
Article:
Alphey, Luke Stephen, Harvey-Samuel, Tim, Kaur, Rajdeep et al. (Accepted: 2026) Sequence mismatch between gene-drive and target-site flanking regions significantly impairs homing efficiency in Culex quinquefasciatus. Genetics. ISSN: 0016-6731 (In Press)

Title: Sequence mismatch between gene-drive and target-site flanking regions significantly impairs homing efficiency in *Culex quinquefasciatus*.

Authors

Harvey-Samuel, T.[1,2], Kaur, R.[2], Leftwich, P.T.[3], Feng, X.[4], Gantz, V.[5], Alphey, L.[6,7]

[1] School of Life Sciences, Keele University, Keele, Staffordshire, ST5 5GB, UK

[2] Arthropod Genetics, The Pirbright Institute, Woking, UK, GU24 0NF, UK

[3] School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK

[4] Key Laboratory of Biology of Crop Pathogens and Insects of Zhejiang Province, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, 310027, China

[5] Section of Cell and Developmental Biology, University of California San Diego, San Diego, California, 92093, USA

[6] Department of Biology, University of York, Wentworth Way, York, YO10 5DD, UK

[7] York Biomedical Research Institute, University of York, Heslington, York, YO10 5DD, UK

Corresponding author: Luke Alphey: luke.alphey@york.ac.uk

Keywords: Homology Directed Repair, gene-drive, Homing Endonuclease Gene, west Nile virus, avian malaria

Running Title: Target mismatch impairs homing-drive efficiency

Abstract

CRISPR/Cas9-based homing gene-drives (homing-drives) hold enormous potential as control tools for mosquito disease-vectors. These genomically-encoded technologies spread themselves through target populations by creating double-stranded DNA breaks on homologous chromosomes, into which the homing-drives are copied ('homed'). Homing is dependent on sequence homology between the genomic regions flanking the transgene insertion and the break site. Homing efficiency (i.e. copying rate) substantially impacts the power of these systems: less efficient homing-drives spread slower, have fewer applications and are more resistance-prone. Understanding what influences homing-drive efficiency is therefore vital to the successful use of these technologies. Here we report a novel mechanism by which a homing-drive's efficiency can be significantly impaired by natural sequence variation within a population into which it is spreading. Using a *kmo*-targeting 'split' homing-drive in the West Nile virus mosquito *Culex quinquefasciatus*, we found that target-site heterology (sequence mismatch between the genomic regions flanking the target cut-site and the homing-drive transgene) of less than 10% reduced homing efficiency by up to 54%. While substantial research effort has been dedicated to increasing homing-drive efficiency through optimisation of within-construct components, our results highlight that the real-world efficacy of these systems may in part depend on variation beyond these controllable factors.

Article Summary

Homing-based gene-drives are transformative pest control tools. They work by spreading through the gene-pool of a target-pest population, affecting its control. This spreading is achieved by 'homing': where the gene-drive copies itself from one homologous chromosome to another. Homing requires DNA sequences either side (i.e. flanking) the gene-drive to match those of the site it is copying into. Using a homing gene-drive in the mosquito *Culex quinquefasciatus* we assess how sequence mismatches between these flanking regions affect homing efficiency. We find even small levels of mismatch can significantly reduce homing and discuss this in the context of real-world gene-drive deployment.

Introduction

CRISPR/Cas9 'homing' gene-drives (henceforth homing-drives) are genomically encoded technologies which hold immense promise for controlling intractable pests, with notable development in disease-vectoring mosquitoes(Alphey, 2014; Anderson et al., 2024; Gantz et al., 2015; Hammond et al., 2016; Harvey-Samuel et al., 2023; Li et al., 2020; Sinkins & Gould, 2006). A homing-drive increases its allele frequency from one

generation to the next by copying ('homing') itself from one homologous chromosome to another in germline cells (Burt, 2003). For pest management, a homing-drive can be designed to simultaneously spread traits beneficial for control e.g. pathogen refractoriness or a genetic load for population suppression/eradication (Champer et al., 2016). The efficiency of the homing reaction substantially influences the population-level efficacy of a homing-drive: higher homing rates enable faster spread and higher tolerance of drive-associated fitness costs (Deredec et al., 2008).

At a molecular level, homing occurs when Cas9 expressed by a homing-drive transgene creates a double-stranded DNA break (henceforth 'cut-site') at a specific sequence on the target chromosome, into which the homing-drive is homed (Esvelt et al., 2014) (Fig 1.). Homing is mediated by the homology-directed-repair (HDR) pathway, requiring the genomic sequences flanking the homing-drive transgene insertion-site and the target allele cut-site to show homology. While substantial effort has been devoted to exploring how 'within-drive' components (e.g. regulatory elements for expressing Cas9 (Anderson et al., 2023; Hammond et al., 2021)) affect homing rates, comparatively little attention has been paid to the role of these genomic flanking regions in influencing homing efficiency. This is surprising as the impact of even minor sequence divergence between loci in substantially reducing homologous recombination rates is well established (Elliott et al., 1998). In the real-world, homing-drives will be required to function in genetically diverse 'wild' populations, which will likely also include pre-existing heterogeneity around the cut-site. Whether sequence mismatch (henceforth 'heterology') between a homing-drive's flanking genomic regions and those flanking its target allele cut-site influences homing rates is thus of considerable importance.

We explored this question using our homing-drive system in the West-Nile-virus vector mosquito *Culex quinquefasciatus* (Harvey-Samuel et al., 2023). This system consists of a 'split-drive' design where the two homing-drive components (germline-expressing Cas9 and gRNA-expressing cassettes) are integrated at independent loci - one of which (here the gRNA-cassette) is in the target allele - and maintained in separate lines (Esvelt et al., 2014). The two lines can then be crossed together to initiate the homing reaction in trans-heterozygotes. By using the *kynurinine 3-monooxygenase* (*kmo*) gene as our target allele (where homozygous null mutations give an easily visualized 'white-eye' phenotype (Purusothaman et al., 2021)) and conducting all comparison crosses in a 'controlled' genetic background, we were able to unambiguously distinguish Cas9-mediated cutting and homing, and attribute variation in these to individual *kmo* target alleles which displayed varying levels of heterology. This experimental design provided unparalleled power to assess this fundamentally important question.

Materials and methods

Ethics

Work followed procedures/protocols approved by The Pirbright Institute Biological Agents & Genetic Modification Safety Committee. All homing assays were conducted at The Pirbright Institute IS4L arthropod containment facility under the necessary safety regulations for gene-drive research.

Mosquito lines used, rearing and maintenance.

The *kmo*-gRNA line was generated previously by integrating a *kmo*-gRNA expression cassette via CRISPR/Cas9 HDR into the *kmo* gene in the TPRI (Tropical Pesticides Research Institute) genetic background. The *kmo* -/- line was generated previously by CRISPR/Cas9 knockout of the *kmo* gene in the CA (California) genetic background. The *Vasa*-Cas9 line was generated previously by integrating a Cas9 ORF under the transcriptional control of *Cx. quinquefasciatus Vasa* regulatory elements into the *Cardinal* gene in the CA genetic background via CRISPR/Cas9 HDR. The Wild-type (WT) line is the unmodified TPRI line. Rearing/maintenance followed procedures previously reported.

Mosquito experiments

Assay 1: Allele-by-allele approach

Homozygous *Vasa*-Cas9 females were first pool-crossed to WT males, producing heterozygous *Vasa*-Cas9 F1 progeny in a 50:50 TPRI/CA genetic background. Female F1 were then pool-crossed to heterozygous *kmo*-gRNA males to give the F2 generation. Trans-heterozygous *Vasa*-Cas9, *kmo*-gRNA female F2 progeny were then pool-crossed to *kmo*-/- males and F3 egg-rafts produced individually isolated, with each egg-raft being from an individual female trans-heterozygote. L3-4 stage progeny from each egg-raft were scored for presence of transgenes/eye-pigmentation (white-eyes), as previously demonstrated.

Genotyping of target allele (Assay 1 only)

After scoring, gDNA was extracted from a single WT or *Vasa*-Cas9 individual from each F3 egg-raft (Machery-Nagel Nucleospin Tissue kit, Düren, Germany). This was used to PCR-genotype the maternally-contributed *kmo* allele (i.e. the allele into which the *kmo*-gRNA had attempted to home in the previous generation). The paternally-contributed *kmo* allele was excluded by designing the reverse primer (PL506) to lie across a region deleted during the generation of the *kmo* -/- line. The primers were designed to bind to all 4 *kmo* alleles identified in a preliminary analysis of the strains utilized (A,B,C,D). PCR conditions as follows 98C-1min, (98C-30s, 72C-15s, 72C-1min) x 35, 72C-2min. 200ng of gDNA used per 50ul Q5 PCR Reaction (New England Biolabs). Primers = F(PL303): CCAACATTCACCTTCACTTCAACCACAAGC and R(PL506):

GTGAGCGTCCTCCCACCGAG. Amplicons extended c.320/540bp upstream/downstream of the cut site – sufficient distance to distinguish the 4 alleles. Reactions were run on a 1% agarose gel and bands extracted using the Machery-nagel Nucleospin Gel and PCR clean-up kit. Purified bands were Sanger sequenced using the forward primer PL303 and sequenced amplicons aligned to the TPRI *kmo* sequence.

