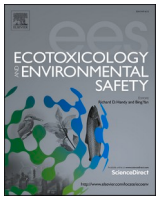
Ecotoxicology and Environmental Safety 282 (2024) 116707

Contents lists available at ScienceDirect

Ecotoxicology and Environmental Safety

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



Predicted multispecies unintended effects from outdoor genome editing

Aline Martins Hoepers [a], Jack A. Heinemann [b], Caroline Bedin Zanatta [a], Philomena Chu [c], Tessa C. Hiscox [b], Sarah Zanon Agapito-Tenfen [c,*]

[a] Crop Science Department, Federal University of Santa Catarina, Florianópolis, Brazil
[b] Centre for Integrated Research in Biosafety and School of Biological Sciences, University of Canterbury, Christchurch, New Zealand
[c] NORCE Norwegian Research Centre AS, Climate & Environment Division, Siva Innovasjonssenter, Sykehusvn 21, Tromsø 9019, Norway

ARTICLE INFO

Edited by Bing Yan

Keywords:
CRISPR/Cas
Gene-editing
Risk assessment
Genetically modified
Non-target organisms

ABSTRACT

CRISPR/Cas9, a potent genetic engineering tool widely adopted in agriculture, is capable of introducing new characteristics into plants on a large scale and without conventional breeding methods. Despite its remarkable efficiency, concerns have arisen regarding unintended consequences in uncontrolled environments. Our aim was to assess potential activity in organisms that could be exposed to genome editing in uncontrolled environments. We developed three scenarios, using irrigation, fumigation and fertilization as delivery methods, based on outdoor uses in agriculture, namely pest and disease control. Using publicly available software (Cas-OFFinder, NCBI Genome Data Viewer and STRING), off-target effects were predicted in multiple species commonly found in the agroecosystem, including humans (16 of 38 (42 %) sampled). Metabolic enrichment analysis (gene IDs), by connecting off-target genes into a physiological network, predicted effects on the development of nervous and respiratory systems. Our findings emphasize the importance of exercising caution when considering the use of this genome editing in uncontrolled environments. Unintended genomic alterations may occur in unintended organisms, underscoring the significance of understanding potential hazards and implementing safety measures to protect human health and the environment.

1. Introduction

Gene/genome editing describes a family of approaches for making site-directed changes in target DNA or RNA polymers (Segal and Meckler, 2013). Usually included in this category are oligonucleotide mutagenesis methods (ODM), TALENs, zinc-finger nucleases (ZFNs), and CRISPR/Cas systems.

Except for ODM, genome editing involves a protein-nucleic acid interaction that results in the change of the nucleic acid at a predefined location. TALENs and ZFNs are proteins with engineered DNA binding domains that confer site-specificity that concentrates an embedded endonuclease activity. CRISPR/Cas uses small oligonucleotides (usually but not exclusively RNA) to guide the Cas nuclease to a predetermined site.

Genome editing makes it easier to manipulate genes within large genomes, such as those of plants, fungi, and animals, and to create null mutations (e.g., gene knockouts). Further refinements now make it possible to also change bases at predetermined sites on a DNA or RNA molecule (Tao et al., 2023). These gene technology efficiency improvements contribute to expectations of developing more productive food organisms and sustainable agroecosystems (Shi et al., 2017; Zhang et al., 2014; Wang et al., 2015).

Genome editing and other gene technology tools are being developed for open air applications, that is, outside of a laboratory (Heinemann and Walker, 2019). The low efficiency of these approaches outside of a laboratory is presently limiting but may improve (Demirer et al., 2021; Vats et al., 2022). Proposed use in uncontrolled environments would also be compatible with deregulation of some applications of genome editing. Deregulation takes a number of forms, including adopting a tiered regulatory framework that defines some uses of genome editing as out of scope with any existing gene technology regulations (Heinemann, 2019; Heinemann et al., 2023). Reagents of out-of-scope techniques become potential environmental mutagens of emerging concern.

CRISPR/Cas is now the pre-eminent genome editing reagent because of its ease of use and the easy access to customized nucleic acid guides. The Cas nuclease is coupled with a crRNA molecule that metaphorically guides the nuclease to a target location in a DNA (e.g. Cas9) or RNA (e.g. Cas13) molecule (Sharma et al., 2022). In their natural context, guides

* Corresponding author.
E-mail address: sarahagapito@protonmail.com (S.Z. Agapito-Tenfen).

https://doi.org/10.1016/j.ecoenv.2024.116707
Received 21 November 2023; Received in revised form 2 July 2024; Accepted 7 July 2024
Available online 12 July 2024

are transcribed from a region called CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats). In a biotechnology context, CRISPR/Cas is a combination of a sequence agnostic Cas9 (or similar) endonuclease activity and a designed oligonucleotide ('guide', e.g. gRNA) that concentrates the endonuclease activity at sequences within the DNA or RNA polymeric substrate through hybridization (Jinek et al., 2012). The gRNA is desired to have a high affinity for a target and a low affinity for non-target binding sites.

The catalytic properties conferred by the specificities of hybridization between gRNA and target DNA/RNA sequence of nucleotides result in an increase in the rate of intended hydrolysis of the nucleic acid polymer in the substrate molecule at an intended location. Derivative of the idea that the guide nucleic acid molecule acts as a reaction catalyst, the strength of the bonding between guide and substrate is presumed to be proportional to the probability that the nuclease reaction will complete.

However, the strength of pairing between the guide and substrate does not describe a digital reaction, as is inherent in the concept of a 'lock and key' metaphoric model for enzymatic catalysts. The interaction space will tolerate mismatches and the number it tolerates will depend on cellular variables such as local ion concentrations and temperature.

Magnets are a more realistic than is the 'lock and key' metaphor for conferring specificity on the reaction. The strength of the interaction will vary between any given magnet and a series of substrates that describe a continuum of attraction potential. Therefore, the catalyst may interact with unintended sequences if given enough time or if in excess concentration. However, the catalyst may act on these secondary targets only after acting on the primary target or if the primary target is saturated with the editing complex, causing the impression of a precise reaction, i.e. off-target effects (Brinkman et al., 2018; Zhang et al., 2018).

Site-specific protein-nucleic acid interactions are also a feature of RNA interference (RNAi), which is associated with gene silencing. RNAi relies on RNA:RNA (sometimes RNA:DNA) hybridization to impart sequence-guided argonaut protein interactions. Small guide RNAs (e.g. siRNA) derived from double-stranded RNA (dsRNA) have an equivalent role to gRNA for CRISPR/Cas.

At efficiencies of effectiveness in uncontrolled environments that meet commercial or other objectives, a fundamentally new challenge for risk assessment will emerge (Li et al., 2021; Mueller, 2021). When designing site-directed biotechnological tools - whether or not the protein relies on a nucleic acid guide - predicted sites of action are firstly, sometimes exclusively, informed by *in silico* searches for nucleotide sequences in known DNA or RNA molecules that may be alternative substrates from those intended.

No in silico search-and-design approach has met the scientific standards necessary to validate in silico methodology as a substitute for in vivo direct testing for off-target effects of site-directed tools (Hanning et. al., 2013). In a recent report on CRISPR/Cas9 use in zebrafish (Höijer et al., 2022), the authors found insertions and deletions of sizes up to several kilobases, at on- and off-target sites, which were not predicted by available computational tools.

In vivo testing of HPV-negative squamous epithelial cells (normal cervix NCx and skin HaCaT cells) transfected with therapeutic siRNAs also showed no correlation between the number of computationally predicted off-target sites and the actual number of off-target mutations (Hanning et al., 2013). On average, only 20.5 % of actual transcriptional off-target sites were computationally predicted whereas unpredicted off-targets included stimulation of innate immune pathways, as well as indirect (downstream) effects of other off-target genes, which affected important cancer-associated pathways. In addition, only 0–5.9 % of differentially expressed genes overlapped between the two cell types, thus showing strong cell-dependent effects.

In another study, plants imbibed water with dsRNA molecules that were insecticidal to hoppers. Consumption of exposed hoppers by spiders caused gene silencing in spider midgut cells. Drinking the plant guttation water caused systemic gene silencing in the spider.

Nevertheless, in silico surveys are useful to predetermine sites that could be of interest and prioritize them for risk assessment. A useful survey will likely generate more false positives than negatives, but when biosafety is the priority, minimizing false positives is not. The in silico tools used to create and troubleshoot unintended effects are only superficially different depending on whether the intention is to create siRNAs or gRNAs. And whether or not the intention behind the use of the RNA as siRNA or gRNA, the molecules can sometimes trigger either pathway (Sharma et al., 2022).

Risk assessment depends on a robust hazard identification process. Hazards only exist in relation to something. Regulatory risk assessment is therefore a process of describing what is to be protected (a protection goal) and formulating hypotheses as to how a hazard adversely affects the protection goal.

We present an *in silico* analysis and a discussion of unintended effects in non-target species through which CRISPR/Cas open air application technology could cause adverse effects. At first it may seem confusing that in a paper on genome editing we chose to model hypothetical gRNAs based on dsRNA molecules developed for producing siRNAs that induce RNAi. However, while the details of the biochemistry of RNAi differ from CRISPR/Cas, the specificities and limits of base-pairing are thought to be necessary if not sufficient to confer target "precision" in both techniques.

For our *in silico* exercise, we have assumed that the chosen gRNAs were developed based on efficacy for site-directed interaction *in vivo*. They would also have benefitted from some exercise to ensure that they were active only on the intended target nucleotide sequence.

