8-2023

Leveraging eco-evolutionary models for gene drive risk assessment

Matthew A. Combs

Andrew J. Golnar

Justin M. Overcash

Alun L. Lloyd

Keith R. Hayes

See next page for additional authors

Authors

Matthew A. Combs, Andrew J. Golnar, Justin M. Overcash, Alun L. Lloyd, Keith R. Hayes, David A. O'Brochta, and Kim M. Pepin

Trends in Genetics

CellPress

Review

Leveraging eco-evolutionary models for gene drive risk assessment

Matthew A. Combs,[1,*] Andrew J. Golnar,[1] Justin M. Overcash,[2] Alun L. Lloyd,[3] Keith R. Hayes,[4] David A. O'Brochta,[5] and Kim M. Pepin[1]

Engineered gene drives create potential for both widespread benefits and irreversible harms to ecosystems. CRISPR-based systems of allelic conversion have rapidly accelerated gene drive research across diverse taxa, putting field trials and their necessary risk assessments on the horizon. Dynamic process-based models provide flexible quantitative platforms to predict gene drive outcomes in the context of system-specific ecological and evolutionary features. Here, we synthesize gene drive dynamic modeling studies to highlight research trends, knowledge gaps, and emergent principles, organized around their genetic, demographic, spatial, environmental, and implementation features. We identify the phenomena that most significantly influence model predictions, discuss limitations of biological complexity and uncertainty, and provide insights to promote responsible development and model-assisted risk assessment of gene drives.

Highlights

As development of gene drive systems accelerates and diversifies, predicting outcomes for target populations and the potential for human and environmental risks requires accounting for numerous eco-evolutionary processes.

Gene drive dynamic models quantify the influence of features across genetics (e.g., resistance development and standing genetic diversity), demographics (e.g., mating systems and inbreeding), spatial ecology (e.g., dispersal and competition), biotic and abiotic environments (e.g., climate variation and landscape structure), and implementation strategies (e.g., introduction size and timing) on gene drive outcomes.

Synthesizing published gene drive models reveals research trends, knowledge gaps, and emergent principles. Modeling limitations and tradeoffs are discussed.

Integrating an iterative modeling approach within the existing phased pathway for gene drive research improves utility for risk assessment.

Dynamic models enable gene drive research and risk analysis

Gene drives are naturally occurring or engineered genetic systems that cause biased inheritance patterns of specific alleles [1]. Most engineered gene drives seek to reduce a **target population's** (see Glossary) abundance (i.e., suppression drives) or modify phenotypes of individuals in a target population (i.e., replacement drives). Technological development has rapidly increased the available strategies for gene drive design and implementation, revolutionizing our capacity to address broad-scale ecological issues [2,3]. Gene drives may improve agricultural security and biodiversity conservation by eradicating pest species [4,5], or advance public health goals by altering insect vectorial capacity [6]. A phased testing pathway guides the development and testing of these genetically modified organisms (Box 1) [7,8], but significant uncertainty remains about the **risks** gene drives may pose to human and natural communities [9].

Risk assessments (RAs; e.g., environmental, socio-economic, epidemiological) guide regulatory decisions governing the testing and implementation of biotechnologies intended for environmental release. They generally seek to identify causal pathways by which sources of harm (i.e., **hazards**) create adverse outcomes, determine the likelihood of experiencing harm in the context of exposure (i.e., **risk**), evaluate strategies to mitigate risk, and provide recommendations for decision-makers. Risk-based decision frameworks are applied across diverse domains including engineering, finance, medicine, and ecotoxicology [10]. Engineered gene drive systems require case-by-case RAs because each construct could act differently across the biological scales of individuals, populations, and communities. The theoretical potential to spread and persist in environments presents several challenges to RAs that differ from other environmental contexts where they are applied. (i) Gene drive traits undergo continuous evolutionary pressures before fixation or loss, as they disperse through the population's natural reproductive behaviors and persist in gene pools for many generations (particularly for low-threshold drives intended

[1]National Wildlife Research Center, United States Department of Agriculture, Animal and Plant Health Inspection Service, Wildlife Services, Fort Collins, CO, 80521, USA
[2]United States Department of Agriculture, Animal and Plant Health Inspection Service, Biotechnology Regulatory Services, 20737, USA
[3]North Carolina State University, Biomathematics Graduate Program and Department of Mathematics, Raleigh, NC, 27695, USA
[4]The Commonwealth Scientific and Industrial Research Organisation, Data 61, Hobart, TAS, 7004, Australia

[5]Foundation for the National Institutes of Health, North Bethesda, MD, 20852, USA

*Correspondence:
matthew.combs@usda.gov
(M.A. Combs).

Box 1. Phased testing pathway for gene drive development

A phased testing pathway has been recommended to guide the development and testing of gene drive products by national and international advisory groups (e.g., World Health Organization, National Academy of Science, Engineering, and Medicine) [6,7]. This pathway promotes the simultaneous evaluation of risks posed by gene drives (e.g. environmental, socioeconomic, and epidemiological) while allowing developers to study the critical factors influencing gene drive outcomes under progressively decreasing confinement and containment. Phase 0 encompasses research preparation, in which a gene drive construct is designed for a specific target population and strategies for risk assessment, management, and mitigation are established. In Phase 1, indoor laboratory-based research begins, as researchers optimize the genetic conversion efficacy of the gene drive system and evaluate fitness costs for transgenic individuals in controlled environments. Phase 1 trials allow for the collection of empirical data over multiple generations while under strict containment, which can be used to inform GDDMs and risk assessments. Progression to Phase 2 involves contained outdoor field trials, using either constructed barriers (e.g., netting or fencing) or geographic isolation (e.g., islands), allowing the study of environmental complexity such as climatic variability, interspecific interactions, and spatial processes. Phase 2 trials allow for more realistic evaluation of transgenic fitness, population outcomes, and ecosystem effects while informing risk assessments and regulatory decisions. Phase 3 involves releasing GDMOs into natural environments without any physical containment. Both Phase 2 and 3 require site-specific regulatory approval from nations in which GDMOs are released. Finally, Phase 4 involves post-release monitoring and surveillance to evaluate true gene drive efficacy and adverse outcomes. For some gene drives, Phase 4 may also involve the release of additional drive constructs to modulate or reverse initial systems. To date, no gene drive projects have been approved for Phase 2 field trials, though several have begun Phase 1 experiments. It is important to recognize that the phased testing pathway is a non-binding recommendation, and regulatory decisions about GDMO introductions will ultimately be nation-specific unless international agreements are ratified into law. Thus, while the phased testing pathway recommendations are designed to provide decision checkpoints between phases, in practice, development phases may overlap.

for population replacement). (ii) Their potential spatial scope (sometimes continental) invokes greater variation in risk-relevant parameters than experienced previously with genetically modified organisms.

Gene drive developers and risk assessors must therefore consider how eco-evolutionary features of gene drive constructs and target populations interact across biological scales, from molecular processes regulating **allelic conversion**, to demographic and ecological processes influencing fitness of transgenic individuals, to environmental contexts and implementation strategies impacting the spatial spread and persistence of gene drive constructs within and among populations. Each application produces unique risks, desired outcomes, and stakeholder priorities, requiring data collection across different contexts. These considerations are difficult to represent, logically organize, and investigate with qualitative or non-mechanistic statistical methods, whereas **dynamic process models** are well suited to this task.

Dynamic process models are mathematical representations of systems over time. They are used to understand many eco-evolutionary processes from gene regulation [11,12], to zoonotic disease emergence, [13,14] to species invasion [15,16]. Dynamic models can also predict the spread of gene-drive-modified organisms (GDMOs) by replicating *in silico* processes that induce changes in population allele frequencies or absolute counts across a variety of biological, temporal, and spatial scales [17]. For example, gene drive dynamic models (GDDMs) can describe how molecular dynamics influence gene drive evolutionary stability [18], how mating systems affect minimum release thresholds [19], or how landscape heterogeneity influences speed of drive spread [20]. For risk assessors and regulators focused on identifying, minimizing, and managing adverse outcomes and meeting desired outcome criteria, GDDMs can provide valuable insight by comparing genetic technologies and implementation strategies under diverse ecological, social, and fiscal contexts [21].