Calculation of homing and cutting rates (Assay 1 only)

Homing rates are defined as the percentage of homologous wild-type chromosomes converted to carry the *kmo*-gRNA transgene (i.e. the percentage of WT alleles which have been affected by homing). It was estimated for each egg-raft as (200 x (no. *kmo*-gRNA/total progeny – 0.5). Cutting rate is defined as the percentage of homologous wild-type chromosomes observed as having been cut by Cas9. This includes DSBs that were repaired via HDR (homing), NHEJ (or other error-prone mechanisms) but excludes repairs which resulted in no observable change to the WT allele (perfect repair). It was estimated for each egg-raft as the previously estimated homing rate + (200 x (no. non-*kmo*-gRNA progeny which showed white-eyes/total progeny – 0.5). That is, the HDR + error-prone pathway rates.

Assay 2: Generalising findings

Homozygous *Vasa*-Cas9 females were first pool-crossed to WT males, producing heterozygous *Vasa*-Cas9 F1 progeny in a 50:50 TPRI/CA genetic background. This was repeated twice to give F3 *Vasa*-Cas9 progeny in a predominantly TPRI genetic background (c. 87.5% TPRI alleles). Male and female *Vasa*-Cas9 heterozygote F3s were pool-crossed to female and male *kmo*-gRNA line heterozygotes, respectively, giving two cohorts in the F4 generation (one from each of the two F3 crosses). Trans-heterozygous male and female F4 progeny from each of these two crosses were pool-crossed to female and male *kmo* -/- individuals, respectively (4 crosses total), and the egg rafts produced individually isolated and scored as for Assay 1.

Statistical analysis

Homing and cutting rates were both analysed with generalized linear mixed models with a binomial (logit) error distribution. Fixed effects for Assay 1 were the genotyped maternally contributed *kmo* allele with individual female egg-batch as a random effect. Minor overdispersion in the cutting analysis models meant all 95% confidence intervals were checked by semi-parametric bootstrapping (1000 iterations). Results are presented with model-based standard errors and 95% confidence intervals; bootstrap validation confirmed these estimates were robust to minor overdispersion and both summaries are presented in supplementary materials (supplementary tables 1 & 2). For Assay 2 fixed effects included the sex of the Cas9-bearing parent and grandparent as well as whether

the data came from the original (i.e. previously published) experiment or post-introgression (i.e. Assay 2). Individual female egg-batches were also included as a random effect. All models and simulations were constructed in R version 4.3.3 with the package lmerTest, model fits were checked with the package DHARMa and model predictions generated with the package emmeans. Data processing and visualization was performed with the tidyverse packages.

Results and Discussion

Our aim was to investigate the effect of varying the heterology of sequences flanking the cut-site of a target allele on a homing-drive's homing efficiency. We assessed this at two distinct levels, each specifically designed to give insight into the behavior of this potential effect. Firstly, we took an 'allele-by-allele' approach – assessing a homing-drive's efficiency as it was paired against four different *kmo* target alleles in a controlled genetic background. Secondly, we extended this analysis to a broader range of cross schemes and sexes to assess the generality of findings.

1: Allele-by-allele approach

In 'split' homing-drives, one of the expression cassettes (here, that expressing the gRNA) is integrated into the target allele at the precise cut-site, forming the 'homing-element'. This can then be combined with the other element, here encoding Cas9, to provide trans-heterozygotes (aka "double heterozygotes") which have one copy each of Cas9 and homing-element and so can catalyse the homing reaction. Furthermore, and crucially for our purpose, in this set-up, the target allele into which a homing-element attempts to home will always come from the line (and therefore the genetic background) which contributed the Cas9 transgene. Preliminary analysis of our highly inbred wild-type (WT) line (in which the *kmo*-gRNA line was generated and is outcrossed to each generation – TPRI genetic background – see methods) and the *Vasa*-Cas9 line (which was generated separately in the CA genetic background – see methods) showed that these two lines possessed highly divergent *kmo* allele sequences. Those carrying the TPRI genetic background contained only a single *kmo* sequence (allele A) which perfectly matched the sequences flanking the integrated *kmo*-gRNA homing-element, while the *Vasa*-Cas9 line contained three different *kmo* allele sequences (alleles B, C and D – back calculated from subsequent genotyping data to be represented at frequencies of 0.15, 0.65 and 0.2, respectively, in that strain) (See figure 2A), each of which showed substantial sequence variation as compared to allele A, especially in the area 5' of the gRNA target site which harbors intronic sequence (See Supplementary Figure 1). We took advantage of this to design an assay in which a Cas9-bearing individual could contribute a variety of target alleles to a homing cross, some but not all of which would match the homing-element flanking sequences but, critically, all of which would end up in the same mixed genetic background

(see figure 2B). These female trans-heterozygotes (*kmo*-gRNA/+ $^{A,B,C\ or\ D}$, *Vasa*-Cas9/+) were then crossed to male *kmo* -/- homozygotes to allow us to detect both homing (deviation in inheritance rate of homing-element from Mendelian expectation of 50%) and the rate of end-joining (individuals inheriting loss-of-function *kmo* alleles from their mother, generated via Cas9 cutting followed by error-prone repair, will have white eyes but lack the fluorescent marker of the homing-element) in their progeny. This assay showed a highly significant effect of the *kmo* target allele (i.e. A,B,C or D) on homing rate (LRT: χ^2_3 = 16.17, p = 0.001), with a 54% difference in estimated homing rates associated with the highest (A = 0.37 [95% CI; 0.31 – 0.42]) and the lowest (D = 0.17 [0.06 – 0.27]) *kmo* target alleles (Figure 3, Supplementary Figure 2, Supplementary Table 1, Supplementary Table 2). Preliminary sequencing confirmed that the gRNA target sequence was present in all four *kmo* alleles and this was confirmed by our comparison of estimated cutting rates (i.e. estimated homing + observed end-joining) between the *kmo* alleles, which did not significantly differ (LRT: χ^2_3 = 5.27, p = 0.15; Supplementary Figure 2, Supplementary Table 1, Supplementary Table 2). This suggests that the differences in homing we observed between the four target *kmo* alleles were not simply driven by Cas9 being able to cut some alleles more efficiently than others. Instead, taken together, these data strongly suggest that the mechanism underlying the observed reduction in homing is impairment of homology-directed repair through drive-target allele flanking sequence mismatch. Of note, this effect was observable despite the fact that the vast majority of sequence heterology occurred only on one side (i.e. 5') of the target site. While the nature of available *kmo* alleles did not allow us to investigate heterology structure beyond this, results from *Drosophila melanogaster* 'transcomplementing' homing-drives (López Del Amo et al., 2020a) suggest that the presence of heterologous sequences on both sides of the target site may further impair homing efficiency. Further experiments with a wider variety of *kmo* alleles, or employing target genes with less conservation in coding regions, would enable exploration of the effect of more evenly distributed heterology on homing.

2: Generalised approach

In the real world, homing-drive transgenes could be inherited from, and be active in, either sex. Previous work has shown that the efficiencies and dynamics of homing-drives can vary dramatically dependent on direction of inheritance (paternal/maternal) in parents and grandparents (e.g. (Anderson et al., 2024; López Del Amo et al., 2020b; Terradas et al., 2021)). As such, we were interested in investigating the extent to which the observed influence of heterology on homing rates was maintained when a wider variety of cross combinations were considered. To achieve this, we first introgressed the *Vasa*-Cas9 transgene into the TPRI genetic background by outcrossing *Vasa*-Cas9 individuals to our TPRI wild-type line for three consecutive generations. This TPRI introgressed *Vasa*-Cas9 line (henceforth *Vasa*-Cas9^{i}) was then crossed to the *kmo*-gRNA line in a full factorial homing-drive cross (i.e. where the *Vasa*-Cas9 and *kmo*-gRNA transgenes were inherited

from both grandparental (F0) sexes and the trans-heterozygous (F1) individuals were of either sex) and results compared to our previously published homing analysis of these two transgenes - in that case where the *Vasa*-Cas9 transgene was in a pure CA genetic background (Harvey-Samuel et al., 2023). We found that in all like-for-like comparisons the homing rates in this study were higher than in our previous study (aggregate difference 57.8% - 69.3%), only in the male-male crosses was this increase not statistically significant (Figure 4, Supplementary Table 3). This significant interaction between experiment and sex could possibly be indicative of a difference in response to flanking-site heterology between sexes, in the context of homing drives. That is, when faced with high heterology, males showed a higher level of homing than females – an effect which largely disappeared when heterology was decreased (i.e. females increased their homing rates to 'male' levels when drive and target-site flanking regions increased in homology). A possible biological explanation for this observation could be a difference in the recognition requirements for engaging HDR in the male and female germlines, however further work would be required to explore this hypothesis. Overall, these results further support our conclusion that target-site/homing-element flanking-sequence heterology can significantly affect homing efficiency.

