structure of the evidence was more informative than any pooled estimate alone. The crude aggregate proportion was dominated by large low-event studies, whereas the random-effects models reflected the strong differences between study contexts.

Possible publication bias is another limitation. Studies that detect off-target mutations may be more likely to emphasize or publish those findings, while studies with null results may report specificity less prominently or may not provide extractable site-level counts. Conversely, large studies reporting high specificity may also be especially visible. With only six heterogeneous records, a funnel plot or statistical test for asymmetry would not provide a reliable assessment. Publication bias therefore could not be quantified or excluded, and

the pooled estimates should be interpreted as summaries of the available eligible evidence rather than unbiased estimates of all completed plant CRISPR/Cas9 experiments.

These limitations also clarify how the thesis differs from broader systematic reviews. Modrzejewski et al. (2020) could make wider claims about mismatch structure because their review included more studies and focused on broader patterns of reported off-target outcomes. The present thesis is narrower but more directly quantitative in one specific sense: it includes only records reporting the number of experimentally tested candidate sites and the number confirmed as mutated. That makes the outcome transparent but reduces the evidence base available for pooling.

Within these limits, the overall conclusion is clear. Off-target activity in plant CRISPR/Cas9 editing was uncommon in the pooled data but was not absent. Confirmed mutations were rare in many routine specificity screens and large whole-genome studies, yet they became more visible in high-homology guide contexts. The most defensible conclusion is therefore that plant CRISPR/Cas9 specificity is conditional: often high, but dependent on guide design, sequence context, and the way off-targets are detected and interpreted. Future work would benefit from standardized reporting, better control of background variation, and larger comparable datasets that permit stronger subgroup analysis and more reliable assessment of publication bias.

5 References

Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: A practical tutorial. *Evidence-Based Mental Health*, 22(4), 153–160. doi:10.1136/ebmental-2019-300117

Chen, K., Wang, Y., Zhang, R., Zhang, H., & Gao, C. (2019). CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annual Review of Plant Biology*, *70*, 667–697. doi:10.1146/annurev-arplant-050718-100049

Feng, Z., Mao, Y., Xu, N., Zhang, B., Wei, P., Yang, D.-L., Wang, Z., Zhang, Z., Zheng, R., Yang, L., Zeng, L., Liu, X., & Zhu, J.-K. (2014). Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 111(12), 4632–4637. doi:10.1073/pnas.1400822111

Gao, C. (2021). Genome engineering for crop improvement and future agriculture. *Cell*, 184(6), 1621–1635. doi:10.1016/j.cell.2021.01.005

Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital

transparency and Open Synthesis. *Campbell Systematic Reviews, 18,* e1230. doi:10.1002/cl2.1230

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing meta-analysis with R: A hands-on guide.* CRC Press. doi:10.1201/9781003107347

Jaganathan, D., Ramasamy, K., Sellamuthu, G., Jayabalan, S., & Venkataraman, G. (2018). CRISPR for crop improvement: An update review. *Frontiers in Plant Science, 9,* 985. doi:10.3389/fpls.2018.00985

Li, J., Manghwar, H., Sun, L., Wang, P., Wang, G., Sheng, H., Zhang, J., Liu, H., Qin, L., Rui, H., Li, B., Lindsey, K., Daniell, H., Jin, S., & Zhang, X. (2019). Whole genome sequencing reveals rare off-target mutations and considerable inherent genetic or/and somaclonal variations in CRISPR/Cas9-edited cotton plants. *Plant Biotechnology Journal, 17*(5), 858–868. doi:10.1111/pbi.13020

Modrzejewski, D., Hartung, F., Sprink, T., Krause, D., Kohl, C., & Wilhelm, R. (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. *Environmental Evidence, 8*, 27. doi:10.1186/s13750-019-0171-5

Modrzejewski, D., Hartung, F., Lehnert, H., Sprink, T., Kohl, C., Keilwagen, J., & Wilhelm, R. (2020). Which factors affect the occurrence of off-target effects caused by the use of CRISPR/Cas: A systematic review in plants. *Frontiers in Plant Science, 11*, 574959. doi:10.3389/fpls.2020.574959

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372,* n71. doi:10.1136/bmj.n71

Peterson, B. A., Haak, D. C., Nishimura, M. T., Teixeira, P. J. P. L., James, S. R., Dangl, J. L., & Nimchuk, Z. L. (2016). Genome-wide assessment of efficiency and specificity in CRISPR/Cas9 mediated multiple site targeting in Arabidopsis. *PLOS ONE, 11*(9), e0162169. doi:10.1371/journal.pone.0162169

Přibylová, A., & Fischer, L. (2024). How to use CRISPR/Cas9 in plants: From target site selection to DNA repair. *Journal of Experimental Botany, 75*(17), 5325–5343. doi:10.1093/jxb/erae147

Sturme, M. H. J., van der Berg, J. P., Bouwman, L. M. S., De Schrijver, A., de Maagd, R. A., Kleter, G. A., & Battaglia-de Wilde, E. (2022). Occurrence and nature of off-target modifications by CRISPR-Cas genome editing in plants. *ACS Agricultural Science & Technology, 2*(2), 192–201. doi:10.1021/acsagscitech.1c00270