In silico tools are accepted as a starting point for identifying high quality genome editing and gene silencing guides often with the suggestion that their specificity will result in high reaction efficiency for intended effects. The latter is referred to as precision. This study examines the other side of precision. That is, do the same *in silico* tools also predict that the guides will not catalyze reactions at locations for which they were not intended?

2. Material and methods

2.1. Description of scenarios

We have performed an *in silico* analysis of potential unintended activities in exposed species to predict the potential ecological impacts of genome editing for pest and disease control. Three cases (plausible scenarios) have been chosen for the purpose of identifying plausible hypotheses of harm to selected protection goals from the use of genome editing to control an agricultural pest. In this context, we examined three target organisms hypothetically capable of harboring the CRISPR/Cas9 complex and a gene with an agronomic trait of interest. The selection of target organisms considers their natural environment and interactions with non-target organisms. For each scenario (Fig. 2) we considered the strategies currently employed to deliver agricultural products also applicable to deliver CRISPR/Cas9 complex solutions or pellets (Fig. 1), that would be plausible for the target insect to be exposed. The target genes used in this analysis were "borrowed" from those previously identified from others who used the same gene as a target in RNAi-based pest and disease control. The targeting of a gene by RNAi also involves hybridization.

In the first scenario, the Western corn rootworm (WCR), *Diabrotica virgifera virgifera* LeConte, was chosen as an insect target. WCR is an important pest in *Zea mays*, where its larvae feed on maize roots, leading to decreased plant growth and reduced yield (Meinke et al., 2009). The insect is considered one of the major pests in the United States (Sappington et al., 2006) and Europe (Miller et al., 2005), mainly because of field-evolved insecticide resistance (Meinke et al., 2021).

In the second scenario, the Red Flour Beetle, *Tribolium castaneum,* was chosen as an insect target. This beetle infests a large diversity of

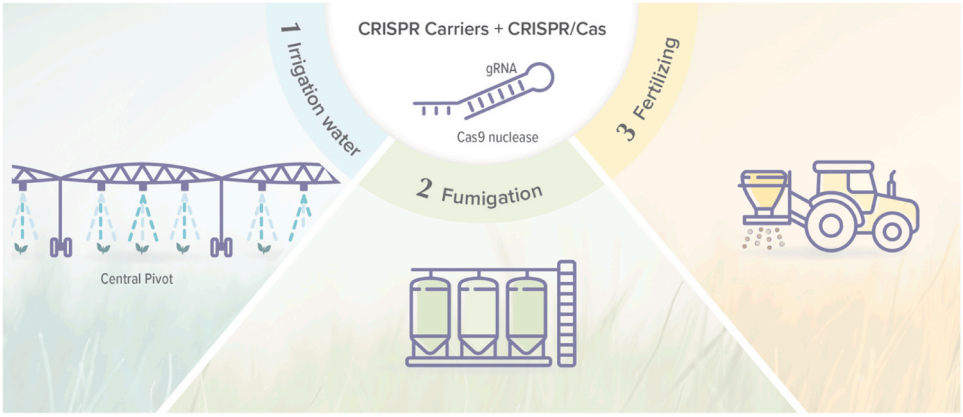


Fig. 1. Graphic presentation of application methods proposed for an open-air technology of CRISPR/Cas. In this study, we consider applications of CRISPR/Cas molecules through irrigation water (sprinkler and or center pivot irrigation) – Scenario #1, fumigation technology (filling an area with gaseous pesticides) – Scenario #2, and fertilizing strategies (direct applications of pellets into soil) – Scenario #3.

cereal flours (Agarwal and Agashe, 2020) and is considered a significant global pest of stored food products (Abd El-Aziz, 2011; Campbell et al., 2021). In addition, *T. castaneum* is considered a model beetle for several other species (Kumar et al., 2018).

In the third scenario, the necrotrophic fungus *Sclerotinia sclerotium* is the target species. This fungus is an major agricultural pathogen, and it can colonize over 400 crops (Bolton et al., 2006), such as *Glycine max, Phaseolus vulgaris, Gossypium hirsutum,* and *Brassica napus* (Boland, Hall, 1994). In favorable environments for the development of the disease, yield losses often surpass 20–35 %, although cases of over 50 % and as high as 80–100 % have been documented in various places, particularly in temperate climates (Alkooranee et al., 2019).

Open air application of CRISPR/Cas results in direct exposures to target and non-target organisms through contact, inhalation, or ingestion. Release of reagents such as aerosols or gels is intended to create contact exposures on target plants or pests. Non-target organisms, such as invertebrate pollinators, may also experience contact exposures. Both target and non-target animals may inhale the aerosol. Plant pests and pest predators may ingest the formulation by eating residues on the surface of exposed plants or prey.

Indirect exposures follow from ingestion of organisms that have been exposed and contain reagents in an active form within their cells. For example, pollinators could ingest pollen and insect predators could ingest pests that have internalized spray. For herbivores, invertebrates, farm animals and soil organisms, the route of exposure is leaves, roots and grains (Bachman et al., 2016).

2.2. Target gene selection and gRNA design

The Snf7 ortholog DvSnf7 gene was chosen for scenario #1 as the target for *Diabrotica virgifera virgifera* (Gene ID: 114337301) because it was previously described as a potential RNAi target for the control of WCR (Bolognesi et al., 2012). A 240 bp RNA transcript from the WCR Snf7 gene was developed as a commercial product. Its spontaneous folding produces a dsRNA that induces RNAi-mediated coleopteran resistance. The 240 bp dsRNA described by Bolognesi and colleagues (2012) was identified in the National Center for Biotechnology Information (NCBI) using the Basic Local Alignment Search Tool (BLAST), against *Diabrotica virgifera virgifera* (taxid:50390) genome.

For case 2, the conserved gene RpII140 (Gene ID: 663520) of *Tribolium castaneum* was selected as target. This gene is part of the DNA-directed RNA polymerase II subunit that catalyzes the transcription of DNA into RNA (Knorr et. al., 2018) and when used as a target for RNAi, silencing kills 88.89 % of exposed *T. castaneum* (Knorr et al., 2018).

For scenario 3, the SS1G_01703 - ABHYDROLASE-3 gene (Gene ID: 5493374) from *Sclerotinia sclerotiorum* gene, which was previously identified as a potential RNAi target (Wytinck et al., 2022), was chosen. This gene is predicted to be involved in aflatoxin biosynthesis and pathogenicity factor.

The gRNAs were designed on CRISPOR (Concordet and Haeussler, 2018) and were selected based on high GC% content (between 40 % and 70 %) and high out-of-frame scores (complete knockout efficiency on-target and specificity score). For each case, input sequences were partial sequences from the genes previously chosen. The full description of input sequences, along with their size and number of possible guide sequences, is available in Supplementary file #1. The number of mismatches at off-target sites in the target genome was also considered to avoid having binding sites in related genus and species regardless of their protection status. gRNAs with lower mismatches levels were selected to provide higher specificity of the gRNA sequences.

2.3. In silico prediction of unintended CRISPR binding

To identify possible unintended activities, the three gRNA sequences were analyzed by the Cas-OFFinder in CRISPR RGEN Tools (Park et al., 2015). CasOFFinder performs DNA searches based on reference genomes from selected species. Here, we selected species based on protection goals established for each scenario. This platform reads the input sequence data, which in our study were gRNA sequences designed for each case, and collects information about specific sites containing PAM sequences. These sequences are then delivered to a comparator, which counts the number of mismatched bases and selects potential off-target sites that have fewer mismatched bases than the parameters given (Bae et al., 2014). In this case, the parameters allowed up to four base mismatches and identical protospacer adjacent motifs (PAM) sites which in our case is 5′-NGG-3′from SpCas9 from *Streptococcus pyogenes* (Smith et al., 2020). CasOFFinder provides an output of potential off-target DNA sequences with mismatched bases, alongside the chromosome number, position, direction, and number of mismatched bases.

The output from Cas-OFFinder was then manually analyzed in the NCBI tool Genome Data Viewer (GDV), to identify matching DNA sequences (Rangwala et al., 2021). The input provided to GDV included chromosome, location, and organism analyzed. The software provides the exon navigator element, which allows the evaluation of whether the analyzed binding sequence corresponds to a gene, and if so, which gene. Putative binding sites that did not match any gene were labeled as “intergenic hits” (Rangwala et al., 2021).

A manual search for each sequence hit was also performed when different reference genome versions were available in the software used (i.e. STRING (see below) and GDV). The sequences that did not have any corresponding gene were not considered for further analysis. Finally, poorly annotated reference genomes also contributed to a lack of gene

Table 1
Description of each scenario investigated in this study.

CASE	Description	Target Species	Target Gene (Genebank ID)	Knock-Out Metabolism
1	Western corn rootworm infestation in maize fields	*Diabrotica virgifera virgifera*  © Russel IPM (source: russellipm.com)	Charged multivesicular body protein 4b (Gene ID: 114337301)	Essential protein involved in intracellular trafficking
2	Red flour beetle infestation in stored cereals	*Tribolium castaneum*  © Georg Goergen (source CABI CPC)	RpII140 (Gene ID: 663520)	Part of the DNA-directed RNA polymerase II subunit that catalyzes the transcription of DNA into RNA
3	White mold infection in soybean, common beans, and cotton fields	*Sclerotinia sclerotium* 	ABHYDROLASE-3 gene (Gene ID: 593374)	Predicted aflatoxin biosynthesis gene and pathogenicity factor

identification.