To date, only one other study has directly explored the influence of significant (i.e. >1%) flanking-sequence heterology on homing-drive performance (in *Anopheles gambiae*) (Pescod et al., 2023). In contrast to the results presented here, that study observed no significant influence of target-site heterology on homing-efficacy, despite similar levels of sequence divergence (5.3-6.6% over c.690bp). What might explain the differences between this finding and ours? We propose three non-exclusive possibilities.

The first involves the genetic backgrounds in which these crosses were conducted. Pescod *et al*. (Pescod et al., 2023)crossed their homing-drive directly into different geographically distinct *An. gambiae* strains, each of which contained divergent target alleles. As such, each of the different homing-drive/target allele assessments occurred in a distinct genetic background. As it has been shown that (as with 'natural' homologous recombination(Mansai et al., 2011)) many loci unlinked to a homing-drive may exert significant influence over homing efficiencies (Champer et al., 2019), this experimental design may have introduced additional variation, potentially obscuring the effects as observed here. We were able to control for these potential epistatic effects by crossing, and therefore partially introgressing, our *Vasa*-Cas9 line into the TPRI background prior to homing assays, ensuring each of our *kmo* target allele comparisons occurred within (on average) the same, mixed CA/TPRI genetic background.

Secondly, there are potentially species-specific explanations for these contrasting results. Homing-drive efficiencies in *Anopheline* mosquitoes are uniquely high (often approaching or reaching 100%) (e.g. (Carballar-Lejarazú et al., 2023)). The precise reason for this is unknown but could include less stringent requirements on engaging the HDR pathway

e.g. higher tolerance to flanking-sequence heterology, or smaller flanking sequences required to initiate homing. Indeed, analysis of homing conversion tract lengths in *An. gambiae* (i.e the degree to which the flanking regions of a target allele are converted to carry heterologous sequences contained within the homing-drive flanking regions) showed that (beyond the homing-drive cassette itself) only small sequences are transferred onto the target chromosome (>80% of conversion tracts <50bp) (Pescod et al., 2024). This is substantially shorter than reported for studies considering 'natural' homologous recombination in other species (typically c. 200bp (Mansai et al., 2011)) and may signify shorter required sequences for RAD51-dependent complementary strand recognition/invasion in *Anophelines*. In contrast, conversion tract lengths in *Aedes aegypti,* another Culicine mosquito, following CRISPR-Cas9 plasmid-based knock-in (also dependent on HDR, though likely occurring in somatic rather than germline cells) were substantially longer (mean of 160bp) (Ang et al., 2022). Interestingly, and in concurrence with our results, that study in *Ae. aegypti* also identified significant negative effects of plasmid/target-site heterology on the efficiency of knock-in. Taken together, these results could suggest that the HDR-pathway in *An. gambiae* is simply less affected by sequence heterology than in other species, including Culicine mosquitoes.

Finally, the near-100% homing rates displayed by the drives tested in Pescod *et al* may have made it difficult to observe anything other than extremely large changes (i.e. in this case reductions) in homing, potentially underestimating or missing 'subtler' modifiers of homing-efficiency – e.g. those caused by target alleles less amenable to homing. This is because in a homing-drive, inheritance-bias (i.e. the level of homing) is described by odds (p/(1-p)) – where p = wild-type allele conversion probability - and the effect of a change in odds on the change in homing rate differs substantially depending on the initial level of homing ((Deredec et al., 2008), concept visualised in Supplementary Figure 3). In essence, for homing-drives which display very high homing-rates, even large changes (e.g. reductions) in odds *(i.e. reductions in the probability of homing)* will have only small effects on the observed levels of homing. This effect is much less pronounced in less efficient homing-drives (i.e. where homing rates are closer to 0), like that demonstrated here. The diminishing return on changes in odds means that, near 100% efficiency, subtle effects on homing may go unnoticed, complicating efforts to accurately evaluate or identify modifiers of highly effective homing-drives.

In conclusion, our results demonstrate a novel factor affecting homing-drive efficiency. We show that heterology between a homing-drive's flanking regions and those sequences flanking the target allele cut-site can significantly reduce that drive's performance. A critical question is to what extent this effect could impact homing-drive performance in the real-world. In most cases it should be assumed that homing-drives could encounter significant levels of sequence variation upon release, at least equal to the variation observed within our inbred laboratory strains, particularly in continental areas where most

gene-drive deployments will likely occur. A caveat here is that the nature and distribution of this heterology will itself vary and not necessarily be represented by that we observed here. In our system we were only able to contrast and compare those pre-existing 'natural' *kmo* alleles present. Here, even the most divergent *kmo* allele tested was still able to be targeted for Cas9 cleavage and successfully converted to carry the homing-drive, albeit at a significantly reduced rate. If the percentage reduction in homing efficiency observed here were to translate to released strains, however, this could indeed substantially impede their spread through a target population (Supplementary Figure 4), with the magnitude of this effect of course dependent on the identity and frequency of heterogenous target alleles in the wild population. A potentially serious and related issue is the generation of novel drive-resistant alleles: in our system, while sequence heterology impaired the gene-drive's ability to home, it did not impair its ability to cut. As such, a drive introduced into a wild population which shows substantial heterology may produce drive-resistance alleles via end-joining pathways at a substantially higher rate than predicted under lab conditions. While naturally high homing rates may mitigate this effect, for the vast majority of species which have been tested so far, reported homing rates are modest (Grunwald et al., 2019; Harvey-Samuel et al., 2023; Li et al., 2020; Meccariello et al., 2024; Yadav et al., 2023) and thus they may be, to a degree, affected by this mechanism. Prior knowledge of these effects for a given species, homing-drive locus and target population may allow future release programs to preempt adverse consequences on project outcomes, e.g. by increasing release rates accordingly to counteract predicted lower drive efficiency, or by introducing the homing-drive construct into a variety of alleles present within a particular targeted population, ahead of releases. Releasing multiple alleles of a gene-drive insertion into a population would also open the opportunity to employ population-level multiplexing strategies, with each of the released drive alleles working to provide resistance management for the others(Edgington et al., 2020). Additionally, our findings highlight the utility of exploring factors which affect homing-drive dynamics and performance in species/model-systems which do not display near-perfect performance. While 'high-performance' homing-drives may be the first in line for field-release, it is precisely their extreme efficiency which may make them less useful for investigating some of the underlying factors which affect the drive mechanism itself.

Our results are particularly significant as gene-drives edge ever closer to field application. They suggest that, at least in some cases, homing-drive efficiencies estimated in genetically homogenous lab populations may be unrepresentative of those which would occur in the real-world. While lab-field discrepancies are well known for other parameters relevant to genetic biocontrol, e.g. mating success, longevity or other factors relating to fitness, this is the first study we are aware of which identifies such a disparity related to the homing mechanism itself. This information will be critical in helping to set realistic expectations of these potentially game-changing systems, if and when they enter the field-testing phase.

Data Availability

The authors affirm that all data necessary for confirming the conclusions of the article are present within the article, additional information files, figures, and tables. Specifically, all data gathered during and analyzed as part of the two experiments (i.e. inheritance rates of different genotypes, estimation of cutting rates, list of sequenced allele-identities) is available in the two additional information files. Strains and plasmids are available upon request.

Author contributions

THS, XF, VG and LA conceived the project. RK, THS and PL contributed to the design of the experiments. THS, RK and PL performed the experiments and contributed to the collection and analysis of data. THS and PL wrote the manuscript. All authors edited the manuscript.

Study Funding

This work was supported by strategic funding from the UK Biotechnology and Biological Sciences Research Council (BBSRC) to The Pirbright Institute (BBS/E/I/00007033, BBS/E/I/00007038, and BBS/E/I/00007039).

