Proceedings of the 5th International Conference on Modern Medicine and Global Health
DOI: 10.54254/2753-8818/2027.37449

An Automated Multi-criteria Computational Workflow for dsRNA Design and Prioritization in Aspergillus flavus

Muhan Tang

The Experimental High School Attached To Beijing Normal University, Beijing, China
muhan.tang@hotmail.com

Abstract. Efficient double-stranded RNA (dsRNA) design is a prerequisite for RNA interference (RNAi)-based control of fungal pathogens, yet a usable design has to satisfy several sequence requirements at once, and comparing candidates by hand is slow, inconsistent between researchers, and hard to audit. We developed an automated computational workflow that spans candidate generation, multi-criteria sequence evaluation, weighted composite scoring, and experimental preparation. Taking Aspergillus flavus, an aflatoxin-producing phytopathogen, as the case study, the workflow produced more than 100 000 dsRNA candidates and ranked them on seven biologically relevant metrics: GC content, sequence complexity, repeat sequence abundance, inverted repeat content, predicted RNA secondary structure stability, estimated effective siRNA abundance following Dicer processing, and potential off-target effects. The highest-scoring candidate reached a composite score of 91.8, and the workflow returned the matching PCR primers and T7 transcription templates alongside it. The framework offers a systematic and reproducible way to prioritize dsRNA candidates for RNAi-based pathogen control before any experimental validation is attempted.

Keywords: RNA interference (RNAi), double-stranded RNA (dsRNA), *Aspergillus flavus*, computational workflow, multi-criteria evaluation

1. Introduction

Fungal plant pathogens threaten global food security on several fronts at once: they reduce crop yields, contaminate harvested grain with mycotoxins, and leave growers dependent on synthetic fungicides. *Aspergillus flavus* is a case in point. This aflatoxin-producing fungus infects maize, peanuts and other nut crops, and the toxins it leaves behind endanger both people and livestock [1]. Fungicides remain the default measure for controlling such pathogens, but resistant pathogen variants have already emerged, and their residues create pollution and food-safety problems of their own. The demand for host-specific, environmentally benign alternatives is therefore pressing [2]. RNA interference (RNAi) meets that description. Its two main agricultural forms, host-induced gene silencing (HIGS) and spray-induced gene silencing (SIGS), deliver dsRNA that silences essential fungal genes, and neither relies on the chemistry that conventional fungicides use [3]. The finding that small RNAs move in both directions between plant hosts and fungal pathogens gave these approaches a mechanistic footing, and with it a reason to expect them to work in the field [4]. What

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2027.37449

has moved to centre stage more recently is not whether RNAi can be deployed, but how the dsRNA itself should be chosen.

How efficiently RNAi silences a target, and how cleanly it does so, depends largely on the sequence of the dsRNA that is introduced. No single property decides the outcome. GC composition, sequence complexity, repetitive element content, secondary conformation, the yield of siRNAs after Dicer processing, and the risk of hybridizing to an unintended transcript all contribute [1, 2]. A workable design therefore has to balance several biological constraints rather than optimize one number. Doing this by hand does not scale. A single coding sequence yields far more candidate fragments than a person can compare reliably, and which fragments get discarded depends as much on the person doing the discarding as on the sequences themselves.

Several tools already cover parts of this problem. E-RNAi designs optimized RNAi constructs and predicts off-targets, while dsCheck performs sensitive off-target searches for dsRNA-mediated interference [5-7]. Each is useful on its own; none of them was built to talk to the others. In practice a user runs one program, saves the output, runs the next, and reconciles the results by hand before a candidate can be judged. Reviewers in this field have been asking for the opposite arrangement: a standardized, automated and reproducible workflow that carries a design from sequence evaluation through candidate prioritization to experimental preparation without leaving the pipeline [8].

We report here a pipeline that answers that request. Given one or more target coding sequences, it scans them end to end, scores every candidate fragment against seven biologically motivated criteria, returns the fragments in ranked order, and writes out the PCR primers and in vitro transcription templates needed to test them. Auxiliary AI modules handle peripheral steps of the workflow; the scoring itself remains deterministic, so any ranking can be traced back to the numbers that produced it.

2. Method

2.1. Data collection

We downloaded the *Aspergillus flavus* reference assembly and its matching gene annotation from a public genome repository, extracted the coding-region nucleotide sequences, and stored them as FASTA files. Those files are the only input the pipeline requires. Most published RNAi design workflows ask the user to narrow the target down to a region of interest before any calculation runs, and that step has to be repeated for every new gene. We removed it altogether. Full-length coding sequences go in as they are, and every nucleotide site is assessed on the same footing regardless of where it sits, which removes the subjective bias that a hand-picked window would introduce and makes the output reproducible from the input alone. The same property holds when the target changes: a new gene, or even a different fungal species, requires no modification to the core computation modules.