Modelers use various methods to construct GDDMs. Deterministic approaches describe average outcomes from biological processes and parameters, ignoring the influence of random events. However, stochastic modeling frameworks capture how random events may shift the trajectory of small populations (e.g., genetic drift and Allee effects), making them particularly useful for

Glossary

Allelic conversion: process by which genetic systems change an individual's genotypic composition, usually with the intent of altering inheritance patterns and/or the viability or fitness of individuals with specific genotypes.
Dispersal: geographic movement of an organism, usually from their natal location to a different location in search of resources or mates.
Dynamic process model: mathematical representation of interrelated processes that describe system properties overtime.
Hazard: potential source of harm expected to illicit adverse outcomes for natural ecosystems, human communities, or other systems over space and time.
Interference drive: gene drive mechanism by which a genetic system restricts or permits the development of offspring depending on presence or absence of specific alleles, also called meiotic drive.
Remediation drive: gene drive mechanism intended to reverse or alter the outcomes of a previously introduced gene drive.
Replicator drive: gene drive mechanism by which a specific allele is replaced with a desirable allele through homology directed repair, often achieved through site-directed cleavage by a CRISPR-Cas9 system.
Resistance development: process by which individuals evolve to prevent allelic conversion by a gene drive mechanism.
Risk: likelihood of adverse outcomes caused by a hazard, given degree of exposure.
Target population: group of conspecific individuals through which a gene drive construct is intended to spread. The geographic scale and genetic specificity of a target population is specific to each gene drive project.

modeling suppression gene drive systems as they approach extinction [22,23]. Often gene drives are modeled as discrete-time systems that record population characteristics across evenly spaced generations, while continuous-time systems allow more realistic simulations of overlapping generations [24]. GDDMs can explore the spatial dynamics of gene drive spread using 1D wave forms [25,26] or through metapopulation or individual-based simulations on a 2D grid [27,28] or across continuous space [29].

Underdominance drive: gene drive mechanism by which heterozygous transgenic individuals exhibit lower fitness than wildtype or homozygous transgenic individuals, allowing a gene drive construct to spread population-wide only when introduced at high proportions.

When robustly validated, GDDMs provide a unique platform for geneticists to illustrate how ecological and evolutionary processes interact to mediate the consequences of novel biotechnologies [30]. However, navigating the diverse genetic approaches, hypotheses, and biological features examined across models remains challenging [31]. Here, we synthesize research trends and knowledge gaps across published GDDMs and provide an organizing framework to improve their utility for RA applications. Review and analysis methods are provided in Supplemental Text 1. Resulting data table and citations of reviewed GDMM publications are provided in Supplemental File 1 and Supplemental Text 2, respectively.

Trends and gaps in gene drive modeling

Across published GDDMs, 34 common biological features informing model structure and parameters emerged, falling into five categories of increasing biological scale: genetics, demography, spatial ecology, biotic and abiotic environment, and implementation strategy (Figure 1). Category descriptions and feature definitions are provided in Supplemental Texts 3 and 4.

GDDM publication rate surged over the past 6 years (2011–2016: 3.7/year, 2017–2022: 13.2/year) (Figure 2A), driven largely by adoption of CRISPR/Cas9 **replicator** strategies over **interference** and **underdominance** strategies. These strategies are not mutually exclusive descriptors and gene drive constructs may use combined strategies [32] (Figure 2B). Although modelers once focused primarily on replacement drives and mosquitoes, drive goals and target taxa have diversified over time (Figure 2C,D), with increasing attention on suppression or **remediation drives** targeting mammals, fruit flies, and other species.

While roughly half of GDDM publications mention relevance of gene drive-associated risks (Figure 2A), to date, most GDDMs have focused on questions of efficacy, theory, and the effects of individual model features rather than on RA applications. Most published GDDMs are relevant to early-phase projects and reasonably ignore features adding unnecessary complexity. Such basic research efforts are necessary and valuable, but as drive projects seek field trials and regulatory approval GDDMs will require increased focus on system behavior and feature interactions specifically related to risk. Still, the dearth of GDDMs incorporating features of biotic and abiotic environments (Figure 3) is striking given the potential risks of unexpected spread through introgression or hybridization, or bottom-up ecosystem changes driven by rapid population declines [5,33]. Understanding gene drive efficacy and spatial spread in variable environments and in the context of interacting populations are important requirements for gene drive RAs [34,35]. Features in this category were recommended for future study more often than they were included, suggesting researchers recognize this research gap. Several modeling platforms facilitate simulation of multispecies eco-evolutionary interactions and seasonal variations (e.g., Skeeter Buster, SLiM v4.0, and MGDrivE2) [36–38]. Modeling the influence of biotic and abiotic environments is hampered by a lack of relevant data describing multispecies interactions and potential for drive spread via hybridization and introgression.

Though **resistance development** to drive mechanisms exhibits strong influence on population outcomes in greater proportion than any other genetic feature, fewer than half of GDDMs account

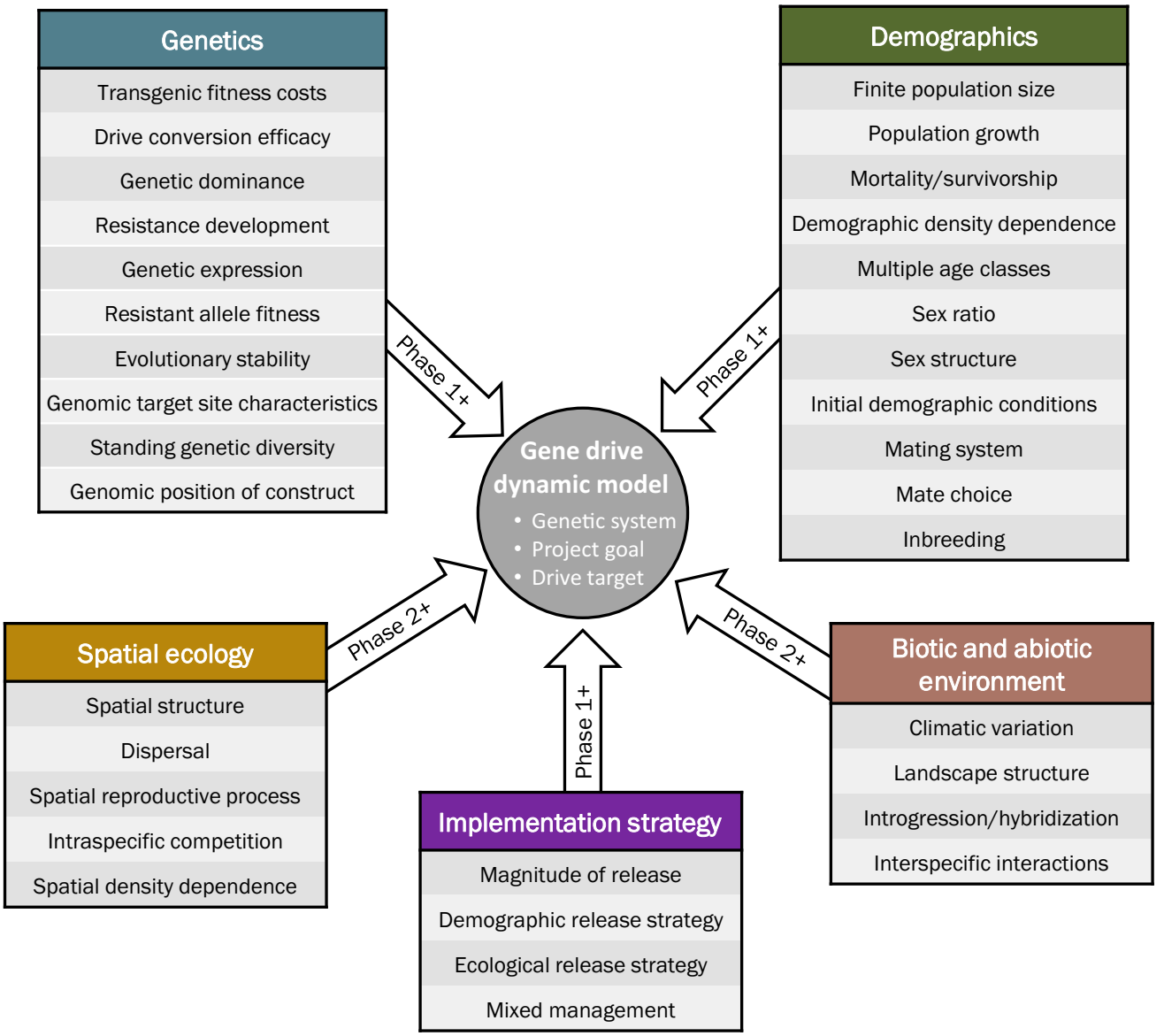

Figure 1. Common features of gene drive dynamic models (GDDMs). Each GDDM will be first developed based on unique project characteristics such as the genetic system of allelic conversion, outcome goals, and targeted population. Different categories of features are likely to be incorporated to reflect study goals along the phased research pathway, described in arrow labels.

for its role (Figure 3). Characterizing and circumventing processes that cause resistance to gene drive systems has been a major concern for developers and risk assessors because strong selective pressure for resistant alleles can cause drive failure, particularly for homing CRISPR/Cas9 systems [39]. GDDMs have been valuable for identifying how evolutionary, life history, and spatial processes interact with resistance development to shape gene drive outcomes [18,40,41]; understanding and incorporating these eco-evolutionary dynamics is important for modeling more complex processes. Many potential hazards recognized by RA involve the causes and consequences of resistance development [8], suggesting its inclusion and exploration should be prioritized across risk-focused GDDM efforts.