Conflict of Interest

VG is a founder of and has equity interests in Symbol, Inc. and Agragene, Inc., companies that may potentially benefit from the research results described in this manuscript. VG also serves on both the company's Scientific Advisory Board and the Board of Directors of Synbal, Inc. The terms of this arrangement have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies. LA is an advisor to, and has financial or equity interest in, Synvect Inc. and Biocentis Ltd., companies operating in the area of genetic control of pest insects. The other authors declare that they have no competing interests

References

Alphey, L. (2014). Genetic control of mosquitoes. *Annual Review of Entomology*, *59*, 205–224. https://doi.org/10.1146/annurev-ento-011613-162002

Anderson, M. A. E., Gonzalez, E., Ang, J. X. D., Shackleford, L., Nevard, K., Verkuijl, S. A. N., Edgington, M. P., Harvey-Samuel, T., & Alphey, L. (2023). Closing the gap to effective gene drive in Aedes aegypti by exploiting germline regulatory elements. *Nature Communications*, *14*(1), 338. https://doi.org/10.1038/s41467-023-36029-7

Anderson, M. A. E., Gonzalez, E., Edgington, M. P., Ang, J. X. D., Purusothaman, D.-K., Shackleford, L., Nevard, K., Verkuijl, S. A. N., Harvey-Samuel, T., Leftwich, P. T., Esvelt, K., & Alphey, L. (2024). A multiplexed, confinable CRISPR/Cas9 gene drive can propagate in caged Aedes aegypti populations. *Nature Communications*, *15*(1), 729. https://doi.org/10.1038/s41467-024-44956-2

Ang, J. X. De, Nevard, K., Ireland, R., Purusothaman, D.-K., Verkuijl, S. A. N., Shackleford, L., Gonzalez, E., Anderson, M. A. E., & Alphey, L. (2022). Considerations for homology-based DNA repair in mosquitoes: Impact of sequence heterology and donor template source. *PLOS Genetics*, *18*(2), e1010060. https://doi.org/10.1371/journal.pgen.1010060

Burt, A. (2003). Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1518), 921–928. https://doi.org/10.1098/rspb.2002.2319

Carballar-Lejarazú, R., Dong, Y., Pham, T. B., Tushar, T., Corder, R. M., Mondal, A., Sánchez C., H. M., Lee, H.-F., Marshall, J. M., Dimopoulos, G., & James, A. A. (2023). Dual effector population modification gene-drive strains of the African malaria mosquitoes, *Anopheles gambiae* and *Anopheles coluzzii*. *Proceedings of the National Academy of Sciences*, *120*(29). https://doi.org/10.1073/pnas.2221118120

Champer, J., Buchman, A., & Akbari, O. S. (2016). Cheating evolution: engineering gene drives to manipulate the fate of wild populations. *Nature Reviews Genetics*, *17*(3), 146–159. https://doi.org/10.1038/nrg.2015.34

Champer, J., Wen, Z., Luthra, A., Reeves, R., Chung, J., Liu, C., Lee, Y. L., Liu, J., Yang, E., Messer, P. W., & Clark, A. G. (2019). CRISPR Gene Drive Efficiency and Resistance Rate Is Highly Heritable with No Common Genetic Loci of Large Effect. *Genetics*, *212*(1), 333–341. https://doi.org/10.1534/genetics.119.302037

Deredec, A., Burt, A., & Godfray, H. C. J. (2008). The Population Genetics of Using Homing Endonuclease Genes in Vector and Pest Management. *Genetics*, *179*(4), 2013–2026. https://doi.org/10.1534/genetics.108.089037

Edgington, M. P., Harvey-Samuel, T., & Alphey, L. (2020). Population-level multiplexing: A promising strategy to manage the evolution of resistance against gene drives targeting a neutral locus. *Evolutionary Applications*, *13*(8), 1939–1948. https://doi.org/10.1111/eva.12945

Elliott, B., Richardson, C., Winderbaum, J., Nickoloff, J. A., & Jasin, M. (1998). Gene Conversion Tracts from Double-Strand Break Repair in Mammalian Cells. *Molecular and Cellular Biology*, *18*(1), 93–101. https://doi.org/10.1128/MCB.18.1.93

Esvelt, K. M., Smidler, A. L., Catteruccia, F., & Church, G. M. (2014). Concerning RNA-guided gene drives for the alteration of wild populations. *ELife*, *3*. https://doi.org/10.7554/eLife.03401

Gantz, V. M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V. M., Bier, E., & James, A. A. (2015). Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*. *Proceedings of the National Academy of Sciences*, *112*(49). https://doi.org/10.1073/pnas.1521077112

Grunwald, H. A., Gantz, V. M., Poplawski, G., Xu, X.-R. S., Bier, E., & Cooper, K. L. (2019). Super-Mendelian inheritance mediated by CRISPR–Cas9 in the female mouse germline. *Nature*, *566*(7742), 105–109. https://doi.org/10.1038/s41586-019-0875-2

Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., Gribble, M., Baker, D., Marois, E., Russell, S., Burt, A., Windbichler, N., Crisanti, A., & Nolan, T. (2016). A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nature Biotechnology*, *34*(1), 78–83. https://doi.org/10.1038/nbt.3439

Hammond, A., Karlsson, X., Morianou, I., Kyrou, K., Beaghton, A., Gribble, M., Kranjc, N., Galizi, R., Burt, A., Crisanti, A., & Nolan, T. (2021). Regulating the expression of gene drives is key to increasing their invasive potential and the mitigation of resistance. *PLOS Genetics*, *17*(1), e1009321. https://doi.org/10.1371/journal.pgen.1009321

Harvey-Samuel, T., Feng, X., Okamoto, E. M., Purusothaman, D.-K., Leftwich, P. T., Alphey, L., & Gantz, V. M. (2023). CRISPR-based gene drives generate super-Mendelian inheritance in the disease vector Culex quinquefasciatus. *Nature Communications*, *14*(1), 7561. https://doi.org/10.1038/s41467-023-41834-1

Li, M., Yang, T., Kandul, N. P., Bui, M., Gamez, S., Raban, R., Bennett, J., Sánchez C, H. M., Lanzaro, G. C., Schmidt, H., Lee, Y., Marshall, J. M., & Akbari, O. S. (2020).

Development of a confinable gene drive system in the human disease vector Aedes aegypti. *ELife*, *9*. https://doi.org/10.7554/eLife.51701

López Del Amo, V., Bishop, A. L., Sánchez C., H. M., Bennett, J. B., Feng, X., Marshall, J. M., Bier, E., & Gantz, V. M. (2020a). A transcomplementing gene drive provides a flexible platform for laboratory investigation and potential field deployment. *Nature Communications*, *11*(1), 352. https://doi.org/10.1038/s41467-019-13977-7

López Del Amo, V., Bishop, A. L., Sánchez C., H. M., Bennett, J. B., Feng, X., Marshall, J. M., Bier, E., & Gantz, V. M. (2020b). A transcomplementing gene drive provides a flexible platform for laboratory investigation and potential field deployment. *Nature Communications*, *11*(1), 352. https://doi.org/10.1038/s41467-019-13977-7

Mansai, S. P., Kado, T., & Innan, H. (2011). The Rate and Tract Length of Gene Conversion between Duplicated Genes. *Genes*, *2*(2), 313–331. https://doi.org/10.3390/genes2020313

Meccariello, A., Hou, S., Davydova, S., Fawcett, J. D., Siddall, A., Leftwich, P. T., Krsticevic, F., Papathanos, P. A., & Windbichler, N. (2024). Gene drive and genetic sex conversion in the global agricultural pest Ceratitis capitata. *Nature Communications*, *15*(1), 372. https://doi.org/10.1038/s41467-023-44399-1

Pescod, P., Bevivino, G., Anthousi, A., Shelton, R., Shepherd, J., Lombardo, F., & Nolan, T. (2023). Measuring the Impact of Genetic Heterogeneity and Chromosomal Inversions on the Efficacy of CRISPR-Cas9 Gene Drives in Different Strains of *Anopheles gambiae*. *The CRISPR Journal*, *6*(5), 419–429. https://doi.org/10.1089/crispr.2023.0029

Pescod, P., Bevivino, G., Anthousi, A., Shepherd, J., Shelton, R., Lombardo, F., & Nolan, T. (2024). Homing gene drives can transfer rapidly between Anopheles gambiae strains with minimal carryover of flanking sequences. *Nature Communications*, *15*(1), 6846. https://doi.org/10.1038/s41467-024-51225-9

Purusothaman, D.-K., Shackleford, L., Anderson, M. A. E., Harvey-Samuel, T., & Alphey, L. (2021). CRISPR/Cas-9 mediated knock-in by homology dependent repair in the West Nile Virus vector Culex quinquefasciatus Say. *Scientific Reports*, *11*(1), 14964. https://doi.org/10.1038/s41598-021-94065-z