2.4. Metabolic enrichment of associated unintended binding sites

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) was used to put candidate unintended binding sites into a metabolic or physiological network context. STRING is an online tool that reports connections between proteins when a specific functional partnership may exist between them (Szklarczyk et al., 2023). This protein–protein association evidence is assessed, quantified, and compared against KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway maps (Kanehisa et al., 2021). KEGG computationally reconstructs biological systems by capturing information from data published in literature and data accumulated in the KEGG PATHWAY database (Kanehisa et al., 2021).

STRING output is a 'confidence score' for each protein association which scores from zero to one and indicates the estimated probability of a given association being true, based on the available evidence in KEGG databases. Confidence scores are first computed separately for each evidence type and then integrated to generate a final confidence score for a given metabolic pathway (Szklarczyk et al., 2023).

The input set of genes included the gene and intergenic hit list for each species in each case study. In total, 12 analyses in nine organisms were performed using STRING. Gene hits were enriched by the platform algorithm to connect metabolic networks and biochemical pathways for the analysis of the following categories: Biological Process (Gene Ontology), Molecular Function (Gene Ontology), Cellular Component (Gene Ontology), Reference publications (PubMed), Local network cluster (STRING), KEGG Pathways, Reactome Pathways, WikiPathways, Tissue expression (TISSUES), Human Phenotype (Monarch), Annotated Keywords (UniProt), Protein Domains (Pfam), Protein Domains and Features (InterPro), Protein Domains (SMART). For some species, disease-gene associations (DISEASES) category was also analyzed. Pathways with FDR < 0.05 were considered as significantly enriched.

3. Results

3.1. Predicted gRNA activity in target organisms

Three top ranked gRNA molecules were chosen for each scenario based on their location within one of the corresponding gene exons, met threshold GC% content and out-of-frame score, and a high MIT specificity score (few unintended hits) (Table 2). For the respective gRNAs selected in scenarios #1 and #3, no unintended binding sites were predicted in the target genomes of *Diabrotica virgifera virgifera* or *Sclerotinia sclerotium*. By allowing up to four mismatches, the top ranked selected gRNA in scenario #2 was predicted to have three unintended binding sites in the target genome of *Tribolium castaneum*. Of these, one unintended site was located in an intergenic region (NC_007424.3), one in an intronic region (NC_007422.5), and one in an exonic region (NC_007421.3) each.

3.2. Identification of protection goals and relevant exposures

We began the hazard identification step by defining our protection goals. Each of the three cases we model has an intended target organism and a receiving environment from which we could identify plausible, overlapping, protection goals (Table 3). Firstly, to be included the protection goal had to be a non-target organism (NTO) that provided an important ecological function in the intended receiving environment, which in our case is a farm agroecosystem. Secondly, a reasonable hypothesis could be made for the protection goal to be exposed to gene-modifying agents. Exposure could be direct or indirect but limited to first level food webs. Thirdly, we limited our study to the investigation of eukaryotes. Rising to the top of our list were people (e.g., the farmers and their families), on-farm animals (e.g., livestock and pets), crops, and beneficial organisms (pollinator, predator, soil invertebrates). See Supplementary File 2 for genomes and RNA reference databases used for these analyses.

The selection of species was constrained by the imbalance of species representation in publicly available databases. For example, *Eisenia andrei, Coleomegilla maculata, Poecilus chalcites* (soil organisms), *Coccinella septempunctata* (invertebrate predator) and *Parthenium hysterophorus and Brachiaria reptans* (wild plants that may also be weeds) are underrepresented. To overcome this limitation, we selected available related species with the same or similar ecological function for each case study (ie *Metarhizium anisoplia; Brassica sp., Sorghum bicolor, Panicum virgatum, Setaria viridis*).

3.3. Predicted gRNA activity sites in NTOs

By definition, any gRNA activity in an NTO is a kind of off-target

Table 2
gRNA design parameters for target organisms in the three case studies. gRNA relevant parameters include: the position in target gene (gene location and strand sense), in which exon the gRNA is located, CG contend (percentage of guanine and cytosine in gRNA sequence), out-of-frame score (0–100; prediction of likelihood of a gene knock-out). The last 5 columns are predicted unintended hits categorized by the number of nucleotide mismatches in relation to the gRNA sequence.

Scenario	Target and Gene ID	gRNA sequence (5′ to 3′)	Position (bp)/ strand	Exon	GC Contents (% without PAM)	Out-of-frame Score	Mismatches				
							0	1	2	3	4
#1	D.v.v Gene ID: 114337301	TGAATAGCCTCCCCAGGACT **AGG**	130/rev	1	55 %	67	0	0	0	0	0
#2	T.c Gene ID: 663520	GTCCCGAATATCCCGAATCA **TGG**	86/rev	2	50 %	55	0	0	0	0	3
#3	S.s Gene ID: 5493374	GGCGCATAACCCTCGGTTGA **TGG**	10/rev	1	60 %	65	0	0	0	0	0

Table 3
Ecological function and non-target species analyzed (indicated by dots) in each case. Organisms not included in a scenario are indicated by dashes.

NTO Category/Ecological function	NTO Species	Scenario #1 **Target** DvSnf7	Scenario #2 **Target** RpII140	Scenario #3 **Target** ABHYDROLASE-3
Farmer	*Homo sapiens* (human being)	•	•	•
Farm animals (grazers and omnivores)	*Bos taurus* (cattle)	•	•	•
	Gallus gallus (chicken)	•	•	•
Pest mammals (rodents)	*Mus musculus* (mouse)	•	•	•
Crops (photosynthetic organisms and nutrient cyclers)	*Zea mays* (maize)	•	•	-
	Glycine max (soybean)	•	-	•
	Phaseolus vulgaris (beans)	-	-	•
	Gossypium hirsutum (cotton)	-	-	•
Wild plants (photosynthetic organisms and nutrient cyclers)	*Brassica sp.*	•	-	•
	Sorghum bicolor	•	-	•
	Panicum virgatum	•	-	•
	Setaria viridis (green bristlegrass)	•	-	•
Pollinator insects	*Apis mellifera* (honey bee)	•	-	•
	Bombus terrestris (buff tailed bumblebee)	•	-	•
Predator invertebrates	*Camponotus floridanus* (Florida carpenter ant)	•	-	•
	Lasius niger (black ant)	•	-	•
Soil invertebrates (nutrient cyclers)	*Metarhizium anisoplia* (green muscardine fungus)	•	-	•
	Caenorhabditis elegans (roundworm)	•	-	•

activity because it is unintended. Off-target activity can be avoided or minimized by using gRNAs that have no potential hybridization activity anywhere in the NTO genomes, or by ensuring that NTOs are not exposed to gRNAs. The latter may not be an option for applications described in Fig. 2.

The likelihood that two organisms share a common DNA (and potential gRNA binding site) sequence is an inference of homology. Our hypothesis was that the longer ago any two species shared a common ancestor, the fewer potential conserved binding sites.

However, the length of evolutionary time between divergence of the NTOs that are protection goals, and the target species was not predictive of the likelihood of unintended hybridization. Relatedness cannot be relied upon as a predictor of susceptibility or resistance to unintended activities.

Despite the evolutionary distance between target (insect and fungi species) and non-target organisms (humans and cattle). No binding sites

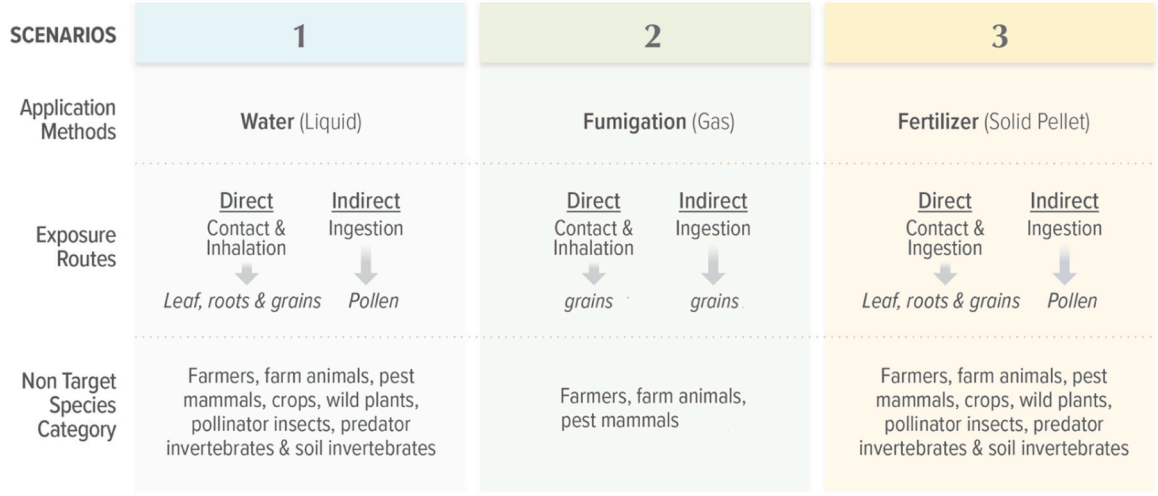


Fig. 2. Schematic representation of direct and indirect exposure routes for the non-target species selected for this study. For each scenario (top row), an application method (second row), exposure route (third row) and non-target exposure (bottom row) are illustrated. Ingestion may be direct (e.g., drinking contaminated nectar) or indirect (e.g., eating pollen that has internalized the reagents). Indirect exposure routes include the linkage via food webs.

were identified in the genomes of the predator and soil invertebrate genomes analyzed.