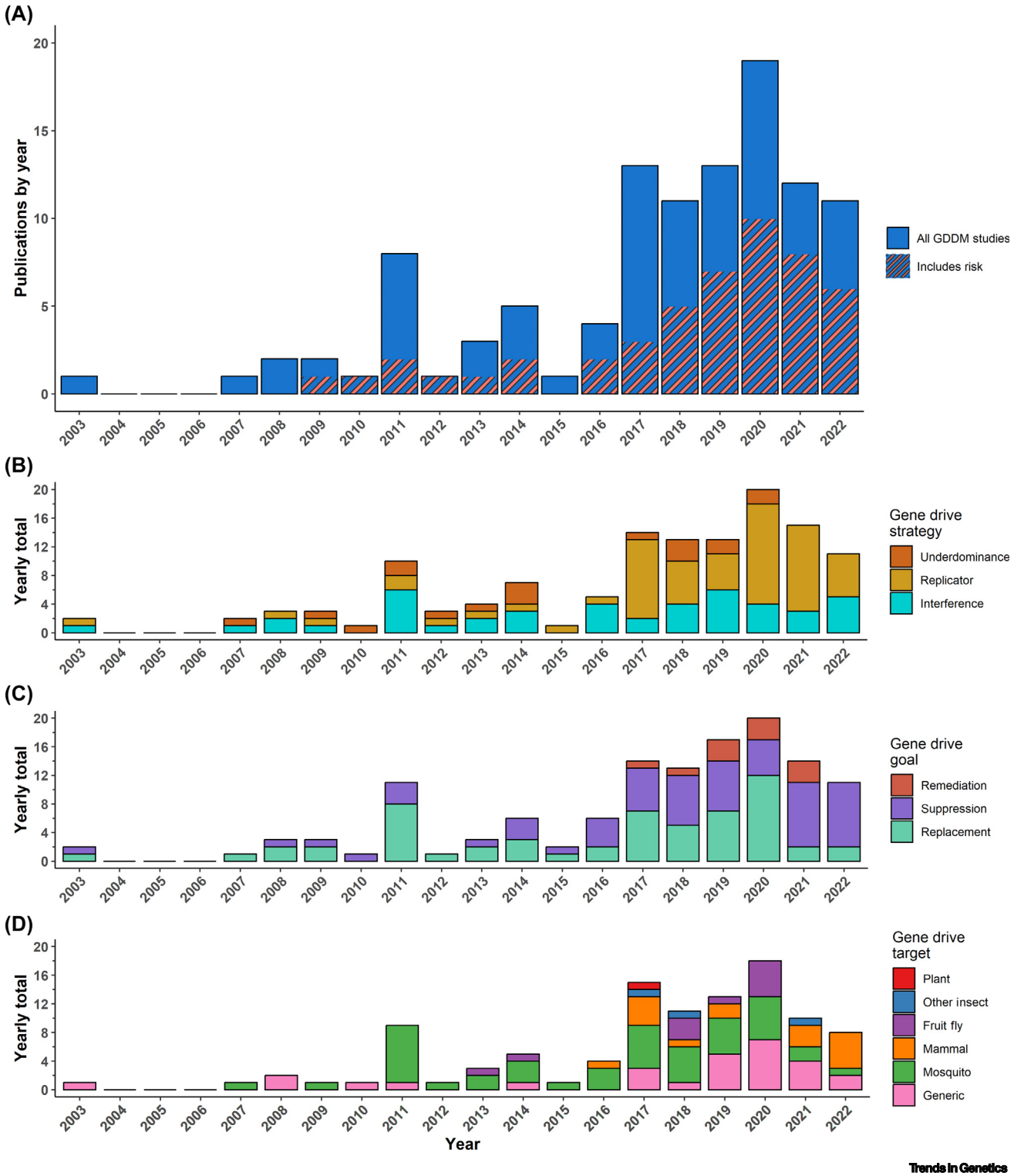

Figure 2. Temporal trends by year across gene drive dynamic models (GDDMs). (A) GDDMs published per year (total bar height) and the subset of those that discussed relevance to drive-associated risks (orange stripes). Yearly breakdown of GDDMs simulating different gene drive genetic strategies (B), drive goals (C), and drive targets (D) across publications. Note that some publications reported multiple models with different gene drive strategies, goals, and/or targets.

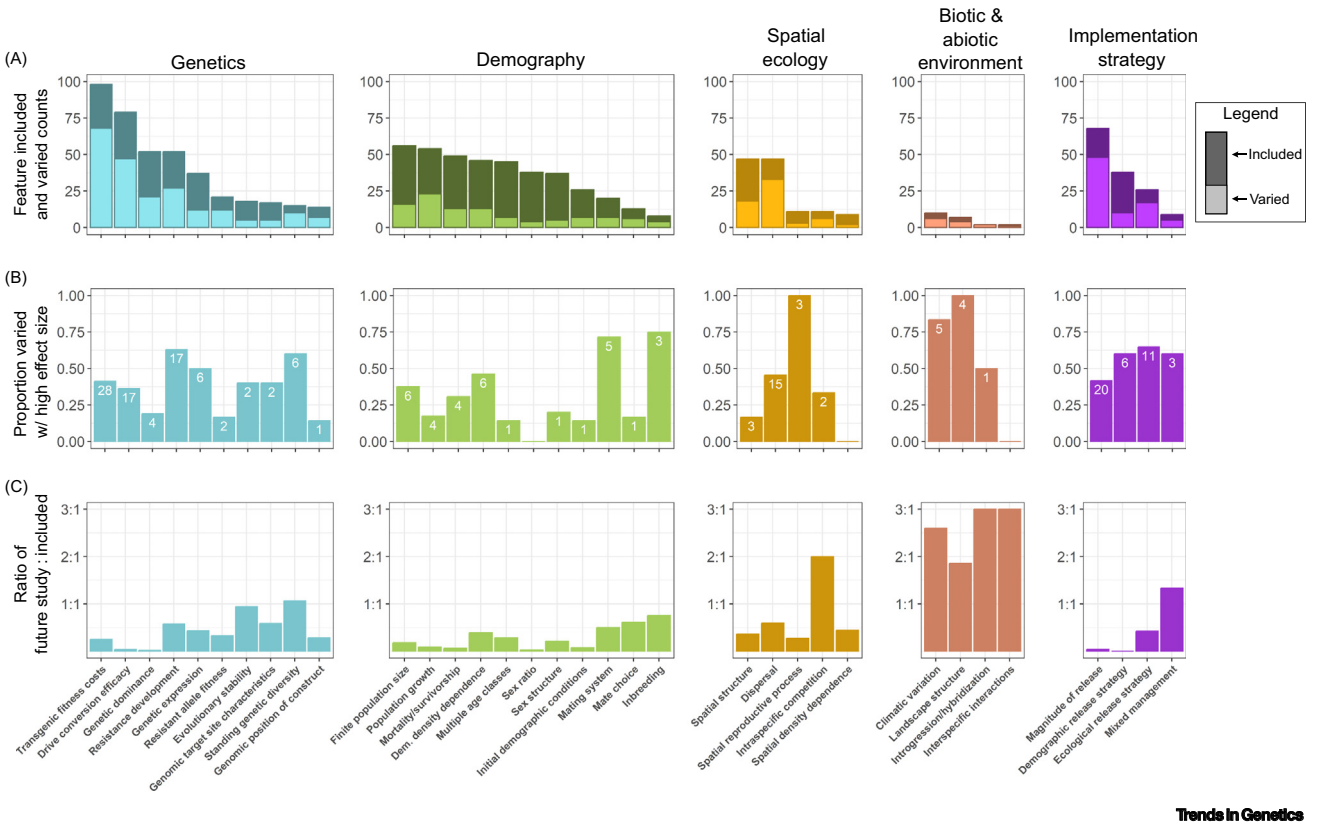

Figure 3. Summary of model features. Frequency of feature inclusion (total bar height) and variation (lighter interior bar) across published gene drive dynamic modeling studies (A). Proportion of studies reporting a high effect size for specific features, scaled by the number of studies for which that feature was varied (B). Bar labels (white text) denote total number of studies reporting high effect size. Ratio of studies mentioning a feature as important for future study compared to the number of studies that included the feature in a gene drive model (C).