Sinkins, S. P., & Gould, F. (2006). Gene drive systems for insect disease vectors. *Nature Reviews Genetics*, *7*(6), 427–435. https://doi.org/10.1038/nrg1870

Terradas, G., Buchman, A. B., Bennett, J. B., Shriner, I., Marshall, J. M., Akbari, O. S., & Bier, E. (2021). Inherently confinable split-drive systems in Drosophila. *Nature Communications*, *12*(1), 1480. https://doi.org/10.1038/s41467-021-21771-7

Yadav, A. K., Butler, C., Yamamoto, A., Patil, A. A., Lloyd, A. L., & Scott, M. J. (2023). CRISPR/Cas9-based split homing gene drive targeting *doublesex* for population suppression of the global fruit pest *Drosophila suzukii*. *Proceedings of the National Academy of Sciences*, *120*(25). https://doi.org/10.1073/pnas.2301525120

Figure Legends

Figure 1: Schematic representations of CRISPR-Cas9 based homing gene-drive technologies. A) A 'global' homing gene-drive in which all the molecular components required to initiate the homing reaction are linked and integrated at a single locus in the genome which conforms to the target site of the integrated gRNA. Expression of the Cas9 enzyme and gRNA thus mediates the creation of a double-stranded break (DSB) at the insertion locus on the homologous chromosome. This DSB can be repaired through the Homology-Dependent Repair (HDR) pathway which results in the copying of the gene-drive transgene onto the homologous chromosome. If this process occurs in germline cells, it will result in super-mendelian inheritance of the gene-drive. The efficiency of HDR is dependent on homology between the sequence used as the repair template and the sequences flanking the DSB. These sequences, however, are not always perfectly matched, as shown in the schematic (specifically, γ and θ flanking sequences show heterology). B) Schematic representation of a 'split' homing gene-drive, here where the gRNA-cassette is located within the homing locus (Locus B) as with a global gene-drive design, but the Cas9-cassette is now located at an unlinked, non-homing, location (Locus A). C) The mosquito line crossing scheme for combining the Cas9-transgene and the gRNA-transgene in a split-drive system. In the 'grandparental' F0 generation, the two lines are crossed. In the 'parental' F1 generation, individuals carrying both transgenes (trans-heterozygotes) are isolated and crossed to wild type, producing the F2 generation where transgene fluorescence inheritance ratios are recorded in order to estimate homing rates. In this representation, Locus B (i.e. that containing the gRNA cassette, aka the homing element) is coloured in blue, while Locus A (i.e. that containing the Cas9 cassette) is coloured in green.

Figure 2: Details of the *kmo* locus sequence targeted and genetic background of lines crossed in Assay 1 (allele-by-allele approach). A) Representations of the four *kmo* alleles identified through Sanger sequencing. Allele A was the only allele observed in the 'TPRI' genetic background and showed 100% homology over the sequenced region to the *kmo*-gRNA integrated homing element (0 mismatches over 776bp). Alleles B, C and D were the only alleles observed in the 'CA' genetic background and showed varying levels of heterology over the sequenced region to the *kmo*-gRNA integrated homing element. B) Schematic showing how introgression of the two genetic backgrounds allowed investigation of the influence of target allele identity on homing rates. In brief, in the F0 generation, individuals carrying the *Vasa*-Cas9 transgene (which resides in the CA genetic background – represented in orange) were first crossed to wild-type individuals possessing a TPRI genetic background (represented in blue). *Vasa*-Cas9 heterozygotes, now in a hybrid CA/TPRI genetic background were then crossed to individuals from the *kmo*-gRNA line (which possesses a TPRI genetic background) (F1 generation). This produced the F2 generation where the *kmo-gRNA* cassette was present opposite all four identified *kmo* alleles, in proportion to their

frequencies in the F1 *Vasa*-Cas9 heterozygotes. This allowed the homing reaction to take place into four separate target alleles, at the same locus, in a controlled genetic background (shown by curved arrows). In this generation, trans-heterozygotes were crossed to individuals from a *kmo* -/- strain (CA genetic background). This produced the F3 generation where, for each egg raft, fluorescence ratios and eye phenotypes were recorded and the target *kmo* allele present In the F2 trans-heterozygote which produced that egg raft identified through PCR.

Figure 3: Results from Assay 1 homing crosses: A) 'Cutting rates' estimated for the four identified *kmo* target alleles in the F2 *Vasa*-Cas9/*kmo*-gRNA trans-heterozygotes. Here cutting was calculated as the sum of NHEJ (signified by white-eye phenotype individuals) and HDR (derived from *kmo*-gRNA inheritance >50%) events observed in the F3 progeny. B) 'Homing rates' estimated for the four identified *kmo* target alleles in the F2 *Vasa*-Cas9/*kmo*-gRNA trans-heterozygotes. Here homing rate was calculated as the % of wild-type alleles converted to bear the homing-drive cassette in the trans-heterozygote germline (derived from level of *kmo*-gRNA inheritance >50%). As such, a homing rate of 50% would represent half the wild-type alleles in the F2 trans-heterozygote germline being converted to homing-element alleles, and an overall inheritance rate of the *kmo*-gRNA transgene of 75%. For both A) and B) Large symbols and error bars (vertical lines) represent mean and 95% confidence intervals calculated by a generalized linear mixed model with a binomial ('logit' link) error distribution, individual points to the right of each estimated mean represent pools of offspring derived from a single female parent. Relative size of the small points is in proportion to the number of individuals recorded for that data point (batch size). C) Raw data values of recorded 'Homing rates' as a percentage of the pooled total of 'Cutting rates' for each of the four identified *kmo* target alleles. Raw data proportions (C) represent unadjusted observations, while model-estimated means (A&B) account for covariates and random effects.

Figure 4: Comparison of estimated homing rates from (Harvey-Samuel et al., 2023) **'Original' (i.e. Harvey-Samuel & Feng, 2023) and Assay 2 (i.e. Generalised approach - this study).** In each case, a full factorial split-drive cross setup was applied (i.e. testing all combinations of grandparental and parental sex on estimated homing rates). Facets separate offspring according to the grandparent and parent from which they inherited the *Vasa*-Cas9 transgene. Large symbols and error bars represent mean and 95% confidence intervals, small points represent raw data. Relative size of the small points is in proportion to the number of individuals recorded for this data point (batch size). Annotated asterisks represent pairwise significant difference tests at (n.s. = non-significant, * <0.05, **<0.01, ***<0.001).

Title: Sequence mismatch between gene-drive and target-site flanking regions significantly impairs homing efficiency in *Culex quinquefasciatus*.

Authors

Harvey-Samuel, T.[1,2], Kaur, R.[2], Leftwich, P.T.[3], Feng, X.[4], Gantz, V.[5], Alphey, L.[6,7]

[1] School of Life Sciences, Keele University, Keele, Staffordshire, ST5 5GB, UK

[2] Arthropod Genetics, The Pirbright Institute, Woking, UK, GU24 0NF, UK

[3] School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK

[4] Key Laboratory of Biology of Crop Pathogens and Insects of Zhejiang Province, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, 310027, China

[5] Section of Cell and Developmental Biology, University of California San Diego, San Diego, California, 92093, USA

[6] Department of Biology, University of York, Wentworth Way, York, YO10 5DD, UK

[7] York Biomedical Research Institute, University of York, Heslington, York, YO10 5DD, UK

Corresponding author: Luke Alphey: luke.alphey@york.ac.uk

Keywords: Homology Directed Repair, gene-drive, Homing Endonuclease Gene, west Nile virus, avian malaria

Running Title: Target mismatch impairs homing-drive efficiency

Abstract

CRISPR/Cas9-based homing gene-drives (homing-drives) hold enormous potential as control tools for mosquito disease-vectors. These genomically-encoded technologies spread themselves through target populations by creating double-stranded DNA breaks on homologous chromosomes, into which the homing-drives are copied ('homed'). Homing is dependent on sequence homology between the genomic regions flanking the transgene insertion and the break site. Homing efficiency (i.e. copying rate) substantially impacts the power of these systems: less efficient homing-drives spread slower, have fewer applications and are more resistance-prone. Understanding what influences homing-drive efficiency is therefore vital to the successful use of these technologies. Here we report a novel mechanism by which a homing-drive's efficiency can be significantly impaired by natural sequence variation within a population into which it is spreading. Using a *kmo*-targeting 'split' homing-drive in the West Nile virus mosquito *Culex quinquefasciatus*, we found that target-site heterology (sequence mismatch between the genomic regions flanking the target cut-site and the homing-drive transgene) of less than 10% reduced homing efficiency by up to 54%. While substantial research effort has been dedicated to increasing homing-drive efficiency through optimisation of within-construct components, our results highlight that the real-world efficacy of these systems may in part depend on variation beyond these controllable factors.