Absence of predicted unintended binding sites in invertebrates (or any particular species of any organism group) should be interpreted with caution. False negative results may arise from several factors. Firstly, it could be due to a small number of genomes from some NTO species in publicly available databases. Secondly, the known intra-species diversity of NTOs may vary. The *Homo sapiens* genome was last revised recently, but the *Caenorhabditis elegans* genome was last revised in 2013. Thirdly, *in silico* models cannot be informed for the many biochemical variables that could influence hybridization or other reaction parameters and are determinative of initiation or completion of an editing reaction. Therefore, an *in silico* analysis such as this may provide a false assurance of a few binding sites in protection goal species underrepresented in the databases.

The total number of predicted binding sites in NTO genomes is presented in Table 4. Out of 18 selected NTOs, we found potential un intended hybridization in 12 species, and these ranged from one (e.g., scenario #1 *Apis mellifera*) to 36 (scenario #2, *Homo sapiens*) sites, depending on the target organism. Where putative binding sites were found, they were in both genic and intergenic regions. We decided to report on intergenic unintended hybridization sites because they may later become associated with some function.

Potential binding sites were identified in nearly half (eight of 16) of the NTOs in scenario #1 (Table 5). In these eight species, the number of potential binding sites ranged from one in soybean and sorghum to 16 in

Table 4
Number of unintended off-target binding sites predicted in NTOs for the 3 scenarios.

Non-target organism (species name)	Scenario #1 (*Diabrotica virgifera virgifera)*	Scenario #2 (*Tribolium castaneum)*	Scenario #3 (*Sclerotinia sclerotium)*
Homo sapiens (humans)	24 unintended sites (8 intergenic regions)	36 unintended sites (15 intergenic regions)	5 unintended sites (2 intergenic regions)
Bos taurus (cattle)	19 unintended sites (11 intergenic regions)	3 unintended sites (1 intergenic region)	-
Gallus gallus (chicken)	9 unintended sites (5 intergenic regions)	-	-
Mus musculus (mouse)	29 unintended sites (17 intergenic regions)	4 unintended sites (1 intergenic region).	4 unintended sites (1 intergenic region)
Zea mays (maize)	4 unintended sites (2 intergenic regions)	1 unintended site (1 intergenic region)	-
Glycine max (soybean)	2 unintended sites (1 intergenic region)	-	1 unintended site (1 intergenic region)
Gossypium hirsutum (cotton)	-	-	1 unintended site (1 intergenic region)
Sorghum bicolor (millet)	3 unintended sites (2 intergenic regions	-	2 unintended sites
Panicum virgatum (switchgrass)	2 unintended sites	-	2 unintended sites (2 intergenic regions)
Setaria viridis (green foxtail)	2 unintended sitess (1 intergenic region)	-	-
Apis mellifera (honeybee)	-	-	1 unintended site
Caenorhabditis elegans (roundworm)	1 unintended site (1 intergenic region)	-	-

Table 5
Unintended hybridization sites for gRNA DvSnf7 in NTOs (Scenario #1).

Organism	Mismatches	Gene name	Gene ID
Homo sapiens	3	MTHFD1L - methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1 like	25902
Homo sapiens	4	EGF like, fibronectin type III and laminin G domains	133584
Homo sapiens	4	LOC105376737	105376737
Homo sapiens	4	neuron navigator 1	89796
Homo sapiens	4	cAMP responsive element binding protein 5	9586
Homo sapiens	4	FLII actin remodeling protein	2314
Homo sapiens	4	ADAMTS like 1	92949
Homo sapiens	4	zinc finger protein 462	58499
Homo sapiens	4	WHRN - whirlin	25861
Homo sapiens	4	NHS like 2	340527
Homo sapiens	4	Neurexin 3	9369
Homo sapiens	4	KHDRBS2 - KH RNA binding domain containing, signal transduction associated 2	202559
Homo sapiens	4	FYN proto-oncogene, Src family tyrosine kinase	2534
Homo sapiens	4	DENN domain containing 2B	6764
Homo sapiens	4	LOC105370324	105370324
Homo sapiens	4	RAB43, member RAS oncogene family	339122
Bos taurus	2	zinc finger protein, FOG family member 1	528338
Bos taurus	4	ALDH4A1 - aldehyde dehydrogenase 4 family member A1	100126042
Bos taurus	4	adrenoceptor alpha 1 A	282134
Bos taurus	4	ryanodine receptor 3	539899
Bos taurus	4	serine/threonine kinase like domain containing 1	614899
Bos taurus	4	rabphilin 3 A like (without C2 domains)	617294
Bos taurus	4	galactosidase beta 1	507188
Bos taurus	4	cytoplasmic polyadenylation element binding protein 3	537016
Gallus gallus	3	GTPase IMAP family member 7-like 2	101747807
Gallus gallus	4	cAMP responsive element binding protein 5	428435
Gallus gallus	4	RasGEF domain family member 1 A	423631
Gallus gallus	4	alpha kinase 1	100858173
Mus musculus	3	Arhgap15 Rho GTPase activating protein 15	76117
Mus musculus	3	Mall, T cell differentiation protein-like	228576
Mus musculus	3	myocardin	214384
Mus musculus	3	TSPO associated protein 1	207777
Mus musculus	4	solute carrier family 10 (sodium/bile acid cotransporter family), member 7	76775
Mus musculus	4	quiescin Q6 sulfhydryl oxidase 1	104009
Mus musculus	4	DENN domain containing 1 A	57706
Mus musculus	4	predicted gene, 30319	102632174
Mus musculus	4	utrophin	22288

(*continued on next page*)

Table 5 (*continued*)

Organism	Mismatches	Gene name	Gene ID
Mus musculus	4	death-associated protein kinase 3	13144
Mus musculus	4	cardiac mesoderm enhancer-associated non-coding RNA	328968
Zea mays	4	LOC100381711 putative WAK receptor-like protein kinase family protein	100381711
Zea mays	4	LOC100283564 transferase/ transferase, transferring glycosyl groups	100283564
Glycine max	4	uncharacterized LOC100801289	100801289
Panicum virgatum	4	LOC120681520	120681520
Panicum virgatum	4	LOC120687334	120687334
Sorghum bicolor	4	loc8082872	8082872

humans. The number of mismatches ranged from two (cattle) to four. Humans were the most likely to be affected by exposure to CRISPR/Cas irrigation water based on the number of potential hybridization sites.

Five NTOs were analyzed in scenario #2, *Homo sapiens, Bos taurus, Mus musculus, Gallus gallus,* and *Zea mays*. Potential binding sites were found in four of them (Table 6), ranging from three in the genome of *Mus musculus* to 21 in *Homo sapiens.* Potential binding sites in the *Homo sapiens* genome included proteins with antimicrobial activity, such as encoded by the defensin beta gene. Other possible binding sites in human genomes include the sarcoglycan delta gene which is produced in skeletal and cardiac muscle. Potential unintended binding sites in the *Mus musculus* genome were in genes involved in the functioning of the Golgi apparatus, ventricular layer, and nervous system development.

Potential binding sites were found in four of seventeen NTO genomes analyzed in scenario #3 (Table 7). They ranged from one found in the *Apis mellifera* genome and three in *Homo sapiens*. One potentially affected gene was matrix metalloproteinase-2 of *Apis mellifera.* This gene is involved in extracellular protein processing.

3.4. Predicted metabolic disturbances in exposed NTOs

A metabolic enrichment pathway analysis was used to predict possible effects of gRNA hybridizations in NTOs. Such analysis is based on assuming that gRNA hybridization at the identified complementary DNA sequences from above would result in loss of the annotated function of the genes through an unintended editing reaction. To begin to construct hypotheses of potential adverse effects from the mutation of these genes, we use KEGG and GenBank databases to retrieve information on metabolism, biological processes, and physiological responses after CRISPR/Cas exposure. From this a snapshot of the top ranked statistically significant metabolic pathways are described in the next sections. The confidence in such analysis is derived from statistical modeling of one-to-one gene and protein networks and the number of genes or proteins affected in a specific network.

Several gene hybridization sites (21) from the above sections were not found by string database search because they are mostly comprised of non-coding RNA (ncRNA), long non-coding (lcRNA), or pseudogenes, and have not been annotated. These genes and all output results can be found in Supplementary Files 3, 4 and 5.

3.5. Human (Homo sapiens)

Using KEGG predictions, the potential binding of the DvSnf7-targeting-gRNA to identified sites in the *Homo sapiens* genome could result in editing activity that altered the function of genes that affect 43 metabolic pathways in scenario #2 (Fig. 3). The RpII140-targeting-gRNA could direct nucleases to genes in three pathways.

Table 6
Unintended hybridizations of gRNA RpII140 in NTOs (Scenario #2).