To date, GDDMs exhibit limited attention to mixed management strategies (i.e., gene drives implemented alongside other management tools) (Figure 3). Population and public health management with gene drives will necessarily involve fiscal, practical, and temporal constraints. Modeling conventional pest management and/or disease reduction strategies in comparative or complementary frameworks improves inferences on the feasibility of accomplishing gene drive goals within these contexts [42,43]. Although many modeling studies reasonably exclude mixed management to simplify study questions, gene drive implementation will not occur in a vacuum, but as a new tool within ongoing management efforts.

Transgenic fitness costs were included in models more than any other feature, indicating their fundamental importance to the trajectory of gene drive systems (Figure 3). Fitness costs appear ubiquitous across engineered drives, where increased fitness costs generally decrease drive success likelihood. Although high effect sizes were reported for less than half the models that varied fitness costs (Figure 3), they were commonly found to influence, and be influenced by, other features such as inbreeding [44], **dispersal** [45], and ecological release strategy [46], highlighting the capability of GDDMs to reveal how eco-evolutionary relationships interact to structure gene drive outcomes. Many studies assume fitness costs are stable across time and environmental variation but empirical data on relative fitness of transgenic individuals for parameterizing models remains limited and system specific [47]. Our understanding of transgenic fitness costs,

particularly in the context of additional system complexity, limits our ability to accurately predict real-world outcomes and for the applicability of gene drive models in RAs.

Network analysis of shared feature inclusion within studies indicated that commonly included features are commonly studied together to provide insight into eco-evolutionary dynamics between critical processes such as spatial structure and transgenic fitness costs or resistance development and drive conversion efficacy (Figure 4). However, several feature pairs appear to deserve increased research scrutiny, as they were commonly included across studies individually but

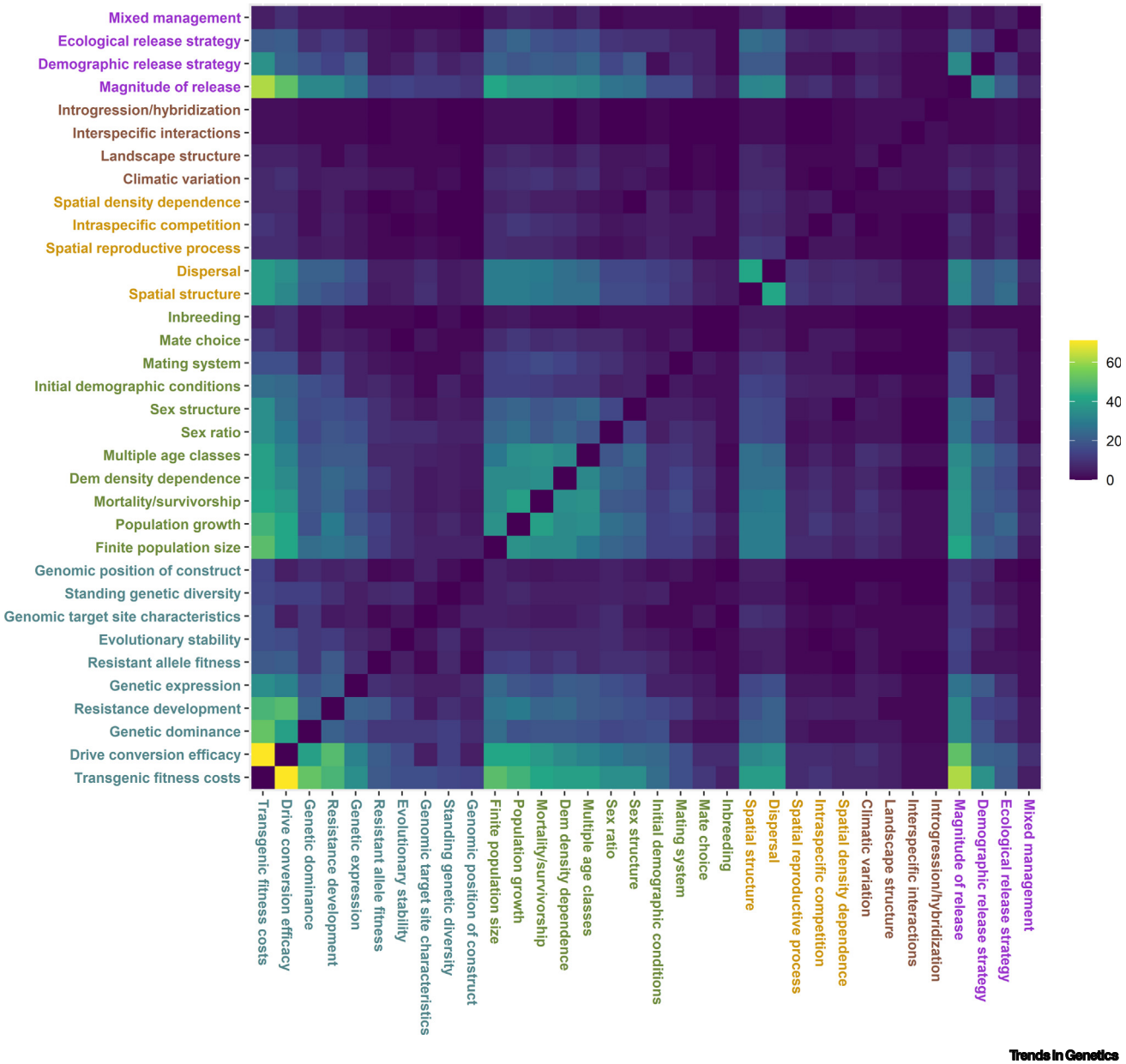

Figure 4. Network heatmap of shared feature inclusion within individual DPGMs. Heatmap color indicates the number of studies in which both features were included. Feature labels are organized and colored based on categories defined in Figure 1.

rarely studied together, despite potential interdependence. These include how genomic target site characteristic interacts with drive conversion efficacy (e.g., via sequence diversity or functional constraints) or how initial demographic conditions interact with demographic release strategy (e.g., via shifting release thresholds).

Emergent principles of gene drive dynamics

Directly comparing model outcomes related to specific GDDM features remains difficult due to the diversity of interactions among taxa-specific traits, receiving environment ecology, gene drive mechanisms, and mathematical assumptions within models. Despite this challenge, several generalizable principles emerged when examining results reported for the top two features in each category, based on proportion of high effect sizes. These features strongly influenced gene drive outcomes in >50% of studies where they were varied (Figure 3).

'Principle 1: resistance development and standing genetic diversity are key limitations to gene drive success'. For homing-based drives, variation at homing sites due to standing genetic diversity in target populations increases the likelihood of resistant allele formation and reduces the likelihood of drive success [43,48–52], prompting research on methods to reduce resistance development. For example, targeting functionally constrained genes can reduce resistant allele formation [53]. While early studies indicated that increasing the number of gRNAs (i.e., multiplexing) causes exponential decreases in resistant allele generation [23], recent models incorporating realistic timing of cleavage and repair indicate an optimal number of gRNAs (usually 2–8) exists for each system [54]. Alternative gene drive architectures such as split-drive systems can reduce exposure to adverse gene drive outcomes while allowing study of resistance formation within wild-type populations [55].