Article Summary

Homing-based gene-drives are transformative pest control tools. They work by spreading through the gene-pool of a target-pest population, affecting its control. This spreading is achieved by 'homing': where the gene-drive copies itself from one homologous chromosome to another. Homing requires DNA sequences either side (i.e. flanking) the gene-drive to match those of the site it is copying into. Using a homing gene-drive in the mosquito *Culex quinquefasciatus* we assess how sequence mismatches between these flanking regions affect homing efficiency. We find even small levels of mismatch can significantly reduce homing and discuss this in the context of real-world gene-drive deployment.

Introduction

CRISPR/Cas9 'homing' gene-drives (henceforth homing-drives) are genomically encoded technologies which hold immense promise for controlling intractable pests, with notable development in disease-vectoring mosquitoes(Alphey, 2014; Anderson et al., 2024; Gantz et al., 2015; Hammond et al., 2016; Harvey-Samuel et al., 2023; Li et al., 2020; Sinkins & Gould, 2006). A homing-drive increases its allele frequency from one

generation to the next by copying ('homing') itself from one homologous chromosome to another in germline cells (Burt, 2003). For pest management, a homing-drive can be designed to simultaneously spread traits beneficial for control e.g. pathogen refractoriness or a genetic load for population suppression/eradication (Champer et al., 2016). The efficiency of the homing reaction substantially influences the population-level efficacy of a homing-drive: higher homing rates enable faster spread and higher tolerance of drive-associated fitness costs (Deredec et al., 2008).

At a molecular level, homing occurs when Cas9 expressed by a homing-drive transgene creates a double-stranded DNA break (henceforth 'cut-site') at a specific sequence on the target chromosome, into which the homing-drive is homed (Esvelt et al., 2014) (Fig 1.). Homing is mediated by the homology-directed-repair (HDR) pathway, requiring the genomic sequences flanking the homing-drive transgene insertion-site and the target allele cut-site to show homology. While substantial effort has been devoted to exploring how 'within-drive' components (e.g. regulatory elements for expressing Cas9 (Anderson et al., 2023; Hammond et al., 2021)) affect homing rates, comparatively little attention has been paid to the role of these genomic flanking regions in influencing homing efficiency. This is surprising as the impact of even minor sequence divergence between loci in substantially reducing homologous recombination rates is well established (Elliott et al., 1998). In the real-world, homing-drives will be required to function in genetically diverse 'wild' populations, which will likely also include pre-existing heterogeneity around the cut-site. Whether sequence mismatch (henceforth 'heterology') between a homing-drive's flanking genomic regions and those flanking its target allele cut-site influences homing rates is thus of considerable importance.

We explored this question using our homing-drive system in the West-Nile-virus vector mosquito *Culex quinquefasciatus* (Harvey-Samuel et al., 2023). This system consists of a 'split-drive' design where the two homing-drive components (germline-expressing Cas9 and gRNA-expressing cassettes) are integrated at independent loci - one of which (here the gRNA-cassette) is in the target allele - and maintained in separate lines (Esvelt et al., 2014). The two lines can then be crossed together to initiate the homing reaction in trans-heterozygotes. By using the *kynurinine 3-monooxygenase* (*kmo*) gene as our target allele (where homozygous null mutations give an easily visualized 'white-eye' phenotype (Purusothaman et al., 2021)) and conducting all comparison crosses in a 'controlled' genetic background, we were able to unambiguously distinguish Cas9-mediated cutting and homing, and attribute variation in these to individual *kmo* target alleles which displayed varying levels of heterology. This experimental design provided unparalleled power to assess this fundamentally important question.

Materials and methods

Ethics

Work followed procedures/protocols approved by The Pirbright Institute Biological Agents & Genetic Modification Safety Committee. All homing assays were conducted at The Pirbright Institute IS4L arthropod containment facility under the necessary safety regulations for gene-drive research.

Mosquito lines used, rearing and maintenance.

The *kmo*-gRNA line was generated previously by integrating a *kmo*-gRNA expression cassette via CRISPR/Cas9 HDR into the *kmo* gene in the TPRI (Tropical Pesticides Research Institute) genetic background. The *kmo* -/- line was generated previously by CRISPR/Cas9 knockout of the *kmo* gene in the CA (California) genetic background. The *Vasa*-Cas9 line was generated previously by integrating a Cas9 ORF under the transcriptional control of *Cx. quinquefasciatus Vasa* regulatory elements into the *Cardinal* gene in the CA genetic background via CRISPR/Cas9 HDR. The Wild-type (WT) line is the unmodified TPRI line. Rearing/maintenance followed procedures previously reported.

Mosquito experiments

Assay 1: Allele-by-allele approach

Homozygous *Vasa*-Cas9 females were first pool-crossed to WT males, producing heterozygous *Vasa*-Cas9 F1 progeny in a 50:50 TPRI/CA genetic background. Female F1 were then pool-crossed to heterozygous *kmo*-gRNA males to give the F2 generation. Trans-heterozygous *Vasa*-Cas9, *kmo*-gRNA female F2 progeny were then pool-crossed to *kmo*-/- males and F3 egg-rafts produced individually isolated, with each egg-raft being from an individual female trans-heterozygote. L3-4 stage progeny from each egg-raft were scored for presence of transgenes/eye-pigmentation (white-eyes), as previously demonstrated.

Genotyping of target allele (Assay 1 only)

After scoring, gDNA was extracted from a single WT or *Vasa*-Cas9 individual from each F3 egg-raft (Machery-Nagel Nucleospin Tissue kit, Düren, Germany). This was used to PCR-genotype the maternally-contributed *kmo* allele (i.e. the allele into which the *kmo*-gRNA had attempted to home in the previous generation). The paternally-contributed *kmo* allele was excluded by designing the reverse primer (PL506) to lie across a region deleted during the generation of the *kmo* -/- line. The primers were designed to bind to all 4 *kmo* alleles identified in a preliminary analysis of the strains utilized (A,B,C,D). PCR conditions as follows 98C-1min, (98C-30s, 72C-15s, 72C-1min) x 35, 72C-2min. 200ng of gDNA used per 50ul Q5 PCR Reaction (New England Biolabs). Primers = F(PL303): CCAACATTCACCTTCACTTCAACCACAAGC and R(PL506):

GTGAGCGTCCTCCCACCGAG. Amplicons extended c.320/540bp upstream/downstream of the cut site – sufficient distance to distinguish the 4 alleles. Reactions were run on a 1% agarose gel and bands extracted using the Machery-nagel Nucleospin Gel and PCR clean-up kit. Purified bands were Sanger sequenced using the forward primer PL303 and sequenced amplicons aligned to the TPRI *kmo* sequence.

Calculation of homing and cutting rates (Assay 1 only)

Homing rates are defined as the percentage of homologous wild-type chromosomes converted to carry the *kmo*-gRNA transgene (i.e. the percentage of WT alleles which have been affected by homing). It was estimated for each egg-raft as (200 x (no. *kmo*-gRNA/total progeny – 0.5). Cutting rate is defined as the percentage of homologous wild-type chromosomes observed as having been cut by Cas9. This includes DSBs that were repaired via HDR (homing), NHEJ (or other error-prone mechanisms) but excludes repairs which resulted in no observable change to the WT allele (perfect repair). It was estimated for each egg-raft as the previously estimated homing rate + (200 x (no. non-*kmo*-gRNA progeny which showed white-eyes/total progeny – 0.5). That is, the HDR + error-prone pathway rates.

Assay 2: Generalising findings

Homozygous *Vasa*-Cas9 females were first pool-crossed to WT males, producing heterozygous *Vasa*-Cas9 F1 progeny in a 50:50 TPRI/CA genetic background. This was repeated twice to give F3 *Vasa*-Cas9 progeny in a predominantly TPRI genetic background (c. 87.5% TPRI alleles). Male and female *Vasa*-Cas9 heterozygote F3s were pool-crossed to female and male *kmo*-gRNA line heterozygotes, respectively, giving two cohorts in the F4 generation (one from each of the two F3 crosses). Trans-heterozygous male and female F4 progeny from each of these two crosses were pool-crossed to female and male *kmo* -/- individuals, respectively (4 crosses total), and the egg rafts produced individually isolated and scored as for Assay 1.