2.2. Automated candidate dsRNA generation

The conventional route through dsRNA design is manual in an important sense: a researcher selects a small set of candidate segments first and then checks them one by one with whatever bioinformatic tools are available. We replaced that with a sliding window, following the automated RNAi design frameworks of Arziman et al. and Horn and Boutros [5, 6]. The window advances along the coding sequence and emits a candidate fragment at each position it occupies, including the ones a researcher would probably not have thought to pick. Nothing is skipped, so regions with

DOI: 10.54254/2753-8818/2027.37449

strong silencing potential are not lost merely because they were never considered. Every candidate then passes through the same set of calculations. That uniformity is what makes the resulting scores comparable with one another, and what makes screening a sequence at this scale computationally feasible in the first place.

2.3. Multi-criteria sequence evaluation

Silencing is not governed by a single molecular property, so no one number can rank a set of dsRNA candidates [1, 2]. We score each fragment on seven criteria instead: GC content, sequence complexity, repeat sequence abundance, inverted repeat content, predicted RNA secondary structure stability, estimated effective siRNA abundance following Dicer processing, and potential off-target effects. Table 1 gives the biological reasoning behind each and the references it rests on. Together the seven cover what a working design needs: duplex thermodynamic stability, sequence uniqueness, accessibility of the target binding site, inherent silencing capacity, and the risk of interfering with non-target transcripts [6, 7]. Assessing all of them together gives a more dependable quality estimate than any single parameter could [2, 3]. Because the raw metrics are measured on very different scales, each was rescaled to the interval 0 to 1 by linear normalization before entering the score. This also puts the seven criteria on a common footing, so that the weights below express relative importance rather than relative scale.

Table 1. Biological evaluation criteria incorporated into the computational workflow

Evaluation criterion	Biological rationale	Supporting references
GC content	Maintains appropriate duplex stability	[1, 2]
Sequence complexity	Avoids low-complexity regions	[6]
Repeat sequence abundance	Improves sequence specificity	[6]
Inverted repeat content	Reduces self-complementary structures	[6]
RNA secondary structure	Improves RNA accessibility	[1, 2]
Effective siRNA abundance	Estimates RNAi silencing potential	[3]
Off-target effects	Reduces unintended gene silencing	[6, 7]

Note: The seven criteria were selected to capture complementary aspects of dsRNA sequence quality, predicted silencing potential, and specificity. Supporting references indicate the biological rationale used for inclusion of each criterion in the computational workflow.

2.4. Weighted multi-criteria scoring and candidate prioritization

Ranking a set of dsRNA candidates requires a single number per candidate. We therefore defined a weighted multi-criteria scoring function that combines the seven normalized metrics linearly, with each metric weighted according to its relative importance (Equation 1).

$$Score = (GC \times 0.12 + Complexity \times 0.13 + Repeat \times 0.13 + Inverted \times 0.12 \\ + Structure \times 0.15 + siRNA \times 0.20 + OffTarget \times 0.15) \times 100 \quad (1)$$

Every variable in Equation (1) is a normalized value, so each lies between 0 and 1. Effective siRNA abundance carries the largest weight at 0.20, because the number of siRNAs a fragment is expected to release after Dicer processing is the closest available proxy for how strongly it should silence its target. RNA secondary structure and off-target risk follow at 0.15 each: the first governs

DOI: 10.54254/2753-8818/2027.37449

whether the target site can be reached, the second how specific the fragment is. Sequence complexity and repeat abundance receive 0.13 each, and GC content and inverted-repeat content 0.12 each. Fragments are then sorted by composite score, highest first. The scheme itself is a standard weighted-sum approach from multi-attribute decision making, widely used for ranking nucleic acid sequences; we adopt it rather than propose it, which keeps the ranking transparent and easy to recompute from the published weights.

2.5. Automated experimental design and report generation

A high score is not the end of the process. For each fragment that ranks well, the pipeline designs a PCR primer pair and assembles the DNA template with its T7 promoter already attached, so that the fragment can go directly into in vitro transcription without further manual design. Quantitative metric values, composite scores, ranks and primer specifications are exported together in three formats: HTML, CSV and JSON. The HTML file is meant to be read, the CSV to be sorted and filtered, and the JSON to be passed to another script. Because generation, scoring, ranking and export all run inside one pipeline, the manual steps that would otherwise sit between them disappear, and with them the transcription errors those steps tend to introduce. The framework is also flexible enough to accept other target genes and other fungal pathogens, which we discuss below.