'Principle 2: deviations from perfectly monogamous and non-assortative mating systems strongly influence population outcomes'. Asymmetric mating systems (e.g., polygyny and polyandry) are almost always found to reduce drive success probability relative to random mating [19,56,57]. An exception to this trend is a twin-drive system that bypasses sperm competition by enabling terminator males to deliver toxic effector seminal proteins that kill or sterilize females, where polygynous mating increases drive success by increasing the probability that females mate with GDMO males [58]. Developers should consider leveraging asymmetric mating systems to improve outcomes in sex biasing drives. Studies including inbreeding always found that increased sibling mating reduces gene drive speed and success likelihood, while highlighting the important role of premating dispersal, which alleviates these effects [40,41,44]. Predicting the evolution of increased inbreeding or the magnitude of Allee effects experienced across small populations remains difficult and target-species dependent.

'Principle 3: dispersal behavior exerts system-dependent effects on gene drive dynamics'. High dispersal rates can accelerate spread of gene drive constructs to less connected subpopulations [59], but also limit drive success when locally extinct subpopulations are recolonized by wild-type individuals [27]. Lower dispersal rates can improve the likelihood of successful remediation drives [60] but also lower the likelihood of target populations experiencing sufficient changes within anticipated time periods [61]. The desirability of these dispersal effects depends on the intended extent of gene drive spread; excess long-distance dispersal hamper drives intended for limited geographic spread, while spatially restricted dispersal slows progression of low-threshold drives intended for unrestricted spread. Models exploring spatial reproductive processes (e.g., premating dispersal) found those traits have a strong influence on drive speed and spread [44,62,63]. The interplay between life history, dispersal, and evolutionary trajectory varies among populations and has been difficult to predict. However, GDDMs provide a platform to explore how

spatial processes influence gene drive outcomes to inform RAs, underscoring the importance of gathering relevant natural history datasets to develop well-informed prior distributions (e.g., dispersal kernels, population genetic structure) for target populations.

'Principle 4: landscape structure and climatic variation are key determinants of drive success and spread'. These environmental features alter local carrying capacity and migration dynamics across space and time [20]. For mosquito systems, seasonality creates major shifts in drive dynamics; studies incorporating these effects help optimize release strategies [38]. RAs for gene drive field trials or staged releases will benefit greatly from case-specific models in which landscape and climate effects are evaluated (e.g., how habitat distribution or seasonality impact disease transmission outcomes), which are likely difficult to simulate in laboratory studies and may not be easily generalizable.

'Principle 5: releasing GDMOs from more sites, over larger areas, and across multiple time points increases effectiveness'. These strategies tend to lower required release thresholds especially for underdominance systems [29] or when fitness costs limit the efficacy of single releases [46]. Conversely, demographic release strategy modeling has uncovered few generalizable trends. Instead, outcomes depend largely on drive mechanism and target life-history characteristics. For example, for conventional killer-rescue mosquito systems, releasing both sexes is most effective [64], but for killer-male rescue-female mosquito systems, an all-male release is necessary [65]. Ultimately, RAs must balance modeling recommendations against the uncertainty posed by practical realities of GDMO introduction, such as whether mosquito sexing accuracy limits capacity to ensure all-male releases.

Integrating dynamic models into RA

As gene drive products seek to transition from laboratory development to field trials (i.e., Phase 1 to Phase 2), decision-makers will rely on RAs [66–68]. Several modeling approaches commonly contribute to RAs, each with strengths and weaknesses. Qualitative models are valuable for communicating generalizable concepts and system-level understanding (e.g., directional relationships among model features) and can be developed quickly. However, these models fail to provide precise estimates required by some RAs, such as specific threshold values for features mediating drive success. Statistical models are commonly used for RAs. They quantify outcomes of interest using empirical data to explain system features. However, their ability to generalize outcomes across unsampled systems or identify emergent system-level behavior is limited. Also, prior to field trials, it may be impossible to collect empirical data on important gene drive dynamics such as interspecific interactions, response to variable climate and landscape heterogeneity, or dispersal, though naturally occurring drives provide safe and useful experimental systems [69].

Dynamic process models (e.g., GDDMs) provide quantitative outputs that reflect system-level behavior by predicting how ecological and evolutionary processes interact to produce population and ecosystem-level changes, making them well-suited to model causal pathways and inform gene drive RAs. Yet, their precision and applicability depend on the accuracy with which the model structure and parameterization represents the gene drive system. GDDMs can quantify key outcomes such as likelihood of complete population eradication/replacement [43,70] and expected spread over space and time [20,28] (i.e., invasion dynamics), impacts on nontarget populations and ecological communities [71,72], and the influence of eco-evolutionary feedback, such as how fitness effects influence spatial spread or how shifting population size influences remediation drive efficacy [25,73]. They can reveal how technical considerations like genetic conversion mechanisms or release strategies modify risk probabilities [74,75]. Risk assessors can

use predicted outcomes from properly evaluated GDDMs to inform analyses of environmental, socioeconomic, and epidemiological risks without introducing GDMOs into the environment.

RAs begin by identifying hazards and their adverse outcomes, often using the problem formulation approach [76,77]. While checklist approaches are commonly used to determine whether known hazards apply to new systems, they may neglect novel hazards [78]. By modeling underlying biological processes, GDDMs can reveal novel hazards as emergent properties of complex systems that are otherwise difficult to predict. For example, [41] highlights chasing dynamics that result when wild-type individuals recolonize areas previously eradicated by gene drives, increasing the potential for resistant allele evolution and reducing likelihood of complete population suppression/replacement. While chasing is a generalizable hazard applicable across gene drive systems, target-specific hazards have been revealed by incorporating species-level genetic and demographic processes. For example, GDDMs illustrated how haplodiploidy and inbreeding effects impede suppression of the honey bee parasite *Varroa destructor* [79].

Dynamic models provide additional utility by characterizing risk mitigation strategies. Regulatory approval for gene drive products may require a locally restricted, reversible, and/or high-threshold drives, given the risk of unintentional geographic spread and to nontarget populations from low-threshold gene drives [59]. A growing suite of models explore the feasibility of threshold-dependent and other gene drive mechanisms intended to reduce risks of uncontrolled spread including locally fixed alleles [80,81], self-eliminating gene drives [82], reversible gene drives [60,83], toxin-antidote drives (e.g., *Medea* systems) [48,84], and split-drive systems [52,85]. These studies provide risk assessors with comparative and proof-of-concept analyses for risk mitigation under both idealized laboratory environments and variable conditions, as demographic, spatial, and environmental complexity is incorporated. Additionally, dynamic process models can benefit post-release monitoring (i.e., Phase 4) by informing sampling efforts, and in turn, monitoring data can be used to evaluate and improve model performance [17].

Complexity, uncertainty, and limitations of gene drive models

Despite their utility, GDDMs face several challenges including sources of model uncertainty and tradeoffs surrounding model complexity, which might limit trust in model predictions. Model structure can generate uncertainty when biological processes and their interactions are represented inaccurately or ignored, often through simplifying assumptions, which can obscure certain outcomes and bias results. For example, biological inferences can change when models vary the order of genetic or spatial processes to reflect biological complexity [44,54]. Processes are sometimes excluded or simplified due to poor data availability, particularly for genetic and life history characteristics of target populations, species abundances across introduction environments, and the probability of important events (e.g., variable DNA repair mechanisms, long-distance migration). Lack of relevant data across gene drive systems also contributes model uncertainty through inaccurate parameterization, which can cause misleading results [86].

As modelers seek to reduce uncertainty, inherent tradeoffs occur as biological complexity is encoded or ignored. Simpler models incorporating fewer biological processes exhibit reduced variance but increased prediction bias, while predictions from more complex models often exhibit wider confidence intervals but are more likely to capture the truth. Simpler GDDMs can be more flexibly applied to diverse systems, making them attractive for risk assessors, but may ignore important system-specific features. By tailoring GDDMs to unique characteristics of each drive system and target population, model complexity enables more accurate biological inference and risk analysis. However, each additional feature increases the difficulty of accurate model construction, parameterization, and validation, creating opportunities for increased model uncertainty, where

inaccurate representations can cascade across biological scales (e.g., genetic expression affects fitness costs, affecting response to climate variation, and spatial spread). Indeed, a single model that sufficiently encompasses all relevant genetic, demographic, spatial, ecological, and implementation features is difficult to construct and interpret, limiting practicality for risk assessment. Balancing the complexity of GDDMs to optimize accuracy, uncertainty, and application utility will be a persistent challenge as the field develops.

The inherent uncertainty and tradeoffs across GDDMs create challenges for their formal incorporation within RA frameworks. If RA practitioners do not understand the assumptions and limitations of modeling approaches, nor participate in model development, their trust and investment in GDDM applications may wane. These issues highlight the lack of guidelines for model development or levels of model uncertainty and accepted protocols for model evaluation within RA frameworks. Extant recommendations largely fail to describe which gene drive features should be examined across research phases, or how system-specific attributes might influence modeling approaches and applications [7,8].