Statistical analysis

Homing and cutting rates were both analysed with generalized linear mixed models with a binomial (logit) error distribution. Fixed effects for Assay 1 were the genotyped maternally contributed *kmo* allele with individual female egg-batch as a random effect. Minor overdispersion in the cutting analysis models meant all 95% confidence intervals were checked by semi-parametric bootstrapping (1000 iterations). Results are presented with model-based standard errors and 95% confidence intervals; bootstrap validation confirmed these estimates were robust to minor overdispersion and both summaries are presented in supplementary materials (supplementary tables 1 & 2). For Assay 2 fixed effects included the sex of the Cas9-bearing parent and grandparent as well as whether

the data came from the original (i.e. previously published) experiment or post-introgression (i.e. Assay 2). Individual female egg-batches were also included as a random effect. All models and simulations were constructed in R version 4.3.3 with the package lmerTest, model fits were checked with the package DHARMa and model predictions generated with the package emmeans. Data processing and visualization was performed with the tidyverse packages.

Results and Discussion

Our aim was to investigate the effect of varying the heterology of sequences flanking the cut-site of a target allele on a homing-drive's homing efficiency. We assessed this at two distinct levels, each specifically designed to give insight into the behavior of this potential effect. Firstly, we took an 'allele-by-allele' approach – assessing a homing-drive's efficiency as it was paired against four different *kmo* target alleles in a controlled genetic background. Secondly, we extended this analysis to a broader range of cross schemes and sexes to assess the generality of findings.

1: Allele-by-allele approach

In 'split' homing-drives, one of the expression cassettes (here, that expressing the gRNA) is integrated into the target allele at the precise cut-site, forming the 'homing-element'. This can then be combined with the other element, here encoding Cas9, to provide trans-heterozygotes (aka "double heterozygotes") which have one copy each of Cas9 and homing-element and so can catalyse the homing reaction. Furthermore, and crucially for our purpose, in this set-up, the target allele into which a homing-element attempts to home will always come from the line (and therefore the genetic background) which contributed the Cas9 transgene. Preliminary analysis of our highly inbred wild-type (WT) line (in which the *kmo*-gRNA line was generated and is outcrossed to each generation – TPRI genetic background – see methods) and the *Vasa*-Cas9 line (which was generated separately in the CA genetic background – see methods) showed that these two lines possessed highly divergent *kmo* allele sequences. Those carrying the TPRI genetic background contained only a single *kmo* sequence (allele A) which perfectly matched the sequences flanking the integrated *kmo*-gRNA homing-element, while the *Vasa*-Cas9 line contained three different *kmo* allele sequences (alleles B, C and D – back calculated from subsequent genotyping data to be represented at frequencies of 0.15, 0.65 and 0.2, respectively, in that strain) (See figure 2A), each of which showed substantial sequence variation as compared to allele A, especially in the area 5' of the gRNA target site which harbors intronic sequence (See Supplementary Figure 1). We took advantage of this to design an assay in which a Cas9-bearing individual could contribute a variety of target alleles to a homing cross, some but not all of which would match the homing-element flanking sequences but, critically, all of which would end up in the same mixed genetic background

(see figure 2B). These female trans-heterozygotes (*kmo*-gRNA/+ $^{A,B,C\ or\ D}$, *Vasa*-Cas9/+) were then crossed to male *kmo* -/- homozygotes to allow us to detect both homing (deviation in inheritance rate of homing-element from Mendelian expectation of 50%) and the rate of end-joining (individuals inheriting loss-of-function *kmo* alleles from their mother, generated via Cas9 cutting followed by error-prone repair, will have white eyes but lack the fluorescent marker of the homing-element) in their progeny. This assay showed a highly significant effect of the *kmo* target allele (i.e. A,B,C or D) on homing rate (LRT: χ^2_3 = 16.17, p = 0.001), with a 54% difference in estimated homing rates associated with the highest (A = 0.37 [95% CI; 0.31 – 0.42]) and the lowest (D = 0.17 [0.06 – 0.27]) *kmo* target alleles (Figure 3, Supplementary Figure 2, Supplementary Table 1, Supplementary Table 2). Preliminary sequencing confirmed that the gRNA target sequence was present in all four *kmo* alleles and this was confirmed by our comparison of estimated cutting rates (i.e. estimated homing + observed end-joining) between the *kmo* alleles, which did not significantly differ (LRT: χ^2_3 = 5.27, p = 0.15; Supplementary Figure 2, Supplementary Table 1, Supplementary Table 2). This suggests that the differences in homing we observed between the four target *kmo* alleles were not simply driven by Cas9 being able to cut some alleles more efficiently than others. Instead, taken together, these data strongly suggest that the mechanism underlying the observed reduction in homing is impairment of homology-directed repair through drive-target allele flanking sequence mismatch. Of note, this effect was observable despite the fact that the vast majority of sequence heterology occurred only on one side (i.e. 5') of the target site. While the nature of available *kmo* alleles did not allow us to investigate heterology structure beyond this, results from *Drosophila melanogaster* 'transcomplementing' homing-drives (López Del Amo et al., 2020a) suggest that the presence of heterologous sequences on both sides of the target site may further impair homing efficiency. Further experiments with a wider variety of *kmo* alleles, or employing target genes with less conservation in coding regions, would enable exploration of the effect of more evenly distributed heterology on homing.

2: Generalised approach

In the real world, homing-drive transgenes could be inherited from, and be active in, either sex. Previous work has shown that the efficiencies and dynamics of homing-drives can vary dramatically dependent on direction of inheritance (paternal/maternal) in parents and grandparents (e.g. (Anderson et al., 2024; López Del Amo et al., 2020b; Terradas et al., 2021)). As such, we were interested in investigating the extent to which the observed influence of heterology on homing rates was maintained when a wider variety of cross combinations were considered. To achieve this, we first introgressed the *Vasa*-Cas9 transgene into the TPRI genetic background by outcrossing *Vasa*-Cas9 individuals to our TPRI wild-type line for three consecutive generations. This TPRI introgressed *Vasa*-Cas9 line (henceforth *Vasa*-Cas9^{i}) was then crossed to the *kmo*-gRNA line in a full factorial homing-drive cross (i.e. where the *Vasa*-Cas9 and *kmo*-gRNA transgenes were inherited

from both grandparental (F0) sexes and the trans-heterozygous (F1) individuals were of either sex) and results compared to our previously published homing analysis of these two transgenes - in that case where the *Vasa*-Cas9 transgene was in a pure CA genetic background (Harvey-Samuel et al., 2023). We found that in all like-for-like comparisons the homing rates in this study were higher than in our previous study (aggregate difference 57.8% - 69.3%), only in the male-male crosses was this increase not statistically significant (Figure 4, Supplementary Table 3). This significant interaction between experiment and sex could possibly be indicative of a difference in response to flanking-site heterology between sexes, in the context of homing drives. That is, when faced with high heterology, males showed a higher level of homing than females – an effect which largely disappeared when heterology was decreased (i.e. females increased their homing rates to 'male' levels when drive and target-site flanking regions increased in homology). A possible biological explanation for this observation could be a difference in the recognition requirements for engaging HDR in the male and female germlines, however further work would be required to explore this hypothesis. Overall, these results further support our conclusion that target-site/homing-element flanking-sequence heterology can significantly affect homing efficiency.

To date, only one other study has directly explored the influence of significant (i.e. >1%) flanking-sequence heterology on homing-drive performance (in *Anopheles gambiae*) (Pescod et al., 2023). In contrast to the results presented here, that study observed no significant influence of target-site heterology on homing-efficacy, despite similar levels of sequence divergence (5.3-6.6% over c.690bp). What might explain the differences between this finding and ours? We propose three non-exclusive possibilities.

The first involves the genetic backgrounds in which these crosses were conducted. Pescod *et al*. (Pescod et al., 2023)crossed their homing-drive directly into different geographically distinct *An. gambiae* strains, each of which contained divergent target alleles. As such, each of the different homing-drive/target allele assessments occurred in a distinct genetic background. As it has been shown that (as with 'natural' homologous recombination(Mansai et al., 2011)) many loci unlinked to a homing-drive may exert significant influence over homing efficiencies (Champer et al., 2019), this experimental design may have introduced additional variation, potentially obscuring the effects as observed here. We were able to control for these potential epistatic effects by crossing, and therefore partially introgressing, our *Vasa*-Cas9 line into the TPRI background prior to homing assays, ensuring each of our *kmo* target allele comparisons occurred within (on average) the same, mixed CA/TPRI genetic background.