3. Results

When run against our Aspergillus flavus target gene, the sliding-window tiling produced over 100 000 distinct dsRNA fragment candidates. Every fragment went through calculation of the seven sequence-based descriptors. After normalizing each metric to 0-1 scale, our weighted formula computed a composite quality score for every candidate fragment. Fragments were ordered by composite score for downstream inspection.

Table 2 lists the top-ten fragments identified by our pipeline. Fragment XM_041285628.1:250-475 obtains the highest composite score at 91.8. Two other fragments, XM_041285628.1:200-425 and XM_041285628.1:350-575, reach scores of 89.2 and 87.1 correspondingly. The score interval for top-10 candidates spans 81.4 up to 91.8. When we look at component metrics for these top-ranked entries, strong performance typically comes from a combination of high effective-siRNA estimates, good secondary-structure properties and satisfactory sequence-complexity values. No single parameter accounts for top-ranking status; competitive fragments perform well across multiple evaluation dimensions.

Table 2. Top ten candidate dsRNAs identified by the computational workflow

Rank	Candidate	GC Content	Sequence Complexity	Repeat Score	Inverted Repeat	RNA Structure	Effective siRNA	Off-target	Composite Score
1	[CandidateXM_041285628.1:250-475]	[57.3]	[0.84]	[1.00]	[1.00]	[0.59]	[13/69]	[1.00]	[91.8]
2	[CandidateXM_041285628.1:200-425]	[56.9]	[0.84]	[1.00]	[1.00]	[0.58]	[9/69]	[1.00]	[89.2]
3	[CandidateXM_041285628.1:350-575]	[56.4]	[0.87]	[1.00]	[1.00]	[0.41]	[9/69]	[1.00]	[87.1]
4	[CandidateXM_041285628.1:400-625]	[56.9]	[0.88]	[1.00]	[1.00]	[0.36]	[9/69]	[1.00]	[86.3]
5	[CandidateXM_041285628.1:750-975]	[59.1]	[0.97]	[1.00]	[0.99]	[0.37]	[8/69]	[1.00]	[85.7]

DOI: 10.54254/2753-8818/2027.37449

Table 2. (continued)

6	[CandidateXM_041285628.1:100-325]	[58.2]	[0.98]	[1.00]	[1.00]	[0.32]	[8/69]	[1.00]	[85.2]
7	[CandidateXM_041285628.1:550-775]	[57.8]	[0.87]	[1.00]	[1.00]	[0.49]	[6/69]	[1.00]	[82.4]
8	[CandidateXM_041285628.1:600-825]	[58.2]	[0.87]	[1.00]	[1.00]	[0.48]	[6/69]	[1.00]	[82.3]
9	[CandidateXM_041285628.1:150-375]	[56.4]	[0.96]	[1.00]	[1.00]	[0.50]	[5/69]	[1.00]	[81.7]
10	[CandidateXM_041285628.1:500-725]	[54.7]	[0.88]	[1.00]	[1.00]	[0.41]	[6/69]	[1.00]	[81.4]

Note: Individual evaluation metrics represent the normalized values used in the weighted scoring model. Composite scores integrate all seven criteria and were used to determine the final candidate ranking.

For each high-score fragment, our pipeline computes corresponding PCR primers and T7-promoter-attached templates for in-vitro transcription. All fragment sequences, individual metric values, composite scores, ranks and primer specifications are saved into HTML, CSV and JSON output files. These exported files preserve full computational records to support follow-up wet-lab validation work.

4. Discussion

RNAi offers a sequence-specific way to suppress fungal phytopathogens, and unlike conventional fungicides it leaves no chemical residue behind. Turning that potential into a usable dsRNA reagent is harder than the underlying biology suggests. A functional dsRNA molecule has to satisfy several sequence-level properties that pull against one another, and under current practice those properties are examined in different places. Researchers typically run several unconnected bioinformatic resources and then consolidate the output by hand before settling on a fragment [6, 8]. The cost of that arrangement shows up twice. It takes time, and it introduces a point at which two people working from the same data can reach different answers, or at which a result cannot be reproduced at all. Neither problem improves as the candidate library grows; both get worse. Running the entire chain inside one workflow removes both. Fragment generation, evaluation against the seven criteria, candidate prioritization and experiment-ready file output are handled by a single automated process, and the ranking follows from published weights rather than from the operator's judgement. Primer sequences, T7-containing templates and the accompanying documentation are produced without anyone retyping them, which is where most errors of that kind would otherwise originate. For a dsRNA development project, the net effect is a shorter path from sequence to bench and a result that another laboratory can reproduce.