Concluding remarks and future perspectives

Gene drive modeling benefits from an iterative and adaptive approach integrated within the phased research pathway (Figure 5). As gene drive systems are designed, early models use simpler representations incorporating genetic and demographic features relevant to drive efficacy and relative fitness. Model evaluation provides rationale for further data collection to validate and/or improve prior distributions of model parameters, enabling more confident incorporation of increasingly complex and accurate biological processes as projects move through development phases. Model outcomes and uncertainty can inform and improve RA with each iterative improvement. The Target Malaria project exemplifies this process, wherein risk assessment priorities help structure experimental research progression [87,88]. Using a nondrive transgenic sterile male system, Yao *et al.* completed recent field trials in Burkina Faso, providing the first empirical data on realized fitness costs for transgenic mosquitoes in nature [89], which were compatible with estimates drawn from large cage experiments [90,91] and probabilistic risk predictions [92].

Over time, the average number of model features included in GDDMs has increased (see Figure S1 in supplemental information online). However, it is important to recognize that building more complex models may not translate to improved capacity for performing RAs, which often evaluate specific endpoints concerning limited numbers of features or interactions. Simpler models may be more appropriate if they adequately inform RA needs for particular applications. By ignoring features less relevant to causal pathways being assessed, parsimonious models may provide clearer predictions about features of interest. For example, assessment of geographical containment for a field trial may ignore certain genetic processes, while analyses of construct evolutionary stability may include multiple genetic processes but ignore effects of environmental heterogeneity. An important avenue is developing consensus among geneticists, modelers, risk assessors, and stakeholders on modeling priorities, model evaluation, and guidance for making decisions given model uncertainty (Figure 5) (see Outstanding questions) [93]. Model-informed RAs can also use ensemble modeling approaches, wherein multiple independent models are synthesized to account for uncertainty within each model to identify more robust and generalizable outcomes [21].

Modelers can help risk assessors embrace GDDMs by delivering tools that enable exploration of biological features and their dynamic interactions across variable parameter spaces. For example, a searchable database of modeling results could be included within a recently proposed gene drive registry [94], helping standardize and contextualize outcomes based on drive goal,

Outstanding questions

What degree of biological complexity is required to sufficiently analyze population outcomes and environmental risks from gene drives, and how does this vary across target populations and species?

How can we improve our probability estimates of rare events like nontarget spread via hybridization?

How accurately can relative fitness of transgenic individuals be predicted before field trials? What approaches to measuring fitness are most useful for informing models and RA?

How do transgenic fitness costs vary across variable environments?

How should modeling and risk assessment expectations vary for localized or nonlocalized, high-threshold or low-threshold drives?

Can exploratory models testing hypotheses about specific features be easily applied to RA?

What model evaluation and validation procedures should be implemented to address uncertainty and acceptability of model predictions for relevant stakeholders and decision makers at different research stages?

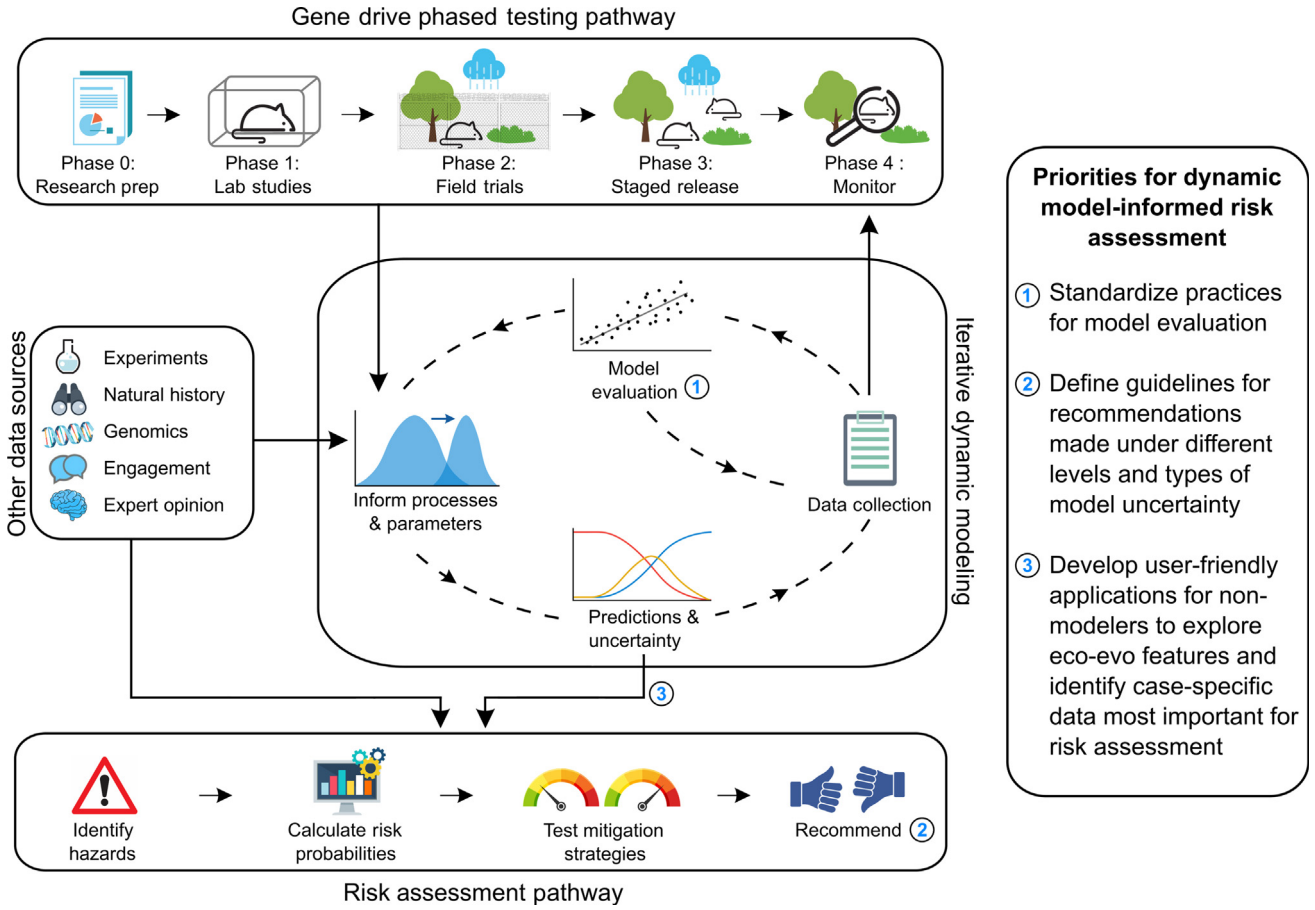

Figure 5. Iterative process of gene drive modeling and application to risk assessment. Dynamic models representing gene drive systems are developed early as projects progress through the phased research pathway. Data from gene drive experiments and other data sources inform model processes and parameters. Model predictions and associated uncertainties inform risk assessments and inform needs for additional data, which can be used to evaluate models and further refine data requirements. New data sources can iteratively update models and risk assessments as the process continues.

technology, target species, eco-evolutionary features, and mathematical assumptions. Similarly, modular simulation platforms like MGDrivE2 can be provided in user-friendly formats to improve intuition about gene drive dynamics [38]. Such tools improve capacity for independent model review if concerns exist about potential conflicts of interest when product developers also develop models required for regulatory approval. These modeling tools have shown promise for RA applications in mosquitoes and mice [42,95]. Adapting such tools into GUI-based programs can further improve approachability and collaboration by increasing accessibility to nontechnical audiences, albeit with more limited modeling options and reduced emphasis on underlying assumptions (e.g., DrMxR [96]). Risk assessors could be encouraged to check the rationale for parameter values proposed in gene drive technology applications, and if necessary, explore the effects of alternative parameter values to identify potential risks, parameters, and eco-evolutionary contexts that may require further investigation before release. Making sense of complex systems is a major benefit of dynamic modeling, given sufficient knowledge of their development and interpretation. Facilitating risk assessors' access to and investment in GDDMs will benefit this nascent field, and ultimately, our potential to translate novel biotechnologies into safe and effective solutions for pressing environmental issues.