Organism	Mismatches	Gene name	Gene ID
Homo sapiens	3	SUFU negative regulator of hedgehog signaling	51684
Homo sapiens	4	defensin beta 107B	503614
Homo sapiens	4	sarcoglycan delta	6444
Homo sapiens	4	solute carrier family 44 member 5	204962
Homo sapiens	4	LOC100419798 gon-4 like pseudogene	100419798
Homo sapiens	4	flavin containing dimethylaniline monoxygenase 9, pseudogene	116123
Homo sapiens	4	complement C3b/C4b receptor 1	1378
Homo sapiens	4	uncharacterized LOC122455341	122455341
Homo sapiens	4	endogenous retrovirus group K member 6 Env polyprotein	124901580
Homo sapiens	4	LOC105369787 endogenous retrovirus group K member 21 Env polyprotein	105369787
Homo sapiens	4	MIR548XHG (MIR548X host gene)	101927797
Homo sapiens	4	G protein subunit alpha q	2776
Homo sapiens	4	meiotic double-stranded break formation protein 4	101928601
Homo sapiens	4	uncharacterized LOC105376655	105376655
Homo sapiens	4	asparaginase and isoaspartyl peptidase 1	80150
Homo sapiens	4	LOC124902738 endogenous retrovirus group K member 25 Env polyprotein	124902738
Homo sapiens	4	LOC124902766 endogenous retrovirus group K member 7 Env polyprotein-like	124902766
Homo sapiens	4	uncharacterized LOC124904606	124904606
Homo sapiens	4	zinc finger protein 420	147923
Homo sapiens	4	LOC124906262 endogenous retrovirus group K member 5 Gag polyprotein	124906262
Homo sapiens	4	ectonucleotide pyrophosphatase/ phosphodiesterase 7 pseudogene 4	100421816
Bos taurus	4	chromosome 5 C12orf75 homolog	100270756
Bos taurus	4	mediator complex subunit 13 L	517336
Mus musculus	3	polypeptide N-acetylgalactosaminyltransferase 9	231605
Mus musculus	4	Mir9–3 hg	101694
Mus musculus	4	erythrocyte membrane protein band 4.1 like 3	13823

Table 7
Unintended gene hits for gRNA ABHYDROLASE-3 in NTOs (Scenario #3).

Organism	Mismatches	Gene name	Gene ID
Homo sapiens	4	zinc finger protein 92	168374
Homo sapiens	4	GDP-mannose 4,6-dehydratase	2762
Homo sapiens	4	Rho/Rac guanine nucleotide exchange factor 18	23370
Mus musculus	4	predicted gene 5364	384931
Mus musculus	4	synaptic vesicle glycoprotein 2c	75209
Mus musculus	4	tet methylcytosine dioxygenase 2	214133
Sorghum bicolor	4	LOC8082218 - protein GAMETE EXPRESSED 3	8082218
Sorghum bicolor	3	LOC8070878 - E4 SUMO-protein ligase PIAL2	8070878
Apis mellifera	4	LOC409383 matrix metalloproteinase-2	409383

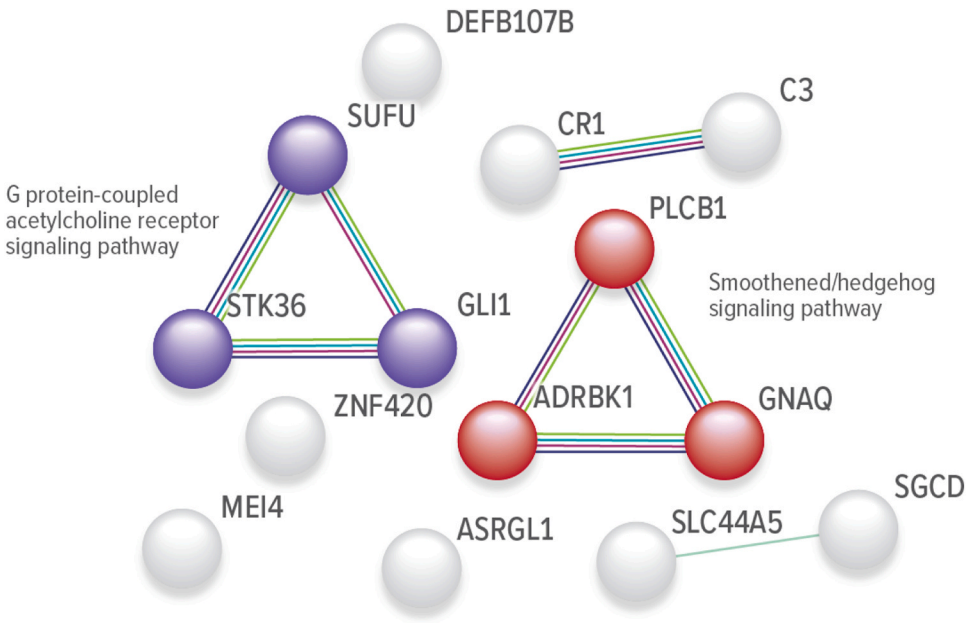


Fig. 3. Interaction network of proteins enriched by string database for scenario #2. Network nodes represent proteins and edges represent protein-protein associations. Blue lines represent known interactions from curated databases, while purple lines represent experimentally determined interactions. Green lines represent gene neighborhood predicted interactions.

Scenarios #2 and #3 describe outcomes where adverse effects include deafness, central nervous system cancer, nevoid basal cell carcinoma syndrome, and medulloblastoma. KEGG pathway enrichment analysis on putative binding sites of the gRNA illustrated in scenario #2 identify potential effects on pathways involved in hedgehog signaling, the glutamatergic synapse, cancer, endocrine system, and calcium reabsorption, among others (Table 8).

3.6. Cattle (*Bos taurus*) and chicken (*Gallus gallus*)

Metabolic enrichment for scenario #1 *Bos taurus* showed potential effects on 13 metabolic pathways (Fig. 4). Possible alterations occur in arginine and proline, alanine, aspartate, and glutamate metabolism (Table 9).

Metabolic enrichment based on genes potentially altered by editing in *Gallus gallus* identified effects on cardiac muscle contraction. The mitophagy pathway was also potentially affected. This pathway is connected to mitochondrial autophagy, a physiological process that contributes to the integrity of a well-functioning mitochondrial network and the prevention of programmed cell death by removing damaged mitochondria.

3.7. Mouse (*Mus musculus*)

Metabolic enrichment for *Mus musculus* scenario #2 identified four affected biological processes that could affect hemostasis, immune response, inflammation, embryogenesis, and development of neuronal tissue (Table 10).

3.8. Western honeybee (*Apis mellifera*)

Metabolic enrichment for *Apis mellifera* scenario #3 revealed 39 affected biological processes (Table 11). The affected pathways are associated with the development of the nervous system, heart respiratory system, larval heart, and the zygote to a multicellular animal.

3.9. Farm crops and wild plants (*Zea mays* and *Sorghum bicolor*)

Metabolic enrichment for *Zea mays* scenario #1 suggested that eight biological processes could be affected, including pathways needed for lipid biosynthesis. Lipid biosynthesis acts in cellular functions such as recognition, cell adhesion, proliferation, and differentiation. Modifications predicted for *Sorghum bicolor* scenario #3 include pathways for ubiquitin proteolysis, involved in cellular processes, and pathways associated with developmental processes involved with reproduction and embryo development (Table 12).

4. Discussion

The issue central to this work was the identification of hazards that might arise from the use of genome editing reagents outside of contained laboratory facilities. Such exposures would not be restricted to single species. We argue that if *in silico* tools are effective at predicting the target DNA sequences where the editing reactions will occur, then they are also useful for predicting some of the unintended binding sites when both target and non-target organisms are exposed.

We assessed the hypothesis that the application of the CRISPR/Cas9 formulated product in an open-air environment can impact other species in three realistic scenarios. Using these tools in uncontrolled environments has the potential to elevate genome editing (and gene silencing) to the class of emerging environmental contaminants.

The selected NTOs were identified as protection goals likely to be exposed. From the predicted potential for hybridization between a gRNA and complementary sequences of nucleotides in the genome of an NTO, many unintended genes could be affected (Tables 4–7). Neither the relatedness of the target to the NTO nor even the biological kingdom of the NTO were predictive of likelihood of unintended activity. Therefore, all species of concern may have to be specifically examined in a risk assessment.

Potential adverse effects were significant. Metabolic enrichment analysis was used to identify these effects on NTOs (Tables 8–12). Unintended activity could result in significant biological effects in all non-target plants and animals examined, such as the effects on immune responses, essential molecule biosynthesis, and the central nervous system. These results, consistent with existing literature on editing techniques, provide a clear rationale for the need to evaluate vulnerable NTOs in any proposed use of spray or topical techniques in the environment.

Table 8
Metabolic pathways ID; description; observed gene count; and false discovery rate for *Homo sapiens* in scenario #2.

Metabolic Pathway ID	Pathway description	Observed gene count	False discovery rate
hsa04340	Hedgehog signaling pathway	3	0.0023
GO:0007213	G protein-coupled acetylcholine receptor signaling pathway	3	0.0057
hsa04724	Glutamatergic synapse	3	0.01
hsa05142	Chagas disease	3	0.01
hsa04062	Chemokine signaling pathway	3	0.0311
hsa05143	African trypanosomiasis	2	0.0311
hsa05200	Pathways in cancer	4	0.0311
hsa04961	Endocrine and other factor-regulated calcium reabsorption	2	0.0381
GO:0007224	Smoothened signaling pathway	3	0.0408
hsa04071	Sphingolipid signaling pathway	2	0.041
hsa04540	Gap junction	2	0.041
hsa04713	Circadian entrainment	2	0.041
hsa04720	Long-term potentiation	2	0.041
hsa04725	Cholinergic synapse	2	0.041
hsa04726	Serotonergic synapse	2	0.041
hsa04730	Long-term depression	2	0.041
hsa04750	Inflammatory mediator regulation of TRP channels	2	0.041
hsa04911	Insulin secretion	2	0.041
hsa04912	GnRH signaling pathway	2	0.041
hsa04916	Melanogenesis	2	0.041
hsa04918	Thyroid hormone synthesis	2	0.041
hsa04922	Glucagon signaling pathway	2	0.041
hsa04924	Renin secretion	2	0.041
hsa04925	Aldosterone synthesis and secretion	2	0.041
hsa04927	Cortisol synthesis and secretion	2	0.041
hsa04928	Parathyroid hormone synthesis, secretion and action	2	0.041
hsa04929	GnRH secretion	2	0.041
hsa04935	Growth hormone synthesis, secretion and action	2	0.041
hsa04970	Salivary secretion	2	0.041
hsa04971	Gastric acid secretion	2	0.041
hsa04972	Pancreatic secretion	2	0.041
hsa05146	Amoebiasis	2	0.041
hsa05217	Basal cell carcinoma	2	0.041
hsa04611	Platelet activation	2	0.0416
hsa04728	Dopaminergic synapse	2	0.0442
hsa04270	Vascular smooth muscle contraction	2	0.0448
hsa04371	Apelin signaling pathway	2	0.0448
hsa04915	Estrogen signaling pathway	2	0.0448
hsa05017	Spinocerebellar ataxia	2	0.0448
hsa04723	Retrograde endocannabinoid signaling	2	0.0487
hsa04261	Adrenergic signaling in cardiomyocytes	2	0.0488
hsa04921	Oxytocin signaling pathway	2	0.0488
hsa04934	Cushing syndrome	2	0.0436

In silico tools are useful for constructing plausible hypotheses of potential adverse effects in NTOs but are not validated because the frequency of false negatives is unknown. Their use provides the *minimum* number of unintended sites that should be examined.