The workflow is not tied to *Aspergillus flavus*. We selected that species because of its agricultural importance and because its reference genome is publicly available and well annotated, but nothing in the pipeline assumes it. No species-specific parameters are preset anywhere in the core modules, so analyzing a different fungal pathogen means supplying a different reference genome and a different target-gene FASTA file and nothing else. The same analytical logic should extend to other RNA-based applications, such as RNAi-based pest control and sequence screening in functional genomics, though we have not yet tested it in those settings [2, 3].

Three limitations deserve to be stated plainly. First, every ranking reported here comes from computation alone. No laboratory validation was carried out within the scope of this study, so a high

DOI: 10.54254/2753-8818/2027.37449

composite score indicates that a fragment ranks well under our model, not that it silences the gene in vivo. Second, the weights in Equation (1) are set from published biological reasoning rather than fitted to data. They should be treated as a defensible starting point; a larger body of measured silencing outcomes would allow them to be calibrated, whether by regression or by machine learning, and we would expect the ranking to shift once that is done. Third, the feature set is incomplete. In vivo target-site accessibility, sequence evolutionary conservation, and empirically measured silencing efficacy are all absent, and any one of them could change which fragments rise to the top. Adding these descriptors is the most obvious direction for a later revision.

One point deserves emphasis. This pipeline does not replace bench work, and it is not intended to. Its purpose is to shrink the candidate pool before experiments are run, so that a limited number of assays are spent on fragments that are likely to work. As RNA-driven techniques move from proof-of-concept studies into routine agricultural use and functional genomics, that kind of pre-screening becomes more valuable: it lowers the cost of entry for groups without a large experimental budget, and it makes designs easier to compare between laboratories.

5. Conclusion

We have described a computational pipeline that takes coding sequences as input and returns ranked dsRNA candidates. Between those two ends it tiles each sequence with a sliding window, scores every fragment against seven criteria, combines those scores using published weights, and exports the primers, templates and records needed at the bench. No part of that chain depends on who runs it. Applied to *Aspergillus flavus*, the pipeline reduced a very large candidate set to a shortlist a laboratory could realistically work through, without the outcome resting on the operator's prior experience with RNAi. The scores remain predictions, and predictions require testing. What the pipeline provides is a dependable basis for deciding what to test first, along with a design that carries over to other fungal pathogens and to RNAi applications elsewhere.

References

[1] Rosa, C., Kuo, Y.-W., Wuriyanghan, H., & Falk, B. W. (2018). RNA interference mechanisms and applications in plant pathology. *Annual Review of Phytopathology, 56*, 581–610. https://doi.org/10.1146/annurev-phyto-080417-050044

[2] Mann, C. W. G., Sawyer, A., Gardiner, D. M., Mitter, N., Carroll, B. J., & Eamens, A. L. (2023). RNA-based control of fungal pathogens in plants. *International Journal of Molecular Sciences, 24*(15), 12391. https://doi.org/10.3390/ijms241512391

[3] Koch, A., & Wassenegger, M. (2021). Host-induced gene silencing—mechanisms and applications. *New Phytologist, 231*(1), 54–59. https://doi.org/10.1111/nph.17364

[4] Qiao, Y., Xia, R., Zhai, J., Hou, Y., Feng, L., et al. (2021). Small RNAs in plant immunity and virulence of filamentous pathogens. *Annual Review of Phytopathology, 59*, 265–288. https://doi.org/10.1146/annurev-phyto-121520-023514

[5] Arziman, Z., Horn, T., & Boutros, M. (2005). E-RNAi: A web application to design optimized RNAi constructs. *Nucleic Acids Research, 33*(Suppl. 2), W582–W588. https://doi.org/10.1093/nar/gki468

[6] Horn, T., & Boutros, M. (2010). E-RNAi: A web application for the multi-species design of RNAi reagents—2010 update. *Nucleic Acids Research, 38*(Suppl. 2), W332–W339. https://doi.org/10.1093/nar/gkq317

[7] Naito, Y., Yamada, T., Matsumiya, T., Ui-Tei, K., Saigo, K., & Morishita, S. (2005). dsCheck: Highly sensitive off-target search software for double-stranded RNA-mediated RNA interference. *Nucleic Acids Research, 33*(Suppl. 2), W589–W591. https://doi.org/10.1093/nar/gki419

[8] Cedden, D. (2026). Emerging experimental and bioinformatic approaches in RNA interference-based pest control research. *Insect Molecular Biology, 35*(4), 364–375. https://doi.org/10.1111/imb.70040