Tang, X., Liu, G., Zhou, J., Ren, Q., You, Q., Tian, L., Xin, X., Zhong, Z., Liu, B., Zheng, X., Zhang, D., Malzahn, A. A., Qi, Y., Zhang, T., & Zhang, Y. (2018). A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 nucleases in rice. *Genome Biology, 19*, 84. doi:10.1186/s13059-018-1458-5

Tsai, S. Q., & Joung, J. K. (2016). Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nature Reviews Genetics, 17*(5), 300–312. doi:10.1038/nrg.2016.28

Wang, X., Tu, M., Wang, Y., Yin, W., Zhang, Y., Wu, H., Gu, Y., Li, Z., Xi, Z., & Wang, X. (2021). Whole-genome sequencing reveals rare off-target mutations in CRISPR/Cas9-edited grapevine. *Horticulture Research, 8*, 114. doi:10.1038/s41438-021-00549-4

Yang, H., Wu, J.-J., Tang, T., Liu, K.-D., & Dai, C. (2017). CRISPR/Cas9-mediated genome editing efficiently creates specific mutations at multiple loci using one sgRNA in Brassica napus. *Scientific Reports, 7*, 7489. doi:10.1038/s41598-017-07871-9

Young, J., Zastrow-Hayes, G., Deschamps, S., Svitashev, S., Zaremba, M., Acharya, A., Paulraj, S., Peterson-Burch, B., Schwartz, C., Djukanovic, V., Lenderts, B., Feigenbutz, L., Wang, L., Alarcon, C., Siksnys, V., May, G., Chilcoat, N. D., & Kumar, S. (2019). CRISPR-Cas9 editing in maize: Systematic evaluation of off-target activity and its relevance in crop improvement. *Scientific Reports, 9*, 6729. doi:10.1038/s41598-019-43141-6

Zhan, X., Lu, Y., Zhu, J.-K., & Botella, J. R. (2021). Genome editing for plant research and crop improvement. *Journal of Integrative Plant Biology, 63*(1), 3–33. doi:10.1111/jipb.13063

Zhang, H., Zhang, J., Wei, P., Zhang, B., Gou, F., Feng, Z., Mao, Y., Yang, L., Zhang, H., Xu, N., & Zhu, J.-K. (2014). The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnology Journal, 12*(6), 797–807. doi:10.1111/pbi.12200

Zhang, N., Roberts, H. M., Van Eck, J., & Martin, G. B. (2020). Generation and molecular characterization of CRISPR/Cas9-induced mutations in 63 immunity-associated genes in tomato reveals specificity and a range of gene modifications. *Frontiers in Plant Science, 11*, 10. doi:10.3389/fpls.2020.00010

Zhang, Q., Xing, H.-L., Wang, Z.-P., Zhang, H.-Y., Yang, F., Wang, X.-C., & Chen, Q.-J. (2018). Potential high-frequency off-target mutagenesis induced by CRISPR/Cas9 in Arabidopsis and its prevention. *Plant Molecular Biology, 96*(4–5), 445–456. doi:10.1007/s11103-018-0709-x

Appendix 1. Search Documentation Table

Database	Platform/interface	Date searched	Exact search string used	Records retrieved
Scopus	Scopus	12.05.2026	TITLE-ABS-KEY ((CRISPR OR "CRISPR/Cas9" OR Cas9) AND ("off-target" OR "off target" OR specificity) AND (plant OR crop OR Arabidopsis OR rice OR maize OR wheat OR tomato OR cotton OR grapevine)) AND PUBYEAR > 2012 AND LANGUAGE(english)	1,304
Web of Science	Web of Science Core Collection	12.05.2026	(CRISPR OR "CRISPR/Cas9" OR Cas9) AND ("off-target" OR specificity OR "unintended mutation") AND (plant OR crop OR Arabidopsis OR rice OR cotton OR grapevine) with English and publication date from 01.01.2013 to May 2026	1,049
PubMed	PubMed	12.05.2026	((CRISPR[Title/Abstract] OR "CRISPR/Cas9"[Title/Abstract] OR Cas9[Title/Abstract]) AND ("off-target"[Title/Abstract] OR specificity[Title/Abstract]) AND (plant[Title/Abstract] OR Arabidopsis[Title/Abstract] OR rice[Title/Abstract] OR crop[Title/Abstract]))	550
ScienceDirect	ScienceDirect	12.05.2026	“CRISPR Cas9 off-target plants”, “CRISPR/Cas9 specificity plants”, “CRISPR Cas9 off-target rice”, “CRISPR/Cas9 specificity Arabidopsis”, and “plant genome editing off-target mutation”	>30,000 total

Google Scholar	Google Scholar	12.05.2026	“CRISPR Cas9 off-target plants”, “CRISPR/Cas9 specificity plants”, “CRISPR Cas9 off-target rice”, “CRISPR/Cas9 specificity Arabidopsis”, and “plant genome editing off-target mutation”	>100,000 total
University of Skövde library	Library discovery tool	12.05.2026	“CRISPR Cas9 off-target plants”, “CRISPR/Cas9 specificity plants”, “CRISPR Cas9 off-target rice”, “CRISPR/Cas9 specificity Arabidopsis”, and “plant genome editing off-target mutation”; English, >2012, Research article	1,205

For formal record accounting, only Scopus, Web of Science, PubMed, and the University of Skövde library discovery tool were included in the screening totals, giving 4,108 identified records. ScienceDirect and Google Scholar were searched as supplementary sources only because their result sets were too broad for reliable inclusion in the formal flow totals.