NTOs have historically not been a risk focus because genetic engineering was performed on the intended organism in a laboratory that minimized the potential for NTOs to be exposed to the gene-modifying procedures. Deregulation of the procedures as is being proposed in some emerging risk frameworks (Kenward, 2023; Heinemann, 2023; European Commission, 2023) would make it possible to use them in either built or open environments that do not control for exposures or for the release of organisms modified following an unintended exposure.

Externally applied reagents that facilitate the penetration of cells for the delivery of genome editing or gene silencing reagents are gaining attention because of their potential to enhance pest management, reduce toxicity of exposures associated with conventional pesticides, and provide benefits to consumers (Heinemann, 2019). The report of the Ad Hoc Technical Expert Group on Synthetic Biology (Secretariat of the convention on biological diversity, 2019) identified topical penetration agents for use in genome editing or (RNAi as amongst several new synthetic biology applications. It is worth noting that due to the rapid advancements in the field, future synthetic biology organisms may fall outside the existing definition of a "living modified organism" as outlined in the Protocol. This raises concerns regarding the assessment and monitoring of such organisms and highlights the need for fluid and evolving GMO risk assessment definitions (Li et al., 2021; Heinemann et al., 2023).

The *in silico* predictions presented in this study provide an overview into the potential effects on organisms inadvertently exposed during the use of topically applied reagents to cause genome editing or RNAi. However, these lists are not exhaustive for several reasons.

Firstly, we based our analyses only on organisms with sequenced and annotated genomes in publicly available databases. Unfortunately, only some of the many protection goals that would inhabit the places where these techniques would be used were represented in the databases.

Small deviations in the activity of keystone species can cause cascading effects. If these species are under-represented, the hazard assessment is compromised. Take, for instance, the role of earthworms in pastures. Earthworms contribute to ecosystem services such as nutrient cycling, soil structure formation, soil quality improvement, and water regulation (Blouin et al., 2013; Fonte et al., 2023). In New Zealand, which is a pasture-based cattle agroecosystem, earthworm activity is valued at $365–440/ha annually (Schon, Gray, Mackay, 2016). A reduction in earthworm activity does not have to stem from the loss of a particular gene but simply from the burden of mutagenesis due to repeat exposures to genome editing or silencing reagents (Miller et al., 1999; Heinemann and Billington, 2004). Changes in earthworm activity at the microscale can translate into large effects at the macroscale, particularly on cattle productivity.

Secondly, the target and non-target organisms considered in this study were limited to eukaryotes. If we were to expand our analysis to include prokaryotes, then a greater depth and breadth of protection goals would need to be considered.

Thirdly, the effects of gene technology used in uncontrolled environments are not limited to unintended changes in NTOs from intended activity. A gRNA designed for a genome editing reaction can upon release into the environment and uptake by cells of either target or NTOs become a reagent for an unintended RNAi or gene editing reaction (Sharma et al., 2022). Furthermore, the co-formulants that will be used with the genome editing or gene silencing reagents will also be potential contaminants due to a direct toxicity or their inherent capacity to capture and transport cargo molecules into cells.

5. Conclusions

Based on the use of the best available *in silico* tools we predict that the use of genome editing in a context where the intended organism is modified *in situ* in the environment will result in unintended changes in the genomes of NTOs. These organisms may provide important ecological functions or be protection goals for other reasons, or they may be undesirable organisms. Whether the NTOs are desired or not, the consequences of modifying them remain unpredictable because of the large number of unintended modifications. gRNAs activity was observed in 12 out of the 18 species of NTOs investigated in this study. These hybridization sites revealed genes with functions in several annotated

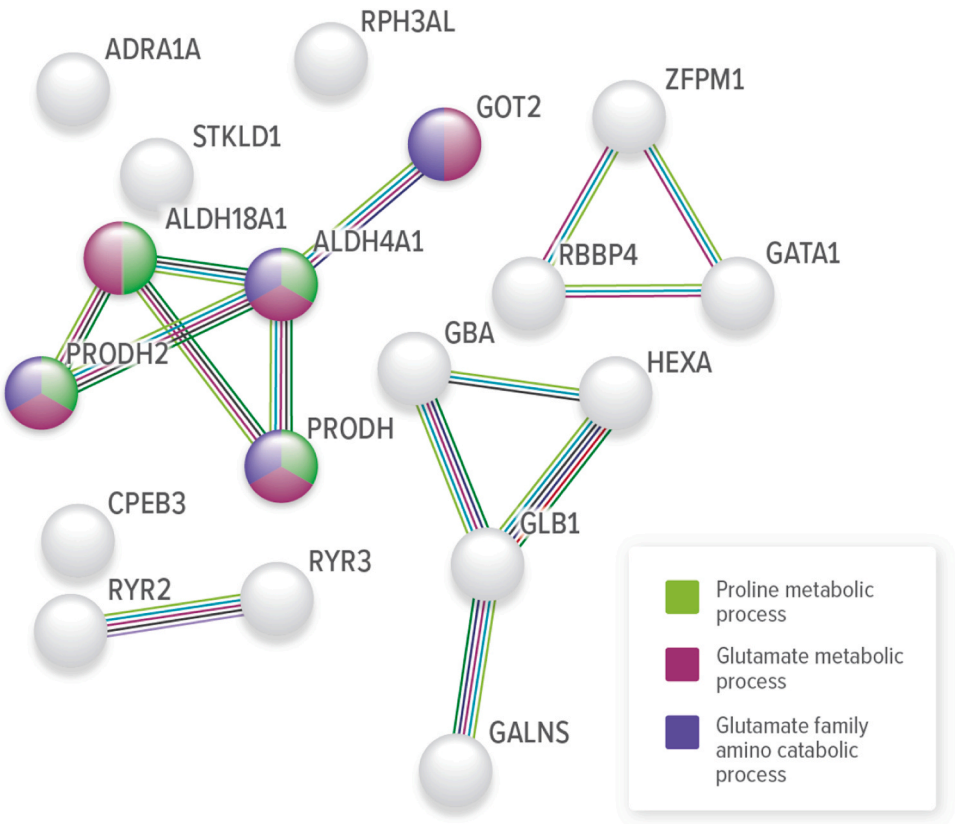


Fig. 4. Interaction network of proteins enriched by string database for *Bos taurus* scenario #1. Network nodes represent proteins and edges represent protein-protein associations. Blue lines represent known interactions from curated databases, while purple lines represent experimentally determined interactions. Green lines represent gene neighborhood predicted interactions.

Table 9
Metabolic pathway ID; process description; observed gene count; and false discovery rate for *Bos taurus* and *Gallus gallus* in case #1.

Metabolic pathway ID	Process description	Observed gene count	False discovery rate
Bos taurus			
bta00330	Arginine and proline metabolism	5	4.52E-08
GO:0006536	Glutamate metabolic process	5	5.56E-08
GO:0006560	Proline metabolic process	4	6.32E-07
GO:0009065	Glutamine family amino acid catabolic process	4	2.68E-06
GO:0010133	Proline catabolic process to glutamate	3	8.44E-06
GO:0010724	Regulation of definitive erythrocyte differentiation	2	0.0023
GO:0071733	Transcriptional activation by promoter-enhancer looping	2	0.0035
bta01100	Metabolic pathways	6	0.0232
GO:0030220	Platelet formation	2	0.0254
GO:0030219	Megakaryocyte differentiation	2	0.0297
GO:0030851	Granulocyte differentiation	2	0.0297
GO:0035162	Embryonic hemopoiesis	2	0.0297
bta00250	Alanine, aspartate and glutamate metabolism	2	0.0332
Gallus gallus			
gga04261	Adrenergic signaling in cardiomyocytes	3	0.0026
gga04137	Mitophagy - animal	2	0.0273

Table 10
Metabolic pathway ID; process description; observed gene count; and false discovery rate for *Mus musculus* in scenario #2.