Acknowledgments

We thank the USDA APHIS Science Fellows Program for funding that supported this research. We thank three anonymous reviewers for their constructive feedback, which improved this manuscript. Free use graphics provided by macrovector, mamewmy, anindyanfitri, brgfx, terdpongvector, and starline on Freepik.

Declaration of interests

The authors declare no conflicts of interest.

Supplemental information

Supplemental information associated with this article can be found online at https://doi.org/10.1016/j.tig.2023.04.004.

References

1. Alphey, L.S. *et al.* (2020) Standardizing the definition of gene drive. *Proc. Natl. Acad. Sci. U. S. A.* 117, 30864–30867
2. Bier, E. (2022) Gene drives gaining speed. *Nat. Rev. Genet.* 23, 5–22
3. Champer, J. *et al.* (2016) Cheating evolution: engineering gene drives to manipulate the fate of wild populations. *Nat. Rev. Genet.* 17, 146–159
4. Scott, M.J. *et al.* (2018) Agricultural production: assessment of the potential use of Cas9-mediated gene drive systems for agricultural pest control. *J. Responsible Innov.* 5, S98–S120
5. Rode, N.O. *et al.* (2019) Population management using gene drive: molecular design, models of spread dynamics and assessment of ecological risks. *Conserv. Genet.* 20, 671–690
6. Macias, V. *et al.* (2017) Gene drive for mosquito control: where did it come from and where are we headed? *IJERPH* 14, 1006
7. World Health Organization, ed (2021) *Guidance Framework for Testing Genetically Modified Mosquitoes*, 2nd ed. World Health Organization
8. National Academies of Sciences, Engineering, and Medicine *et al.* (2016) *Gene Drives on the Horizon: Advancing Science, Navigating Uncertainty, and Aligning Research with Public Values*, National Academies Press
9. Min, J. *et al.* (2018) Harnessing gene drive. *J. Responsible Innov.* 5, S40–S65
10. Renn, O. (2008) White paper on risk governance: toward an integrative framework. In *Global Risk Governance 1* (Renn, O. and Walker, K.D., eds), pp. 3–73, Springer, Netherlands
11. Chaos, Á. *et al.* (2006) From genes to flower patterns and evolution: dynamic models of gene regulatory networks. *J. Plant Growth Regul.* 25, 278–289
12. Geard, N. and Willadsen, K. (2009) Dynamical approaches to modeling developmental gene regulatory networks. *Birth Defects Res. C Embryo Today Rev.* 87, 131–142
13. Riley, S. (2007) Large-scale spatial-transmission models of infectious disease. *Science* 316, 1298–1301
14. Lloyd-Smith, J.O. *et al.* (2009) Epidemic dynamics at the human-animal interface. *Science* 326, 1362–1367
15. Corrales, X. *et al.* (2020) Advances and challenges in modelling the impacts of invasive alien species on aquatic ecosystems. *Biol. Invasions* 22, 907–934
16. Buchadas, A. *et al.* (2017) Dynamic models in research and management of biological invasions. *J. Environ. Manag.* 196, 594–606
17. Golnar, A.J. *et al.* (2021) Embracing dynamic models for gene drive management. *Trends Biotechnol.* 39, 211–214
18. Noble, C. *et al.* (2017) Evolutionary dynamics of CRISPR gene drives. *Sci. Adv.* 3, e1601964
19. Manser, A. *et al.* (2019) Controlling invasive rodents via synthetic gene drive and the role of polyandry. *Proc. R. Soc. B* 286, 20190852
20. Beeton, N.J. *et al.* (2022) Spatial modelling for population replacement of mosquito vectors at continental scale. *PLoS Comput. Biol.* 18, e1009526
21. Calder, M. *et al.* (2018) Computational modelling for decision-making: where, why, what, who and how. *R. Soc. Open Sci.* 5, 172096
22. Marshall, J.M. and Hay, B.A. (2014) Medusa: a novel gene drive system for confined suppression of insect populations. *PLoS ONE* 9, e102694
23. Marshall, J.M. *et al.* (2017) Overcoming evolved resistance to population-suppressing homing-based gene drives. *Sci. Rep.* 7, 3776
24. Alphey, N. and Bonsall, M.B. (2014) Interplay of population genetics and dynamics in the genetic control of mosquitoes. *J. R. Soc. Interface* 11, 20131071
25. Girardin, L. and Débarre, F. (2021) Demographic feedbacks can hamper the spatial spread of a gene drive. *J. Math. Biol.* 83, 67
26. Paril, J.F. and Phillips, B.L. (2022) Slow and steady wins the race: spatial and stochastic processes and the failure of suppression gene drives. *Mol. Ecol.* Published online July 5, 2022. https://doi.org/10.1111/mec.16598
27. North, A.R. *et al.* (2019) Modelling the potential of genetic control of malaria mosquitoes at national scale. *BMC Biol.* 17, 26
28. Sánchez, C. *et al.* (2020) Modeling confinement and reversibility of threshold-dependent gene drive systems in spatially-explicit Aedes aegypti populations. *BMC Biol.* 18, 50
29. Champer, J. *et al.* (2020) Population dynamics of underdominance gene drive systems in continuous space. *ACS Synth. Biol.* 9, 779–792
30. Kim, J. *et al.* (2023) Incorporating ecology into gene drive modeling. *Ecol. Lett.* (in press)
31. Frieß, J.L. *et al.* (2023) Review of gene drive modelling and implications for risk assessment of gene drive organisms. *Ecol. Model.* 478, 110285
32. EFSA Panel on GMOs *et al.* (2020) Adequacy and sufficiency evaluation of existing EFSA guidelines for the molecular characterisation, environmental risk assessment and post-market environmental monitoring of genetically modified insects containing engineered gene drives. *EFSA J.* 18
33. Then, C. *et al.* (2020) Spatiotemporal controllability and environmental risk assessment of genetically engineered gene drive organisms from the perspective of European Union genetically modified organism regulation. *Integr. Environ. Assess. Manag.* 16, 555–568
34. EFSA Panel on Genetically Modified Organisms (ESFA) (2013) Guidance on the environmental risk assessment of genetically modified animals. *EFS2* 11, 3200
35. Secretariat of the Convention on Biological Diversity (SCBD) (2016) *Guidance on Risk Assessment of Living Modified Organisms and Monitoring in the Context of Risk Assessment*, United Nations Environmetal Programme
36. Magori, K. *et al.* (2009) Skeeter Buster: a stochastic, spatially explicit modeling tool for studying *Aedes aegypti* population replacement and population suppression strategies. *PLoS Negl. Trop. Dis.* 3, e508
37. Haller, B.C. and Messer, P.W. (2023) SLiM 4: multispecies eco-evolutionary modeling. *Am. Nat.* 201, E000
38. Wu, S.L. *et al.* (2021) MGDrivE 2: a simulation framework for gene drive systems incorporating seasonality and epidemiological dynamics. *PLoS Comput. Biol.* 17, e1009030
39. Deredec, A. *et al.* (2008) The population genetics of using homing endonuclease genes in vector and pest management. *Genetics* 179, 2013–2026