Metabolic pathway id	Process description	Observed gene count	False discovery rate
GO:0071205	Protein localization to juxtaparanode region of axon	3	1.41E-05
mmu04514	Cell adhesion molecules	3	0.0041
GO:0002175	Protein localization to paranode region of axon	2	0.0078
GO:0030913	Paranodal junction assembly	2	0.01

metabolism, from central nervous system morphogenesis in honeybee to several pathways related to cancer and hormone metabolism in humans. In total, 155 metabolic pathways were enriched for the three gRNA scenarios in the 12 species with the majority of hits in the human genome. For proposed uses of the reagents of genetic engineering outside of contained laboratories, new risk assessment frameworks are needed (Heinemann, 2019) and should also include the assessment of their environmental persistence. The work here justifies their inclusion in legal instruments that manage the risk of gene technology and the impact of gene technology reagents in the context of emerging contaminants and potential hazards, as threats to human and environmental health.

CRediT authorship contribution statement

Tessa Hiscox: Writing – review & editing, Writing – original draft,

Table 11
Metabolic pathway ID; process description; observed gene count; and false discovery rate for *Apis mellifera* in scenario #3.

Metabolic pathways id	Process description	Observed gene count	False discovery rate
GO:0030198	Extracellular matrix organization	5	4.07E-06
GO:0030574	Collagen catabolic process	3	8.59E-05
GO:0007275	Multicellular organism development	9	0.00065
GO:0003344	Pericardium morphogenesis	2	0.002
GO:0016043	Cellular component organization	9	0.002
GO:0016340	Calcium-dependent cell-matrix adhesion	2	0.002
GO:0090129	Positive regulation of synapse maturation	2	0.0022
GO:0007155	Cell adhesion	4	0.0024
GO:0006930	Substrate-dependent cell migration, cell extension	2	0.0031
GO:0007419	Ventral cord development	3	0.0031
GO:0007508	Larval heart development	2	0.0031
GO:0007229	Integrin-mediated signaling pathway	2	0.0035
GO:0060560	Developmental growth involved in morphogenesis	3	0.0058
GO:0007399	Nervous system development	6	0.0062
GO:0048731	System development	7	0.0065
GO:0030182	Neuron differentiation	5	0.0088
GO:0042330	Taxis	4	0.0088
GO:0051492	Regulation of stress fiber assembly	2	0.0088
GO:0021551	Central nervous system morphogenesis	2	0.0096
GO:0034446	Substrate adhesion-dependent cell spreading	2	0.0096
GO:0007157	Heterophilic cell-cell adhesion via plasma membrane cell adhesion molecules	2	0.0153
GO:0016339	Calcium-dependent cell-cell adhesion via plasma membrane cell adhesion molecules	2	0.0153
GO:0048808	Male genitalia morphogenesis	2	0.0153
GO:0007015	Actin filament organization	3	0.0163
GO:0030336	Negative regulation of cell migration	2	0.0163
GO:0071711	Basement membrane organization	2	0.0173
GO:0033627	Cell adhesion mediated by integrin	2	0.018
GO:0035162	Embryonic hemopoiesis	2	0.018
GO:0010669	Epithelial structure maintenance	2	0.0187
GO:0007494	Midgut development	2	0.0215
GO:0007444	Imaginal disc development	4	0.0217
GO:0048666	Neuron development	4	0.0292
GO:0009987	Cellular process	11	0.0321
GO:0045214	Sarcomere organization	2	0.0324
GO:0060541	Respiratory system development	3	0.034
GO:0016203	Muscle attachment	2	0.036
GO:0048513	Animal organ development	5	0.0438
GO:0007411	Axon guidance	3	0.0486
GO:0007608	Sensory perception of smell	2	0.0486

Table 12
Metabolic pathways ID; process description; observed gene count; and false discovery rate for *Zea mays* and *Sorghum bicolor*.

Metabolic Pathway ID	Process description	Observed gene count	False discovery rate
Zea mays			
zma00603	Glycosphingolipid biosynthesis - globo and isoglobo series	6	2.19E-18
zma00604	Glycosphingolipid biosynthesis - ganglio series	5	8.08E-16
zma00531	Glycosaminoglycan degradation	5	2.37E-14
zma00511	Other glycan degradation	5	2.94E-14
zma00513	Various types of N-glycan biosynthesis	5	6.72E-13
zma00520	Amino sugar and nucleotide sugar metabolism	5	3.79E-10
zma01100	Metabolic pathways	6	6.17E-06
GO:0005975	Carbohydrate metabolic process	5	0.0051
Sorghum bicolor			
GO:0016925	Protein sumoylation	4	5.08E-08
sbi04120	Ubiquitin mediated proteolysis	3	0.00027
sbi03013	RNA transport	3	0.00027
GO:0003006	Developmental process involved in reproduction	4	0.0141
GO:0009793	Embryo development ending in seed dormancy	3	0.0141
GO:0009737	Response to abscisic acid	3	0.0178

Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Philomena Chu:** Writing – review & editing, Writing – original draft, Conceptualization. **Sarah Zanon Agapito-Tenfen:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Aline Martins Hoepers:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Formal analysis, Data curation. **Caroline Bedin Zanatta:** Writing – review & editing, Supervision, Software, Methodology, Formal analysis, Data curation. **Jack A. Heinemann:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data is available in Supplementary files.

Acknowledgements

The authors would like to thank Suet Chan for the art creation of all three graphics. We are also thankful to The Research Council of Norway for their funding support (Project Number: 301911—FOODPRINT: Traceability and labeling of gene-edited products in the food chain).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ecoenv.2024.116707.

References

Abd El-Aziz, S.E., 2011. Control strategies of stored product pests. J. Entomol. 8 (2), 101–122. https://doi.org/10.3923/je.2011.101.122.

Agarwal, A., Agashe, D., 2020. The red flour beetle Tribolium castaneum: a model for host-microbiome interactions. PLoS ONE 15 (10 October). https://doi.org/10.1371/journal.pone.0239051.

Alkooranee, J.T., Aledan, T.R., Ali, A.K., Lu, G., Zhang, X., Wu, J., Fu, C., Li, M., 2019. Detecting the hormonal pathways in oilseed rape behind induced systemic resistance by Trichoderma harzianum TH12 to Sclerotinia Sclerotiorum. PLoS ONE 12 (1). https://doi.org/10.1371/journal.pone.0168850.

Bachman, P.M., Huizinga, K.M., Jensen, P.D., Mueller, G., Tan, J., Uffman, J.P., Levine, S.L., 2016. Ecological risk assessment for DvSnf7 RNA: a plant-incorporated protectant with targeted activity against western corn rootworm. Regul. Toxicol. Pharmacol. 81, 77–88. https://doi.org/10.1016/j.yrtph.2016.08.001.

Bae, S., Kweon, J., Kim, H.S., Kim, J.S., 2014. Microhomology-based choice of Cas9 nuclease target sites. In: In Nature Methods, Vol. 11. Nature Publishing Group, pp. 705–706. https://doi.org/10.1038/nmeth.3015.

Blouin, M., Hodson, M.E., Delgado, E.A., Baker, G., Brussaard, L., Butt, K.R., Dai, J., Dendooven, L., Peres, G., Tondoh, J.E., Cluzeau, D., Brun, J.J., 2013. A review of earthworm impact on soil function and ecosystem services. Eur. J. Soil Sci. Vol. 64 (Issue 2), 161–182. https://doi.org/10.1111/ejss.12025.

Boland, G.J., Hall, R., 1994. Index of plant hosts of sclerotinia sclerotiorum. Can. J. Plant Pathol. 16 (2), 93–108. https://doi.org/10.1080/07060669409500766.

Bolognesi, R., Ramaseshadri, P., Anderson, J., Bachman, P., Clinton, W., Flannagan, R., Ilagan, O., Lawrence, C., Levine, S., Moar, W., Mueller, G., Tan, J., Uffman, J., Wiggins, E., Heck, G., Segers, G., 2012. Characterizing the mechanism of action of double-stranded RNA activity against Western Corn Rootworm (Diabrotica virgifera virgifera LeConte). PLoS ONE 7 (10). https://doi.org/10.1371/journal.pone.0047534.

Bolton, M.D., Thomma, B.P.H.J., Nelson, B.D., 2006. Sclerotinia sclerotiorum (Lib.) de Bary: biology and molecular traits of a cosmopolitan pathogen. Mol. Plant Pathol. Vol. 7 (Issue 1), 1–16. https://doi.org/10.1111/j.1364-3703.2005.00316.x.

Brinkman, E.K., Chen, T., de Haas, M., Holland, H.A., Akhtar, W., van Steensel, B., 2018. Kinetics and fidelity of the repair of Cas9-induced double-strand DNA breaks. Mol. Cell 70 (5), 801–813.e6. https://doi.org/10.1016/j.molcel.2018.04.016.

Campbell, J.F., Athanassiou, C.G., Hagstrum, D.W., & Zhu, K.Y. (2021). Annual Review of Entomology Tribolium castaneum: A Model Insect for Fundamental and Applied Research. https://doi.org/10.1146/annurev-ento-080921.

Concordet, J.P., Haeussler, M., 2018. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res. 46 (W1), W242–W245. https://doi.org/10.1093/NAR/GKY354.

Demirer, G.S., Silva, T.N., Jackson, C.T., Thomas, J.B., W. Ehrhardt, D., Rhee, S.Y., Mortimer, J.C., Landry, M.P., 2021. Nanotechnology to advance CRISPR–Cas genetic engineering of plants. Nat. Nanotechnol. 16 (3), 243–250. https://doi.org/10.1038/s41565-021-00854-y.

European Commission. (2023). Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on plants obtained by certain new genomic techniques and their food and feed, and amending Regulation (EU) 2017/625.