40. Bull, J.J. *et al.* (2019) Gene-drive-mediated extinction is thwarted by population structure and evolution of sib mating. *Evol. Med. Public Health* 2019, 66–81
41. Champer, J. *et al.* (2021) Suppression gene drive in continuous space can result in unstable persistence of both drive and wild-type alleles. *Mol. Ecol.* 30, 1086–1101
42. Brown, E.A. *et al.* (2022) Bayesian network-based risk assessment of synthetic biology: Simulating CRISPR-Cas9 gene drive dynamics in invasive rodent management. *Risk Anal.* 42, 2835–2846
43. Leung, S. *et al.* (2022) Population replacement gene drive characteristics for malaria elimination in a range of seasonal transmission settings: a modelling study. *Malar. J.* 21, 226
44. Beaghton, P.J. and Burt, A. (2022) Gene drives and population persistence vs elimination: the impact of spatial structure and inbreeding at low density. *Theor. Popul. Biol.* 145, 109–125
45. Liu, Y. and Champer, J. (2022) Modelling homing suppression gene drive in haplodiploid organisms. *Proc. R. Soc. B* 289, 10
46. Kandul, N.P. *et al.* (2021) A confinable home-and-rescue gene drive for population modification. *eLife* 10, e65939
47. Lindholm, A.K. *et al.* (2016) The ecology and evolutionary dynamics of meiotic drive. *Trends Ecol. Evol.* 31, 315–326
48. Buchman, A. *et al.* (2018) Synthetically engineered *Medea* gene drive system in the worldwide crop pest *Drosophila suzukii*. *Proc. Natl. Acad. Sci. U. S. A.* 115, 4725–4730
49. Drury, D.W. *et al.* (2017) CRISPR/Cas9 gene drives in genetically variable and nonrandomly mating wild populations. *Sci. Adv.* 3, e1601910
50. Eckhoff, P.A. *et al.* (2017) Impact of mosquito gene drive on malaria elimination in a computational model with explicit spatial and temporal dynamics. *Proc. Natl. Acad. Sci. U. S. A.* 114, E255–E264
51. Prowse, T.A.A. *et al.* (2017) Dodging silver bullets: good CRISPR gene-drive design is critical for eradicating exotic vertebrates. *Proc. R. Soc. B* 284, 20170799
52. Nash, A. *et al.* (2018) Integral gene drives for population replacement. *Biol. Open* 8, bio037762
53. Kyrou, K. *et al.* (2018) A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged *Anopheles gambiae* mosquitoes. *Nat. Biotechnol.* 36, 1062–1066
54. Champer, S.E. *et al.* (2020) Computational and experimental performance of CRISPR homing gene drive strategies with multiplexed gRNAs. *Sci. Adv.* 6, eaaz0825
55. Champer, J. *et al.* (2019) Molecular safeguarding of CRISPR gene drive experiments. *eLife* 8, e41439
56. Wilkins, K.E. *et al.* (2018) Pest demography critically determines the viability of synthetic gene drives for population control. *Math. Biosci.* 305, 160–169
57. Gierus, L. *et al.* (2022) Leveraging a natural murine meiotic drive to suppress invasive populations. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2213308119
58. Hurtado, J. *et al.* (2022) Propagation of seminal toxins through binary expression gene drives could suppress populations. *Sci. Rep.* 12, 6332
59. Noble, C. *et al.* (2018) Current CRISPR gene drive systems are likely to be highly invasive in wild populations. *eLife* 7, e33423
60. Chae, D. *et al.* (2020) Chemical controllable gene drive in *Drosophila*. *ACS Synth. Biol.* 9, 2362–2377
61. Huang, Y. *et al.* (2011) Gene-drive into insect populations with age and spatial structure: a theoretical assessment: theoretical assessment of gene-drive. *Evol. Appl.* 4, 415–428
62. Akbari, O.S. *et al.* (2013) A synthetic gene drive system for local, reversible modification and suppression of insect populations. *Curr. Biol.* 23, 671–677
63. Beaghton, A. *et al.* (2016) Gene drive through a landscape: Reaction–diffusion models of population suppression and elimination by a sex ratio distorter. *Theor. Popul. Biol.* 108, 51–69
64. Legros, M. *et al.* (2013) Modeling the dynamics of a non-limited and a self-limited gene drive system in structured *Aedes aegypti* populations. *PLoS ONE* 8, e83354
65. Marshall, J.M. *et al.* (2011) *Semele* : A killer-male, rescue-female system for suppression and replacement of insect disease vector populations. *Genetics* 187, 535–551
66. Devos, Y. *et al.* (2022) Risk management recommendations for environmental releases of gene drive modified insects. *Biotechnol. Adv.* 54, 107807
67. Connolly, J.B. *et al.* (2022) Recommendations for environmental risk assessment of gene drive applications for malaria vector control. *Malar. J.* 21, 152
68. Devos, Y. *et al.* (2022) Potential use of gene drive modified insects against disease vectors, agricultural pests and invasive species poses new challenges for risk assessment. *Crit. Rev. Biotechnol.* 42, 254–270
69. Manser, A. *et al.* (2020) Polyandry blocks gene drive in a wild house mouse population. *Nat. Commun.* 11, 5590
70. Birand, A. *et al.* (2022) Scalability of genetic biocontrols for eradicating invasive alien mammals. *NB* 74, 93–103
71. Greenbaum, G. *et al.* (2021) Designing gene drives to limit spillover to non-target populations. *PLoS Genet.* 17, e1009278
72. Backus, G.A. and Gross, K. (2016) Genetic engineering to eradicate invasive mice on islands: modeling the efficiency and ecological impacts. *Ecosphere* 7, e01589
73. Rode, N.O. *et al.* (2020) Can a population targeted by a CRISPR-based homing gene drive be rescued? *G3 Genes|Genomes|Genetics* 10, 3403–3415
74. Dhole, S. *et al.* (2018) Invasion and migration of spatially self-limiting gene drives: a comparative analysis. *Evol. Appl.* 11, 794–808
75. Edgington, M.P. and Alphey, L.S. (2017) Conditions for success of engineered underdominance gene drive systems. *J. Theor. Biol.* 430, 128–140
76. Roberts, A. *et al.* (2017) Results from the workshop "problem formulation for the use of gene drive in mosquitoes.". *Am. J. Trop. Med. Hyg.* 96, 530–533
77. Connolly, J.B. *et al.* (2021) Systematic identification of plausible pathways to potential harm via problem formulation for investigational releases of a population suppression gene drive to control the human malaria vector *Anopheles gambiae* in West Africa. *Malar. J.* 20, 170
78. Hayes, K.R. *et al.* (2018) Identifying and detecting potentially adverse ecological outcomes associated with the release of gene-drive modified organisms. *J. Responsible Innov.* 5, S139–S158
79. Faber, N.R. *et al.* (2021) A gene drive does not spread easily in populations of the honey bee parasite Varroa destructor. *Apidologie* 52, 1112–1127
80. Sudweeks, J. *et al.* (2019) Locally fixed alleles: a method to localize gene drive to island populations. *Sci. Rep.* 9, 15821
81. Willis, K. and Burt, A. (2021) Double drives and private alleles for localised population genetic control. *PLoS Genet.* 17, e1009333
82. Zapletal, J. *et al.* (2021) Making gene drive biodegradable. *Philos. Trans. R. Soc. B* 376, 20190804
83. Heffel, M.G. and Finnigan, G.C. (2019) Mathematical modeling of self-contained CRISPR gene drive reversal systems. *Sci. Rep.* 9, 20050
84. Champer, J. *et al.* (2020) A toxin-antidote CRISPR gene drive system for regional population modification. *Nat. Commun.* 11, 1082
85. Edgington, M.P. *et al.* (2020) Split drive killer-rescue provides a novel threshold-dependent gene drive. *Sci. Rep.* 10, 20520
86. Xu, C. *et al.* (2010) Understanding uncertainties in model-based predictions of Aedes aegypti population dynamics. *PLoS Negl. Trop. Dis.* 4, e830
87. Hayes, K.R. *et al.* (2018) *Risk Assessment for Controlling Mosquito Vectors with Engineered Nucleases: Controlled field release for Sterile Male Construct, Risk assessment final report*, The Commonwealth Scientific and Industrial Research Organisation
88. Guissou, C. *et al.* (2022) Preparing an insectary in Burkina Faso to support research in genetic technologies for malaria control. *Vector-Borne Zoonotic Dis.* 22, 18–28
89. Yao, F.A. *et al.* (2022) Mark-release-recapture experiment in Burkina Faso demonstrates reduced fitness and dispersal of genetically-modified sterile malaria mosquitoes. *Nat. Commun.* 13, 796
90. Klein, T.A. *et al.* (2012) Infertility resulting from transgenic I-PpoI male *Anopheles gambiae* in large cage trials. *Pathog. Glob. Health* 106, 20–31
91. Valerio, L. *et al.* (2016) Comparison of model predictions and laboratory observations of transgene frequencies in continuously-breeding mosquito populations. *Insects* 7, 47
92. Ickowicz, A. *et al.* (2021) Predicting the spread and persistence of genetically modified dominant sterile male mosquitoes. *Parasit. Vectors* 14, 480

93. Augusiak, J. *et al.* (2014) Merging validation and evaluation of ecological models to 'evaludation': a review of terminology and a practical approach. *Ecol. Model.* 280, 117–128
94. Taitingfong, R.I. *et al.* (2023) Exploring the value of a global gene drive project registry. *Nat. Biotechnol.* 41, 9–13
95. Li, M. *et al.* (2020) Development of a confinable gene drive system in the human disease vector *Aedes aegypti*. *eLife* 9, e51701
96. Verma, P. *et al.* (2021) A common gene drive language eases regulatory process and eco-evolutionary extensions. *BMC Ecol. Evo.* 21, 156