Fonte, S.J., Hsieh, M., Mueller, N.D., 2023. Earthworms contribute significantly to global food production. Nat. Commun. 14 (1) https://doi.org/10.1038/s41467-023-41286-7.

Hanning, J.E., Saini, H.K., Murray, M.J., van Dongen, S., Davis, M.P.A., Barker, E.M., Ward, D.M., Scarpini, C.G., Enright, A.J., Pett, M.R., Coleman, N., 2013. Lack of correlation between predicted and actual off-target effects of short-interfering RNAs targeting the human papillomavirus type 16 E7 oncogene. Br. J. Cancer 108 (2), 450–460. https://doi.org/10.1038/bjc.2012.564.

Heinemann, J.A., 2019. Should dsRNA treatments applied in outdoor environments be regulated? Environ. Int. 132, 104856 https://doi.org/10.1016/J.ENVINT.2019.05.050.

Heinemann, J.A. (2023). Submission on information and supporting documentation relevant to the trends in new technological developments in synthetic biology and any additional trend that should be considered. Open-ended Online Forum on Synthetic Biology. 20/03/2023-31/03/2023.

Heinemann, J.A., Billington, C., 2004. How do genomes emerge from genes? ASM N. 70, 464–471.

Heinemann, J.A., Walker, S., 2019. Environmentally applied nucleic acids and proteins for purposes of engineering changes to genes and other genetic material. Biosaf. Health 1 (3), 113–123. https://doi.org/10.1016/j.bsheal.2019.09.003.

Heinemann, J.A., Clark, K., Hiscox, T.C., McCabe, A.W., Agapito-Tenfen, S.Z., 2023. Are null segregants new combinations of heritable material and should they be regulated? Front. Genome Ed. 4 https://doi.org/10.3389/fgeed.2022.1064103.

Höijer, I., Emmanouilidou, A., Östlund, R., van Schendel, R., Bozorgpana, S., Tijsterman, M., Feuk, L., Gyllensten, U., den Hoed, M., Ameur, A., 2022. CRISPR-Cas9 induces large structural variants at on-target and off-target sites in vivo that segregate across generations. Nat. Commun. 13 (1), 627. https://doi.org/10.1038/s41467-022-28244-5.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science 337 (6096), 816–821. https://doi.org/10.1126/science.1225829.

Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., Tanabe, M., 2021. KEGG: integrating viruses and cellular organisms. Nucleic Acids Res. 49 (D1), D545–D551. https://doi.org/10.1093/NAR/GKAA970.

Kenward, S. (2023). Interim Regulatory Impact Statement: Improving Our Gmo Regulations for Laboratory and Biomedical Research.

Knorr, E., Fishilevich, E., Tenbusch, L., Frey, M.L.F., Rangasamy, M., Billion, A., Worden, S.E., Gandra, P., Arora, K., Lo, W., Schulenberg, G., Valverde-Garcia, P., Vilcinskas, A., Narva, K.E., 2018. Gene silencing in Tribolium castaneum as a tool for the targeted identification of candidate RNAi targets in crop pests, 2018 8:1 Sci. Rep. 8 (1), 1–15. https://doi.org/10.1038/s41598-018-20416-y.

Kumar, H., Panigrahi, M., Chhotaray, S., Bhanuprakash, V., Shandilya, R., Sonwane, A., Bhushan, B., 2018. Red flour beetle (Tribolium castaneum): from population genetics to functional genomics. Vet. World 11 (8), 1043–1046. https://doi.org/10.14202/VETWORLD.2018.1043-1046.

Li, J., Zhao, H., Zheng, L., An, W., 2021. Advances in synthetic biology and biosafety governance. Front. Bioeng. Biotechnol. 9 https://doi.org/10.3389/fbioe.2021.598087.

Meinke, L.J., Sappington, T.W., Onstad, D.W., Guillemaud, T., Miller, N.J., Komáromi, J., Levay, N., Furlan, L., Kiss, J., Toth, F., 2009. Western corn rootworm (Diabrotica virgifera virgifera LeConte) population dynamics. Agric. For. Entomol. 11 (1), 29–46. https://doi.org/10.1111/J.1461-9563.2008.00419.X.

Meinke, L.J., Souza, D., Siegfried, B.D., 2021. The use of insecticides to manage the Western Corn Rootworm, Diabrotica virgifera virgifera, LeConte: history, field-evolved resistance, and associated mechanisms. Insects 12 (2), 1–22. https://doi.org/10.3390/INSECTS12020112.

Miller, J.H., Suthar, A., Tai, J., Yeung, A., Truong, C., Stewart, J.L., 1999. Direct selection for mutators in Escherichia coli. J. Bacteriol. 181 (5), 1576–1584. https://doi.org/10.1128/JB.181.5.1576-1584.1999.

Miller, N., Estoup, A., Toepfer, S., Bourguet, D., Lapchin, L., Derridj, S., Kim, K.S., Reynaud, P., Furlan, L., Guillemaud, T., 2005. Ecology: multiple transatlantic introductions of the western corn rootworm. Science 310 (5750), 992. https://doi.org/10.1126/SCIENCE.1115871/SUPPL_FILE/MILLER.SOM.PDF.

Mueller, S., 2021. Facing the 2020 pandemic: what does cyberbiosecurity want us to know to safeguard the future? Biosaf. Health 3 (1), 11–21. https://doi.org/10.1016/j.bsheal.2020.09.007.

Park, J., Bae, S., Kim, J.S., 2015. Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites. Bioinforma. (Oxf., Engl.) 31 (24), 4014–4016. https://doi.org/10.1093/BIOINFORMATICS/BTV537.

Rangwala, S.H., Kuznetsov, A., Ananiev, V., Asztalos, A., Borodin, E., Evgeniev, V., Joukov, V., Lotov, V., Pannu, R., Rudnev, D., Shkeda, A., Weitz, E.M., Schneider, V. A., 2021. Accessing NCBI data using the NCBI sequence viewer and genome data viewer (GDV). Genome Res. 31 (1), 159–169. https://doi.org/10.1101/GR.266932.120.

Sappington, T.W., Siegfried, B.D., Guillemaud, T., 2006. Coordinated diabrotica genetics research: accelerating progress on an urgent insect pest problem. Am. Èntomol. 52 (2), 90–97. https://doi.org/10.1093/ae/52.2.90.

Schon, N., Gray, R., Mackay, A., 2016. Earthworms stimulate pasture production in sheep and beef systems: their economic value. J. N. Z. Grassl. 78, 89–92. https://doi.org/10.33584/jnzg.2016.78.523.

Secretariat of the convention on biological diversity (2019). Report of the Ad Hoc Technical Expert Group on Synthetic Biology. Montrea. Convention on Biological Diversity. CBD/SYNBIO/AHTEG/2019/1/3.

Segal, D.J., Meckler, J.F., 2013. Genome engineering at the dawn of the golden age. Annu. Rev. Genom. Hum. Genet. 14 (1), 135–158. https://doi.org/10.1146/annurev-genom-091212-153435.

Sharma, V.K., Marla, S., Zheng, W., Mishra, D., Huang, J., Zhang, W., Morris, G.P., Cook, D.E., 2022. CRISPR guides induce gene silencing in plants in the absence of Cas. Genome Biol. 23 (1), 6. https://doi.org/10.1186/s13059-021-02586-7.

Shi, J., Gao, H., Wang, H., Lafitte, H.R., Archibald, R.L., Yang, M., Hakimi, S.M., Mo, H., Habben, J.E., 2017. ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. Plant Biotechnol. J. 15 (2), 207–216. https://doi.org/10.1111/pbi.12603.

Smith, R.H., Chen, Y.C., Seifuddin, F., Hupalo, D., Alba, C., Reger, R., Tian, X., Araki, D., Dalgard, C.L., Childs, R.W., Pirooznia, M., Larochelle, A., 2020. Genome-wide analysis of off-target CRISPR/Cas9 activity in single-cell-derived human hematopoietic stem and progenitor cell clones. Genes 11 (12), 1–16. https://doi.org/10.3390/genes11121501.

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N.T., Pyysalo, S., Bork, P., Jensen, L.J., von Mering, C., 2023. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 51 (D1), D638–D646. https://doi.org/10.1093/NAR/GKAC1000.

Tao, J., Bauer, D.E., Chiarle, R., 2023. Assessing and advancing the safety of CRISPR-Cas tools: from DNA to RNA editing. Nat. Commun. 14 (1), 212. https://doi.org/10.1038/s41467-023-35886-6.

Wang, S., Zhang, S., Wang, W., Xiong, X., Meng, F., Cui, X., 2015. Efficient targeted mutagenesis in potato by the CRISPR/Cas9 system. Plant Cell Rep. 34 (9), 1473–1476. https://doi.org/10.1007/s00299-015-1816-7.

Wytinck, N., Ziegler, D.J., Walker, P.L., Sullivan, D.S., Biggar, K.T., Khan, D., Sakariyahu, S.K., Wilkins, O., Whyard, S., Belmonte, M.F., 2022. Host induced gene silencing of the Sclerotinia sclerotiorum ABHYDROLASE-3 gene reduces disease severity in Brassica napus. PLoS ONE 17 (8 August). https://doi.org/10.1371/journal.pone.0261102.

Zhang, F., Wen, Y., Guo, X., 2014. CRISPR/Cas9 for genome editing: progress, implications and challenges. Hum. Mol. Genet. 23 (R1), R40–R46. https://doi.org/10.1093/hmg/ddu125.

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 Mol. Biol. 96 (4–5), 445–456. https://doi.org/10.1007/s11103-018-0709-x.