Secondly, there are potentially species-specific explanations for these contrasting results. Homing-drive efficiencies in *Anopheline* mosquitoes are uniquely high (often approaching or reaching 100%) (e.g. (Carballar-Lejarazú et al., 2023)). The precise reason for this is unknown but could include less stringent requirements on engaging the HDR pathway

e.g. higher tolerance to flanking-sequence heterology, or smaller flanking sequences required to initiate homing. Indeed, analysis of homing conversion tract lengths in *An. gambiae* (i.e the degree to which the flanking regions of a target allele are converted to carry heterologous sequences contained within the homing-drive flanking regions) showed that (beyond the homing-drive cassette itself) only small sequences are transferred onto the target chromosome (>80% of conversion tracts <50bp) (Pescod et al., 2024). This is substantially shorter than reported for studies considering 'natural' homologous recombination in other species (typically c. 200bp (Mansai et al., 2011)) and may signify shorter required sequences for RAD51-dependent complementary strand recognition/invasion in *Anophelines*. In contrast, conversion tract lengths in *Aedes aegypti,* another Culicine mosquito, following CRISPR-Cas9 plasmid-based knock-in (also dependent on HDR, though likely occurring in somatic rather than germline cells) were substantially longer (mean of 160bp) (Ang et al., 2022). Interestingly, and in concurrence with our results, that study in *Ae. aegypti* also identified significant negative effects of plasmid/target-site heterology on the efficiency of knock-in. Taken together, these results could suggest that the HDR-pathway in *An. gambiae* is simply less affected by sequence heterology than in other species, including Culicine mosquitoes.

Finally, the near-100% homing rates displayed by the drives tested in Pescod *et al* may have made it difficult to observe anything other than extremely large changes (i.e. in this case reductions) in homing, potentially underestimating or missing 'subtler' modifiers of homing-efficiency – e.g. those caused by target alleles less amenable to homing. This is because in a homing-drive, inheritance-bias (i.e. the level of homing) is described by odds (p/(1-p)) – where p = wild-type allele conversion probability - and the effect of a change in odds on the change in homing rate differs substantially depending on the initial level of homing ((Deredec et al., 2008), concept visualised in Supplementary Figure 3). In essence, for homing-drives which display very high homing-rates, even large changes (e.g. reductions) in odds *(i.e. reductions in the probability of homing)* will have only small effects on the observed levels of homing. This effect is much less pronounced in less efficient homing-drives (i.e. where homing rates are closer to 0), like that demonstrated here. The diminishing return on changes in odds means that, near 100% efficiency, subtle effects on homing may go unnoticed, complicating efforts to accurately evaluate or identify modifiers of highly effective homing-drives.

In conclusion, our results demonstrate a novel factor affecting homing-drive efficiency. We show that heterology between a homing-drive's flanking regions and those sequences flanking the target allele cut-site can significantly reduce that drive's performance. A critical question is to what extent this effect could impact homing-drive performance in the real-world. In most cases it should be assumed that homing-drives could encounter significant levels of sequence variation upon release, at least equal to the variation observed within our inbred laboratory strains, particularly in continental areas where most

gene-drive deployments will likely occur. A caveat here is that the nature and distribution of this heterology will itself vary and not necessarily be represented by that we observed here. In our system we were only able to contrast and compare those pre-existing 'natural' *kmo* alleles present. Here, even the most divergent *kmo* allele tested was still able to be targeted for Cas9 cleavage and successfully converted to carry the homing-drive, albeit at a significantly reduced rate. If the percentage reduction in homing efficiency observed here were to translate to released strains, however, this could indeed substantially impede their spread through a target population (Supplementary Figure 4), with the magnitude of this effect of course dependent on the identity and frequency of heterogenous target alleles in the wild population. A potentially serious and related issue is the generation of novel drive-resistant alleles: in our system, while sequence heterology impaired the gene-drive's ability to home, it did not impair its ability to cut. As such, a drive introduced into a wild population which shows substantial heterology may produce drive-resistance alleles via end-joining pathways at a substantially higher rate than predicted under lab conditions. While naturally high homing rates may mitigate this effect, for the vast majority of species which have been tested so far, reported homing rates are modest (Grunwald et al., 2019; Harvey-Samuel et al., 2023; Li et al., 2020; Meccariello et al., 2024; Yadav et al., 2023) and thus they may be, to a degree, affected by this mechanism. Prior knowledge of these effects for a given species, homing-drive locus and target population may allow future release programs to preempt adverse consequences on project outcomes, e.g. by increasing release rates accordingly to counteract predicted lower drive efficiency, or by introducing the homing-drive construct into a variety of alleles present within a particular targeted population, ahead of releases. Releasing multiple alleles of a gene-drive insertion into a population would also open the opportunity to employ population-level multiplexing strategies, with each of the released drive alleles working to provide resistance management for the others(Edgington et al., 2020). Additionally, our findings highlight the utility of exploring factors which affect homing-drive dynamics and performance in species/model-systems which do not display near-perfect performance. While 'high-performance' homing-drives may be the first in line for field-release, it is precisely their extreme efficiency which may make them less useful for investigating some of the underlying factors which affect the drive mechanism itself.

Our results are particularly significant as gene-drives edge ever closer to field application. They suggest that, at least in some cases, homing-drive efficiencies estimated in genetically homogenous lab populations may be unrepresentative of those which would occur in the real-world. While lab-field discrepancies are well known for other parameters relevant to genetic biocontrol, e.g. mating success, longevity or other factors relating to fitness, this is the first study we are aware of which identifies such a disparity related to the homing mechanism itself. This information will be critical in helping to set realistic expectations of these potentially game-changing systems, if and when they enter the field-testing phase.

Data Availability

The authors affirm that all data necessary for confirming the conclusions of the article are present within the article, additional information files, figures, and tables. Specifically, all data gathered during and analyzed as part of the two experiments (i.e. inheritance rates of different genotypes, estimation of cutting rates, list of sequenced allele-identities) is available in the two additional information files. Strains and plasmids are available upon request.

Author contributions

THS, XF, VG and LA conceived the project. RK, THS and PL contributed to the design of the experiments. THS, RK and PL performed the experiments and contributed to the collection and analysis of data. THS and PL wrote the manuscript. All authors edited the manuscript.

Study Funding

This work was supported by strategic funding from the UK Biotechnology and Biological Sciences Research Council (BBSRC) to The Pirbright Institute (BBS/E/I/00007033, BBS/E/I/00007038, and BBS/E/I/00007039).

Conflict of Interest

VG is a founder of and has equity interests in Symbol, Inc. and Agragene, Inc., companies that may potentially benefit from the research results described in this manuscript. VG also serves on both the company's Scientific Advisory Board and the Board of Directors of Synbal, Inc. The terms of this arrangement have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies. LA is an advisor to, and has financial or equity interest in, Synvect Inc. and Biocentis Ltd., companies operating in the area of genetic control of pest insects. The other authors declare that they have no competing interests

References

Alphey, L. (2014). Genetic control of mosquitoes. *Annual Review of Entomology*, *59*, 205–224. https://doi.org/10.1146/annurev-ento-011613-162002

Anderson, M. A. E., Gonzalez, E., Ang, J. X. D., Shackleford, L., Nevard, K., Verkuijl, S. A. N., Edgington, M. P., Harvey-Samuel, T., & Alphey, L. (2023). Closing the gap to effective gene drive in Aedes aegypti by exploiting germline regulatory elements. *Nature Communications*, *14*(1), 338. https://doi.org/10.1038/s41467-023-36029-7

Anderson, M. A. E., Gonzalez, E., Edgington, M. P., Ang, J. X. D., Purusothaman, D.-K., Shackleford, L., Nevard, K., Verkuijl, S. A. N., Harvey-Samuel, T., Leftwich, P. T., Esvelt, K., & Alphey, L. (2024). A multiplexed, confinable CRISPR/Cas9 gene drive can propagate in caged Aedes aegypti populations. *Nature Communications*, *15*(1), 729. https://doi.org/10.1038/s41467-024-44956-2

Ang, J. X. De, Nevard, K., Ireland, R., Purusothaman, D.-K., Verkuijl, S. A. N., Shackleford, L., Gonzalez, E., Anderson, M. A. E., & Alphey, L. (2022). Considerations for homology-based DNA repair in mosquitoes: Impact of sequence heterology and donor template source. *PLOS Genetics*, *18*(2), e1010060. https://doi.org/10.1371/journal.pgen.1010060

Burt, A. (2003). Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1518), 921–928. https://doi.org/10.1098/rspb.2002.2319

Carballar-Lejarazú, R., Dong, Y., Pham, T. B., Tushar, T., Corder, R. M., Mondal, A., Sánchez C., H. M., Lee, H.-F., Marshall, J. M., Dimopoulos, G., & James, A. A. (2023). Dual effector population modification gene-drive strains of the African malaria mosquitoes, *Anopheles gambiae* and *Anopheles coluzzii*. *Proceedings of the National Academy of Sciences*, *120*(29). https://doi.org/10.1073/pnas.2221118120

Champer, J., Buchman, A., & Akbari, O. S. (2016). Cheating evolution: engineering gene drives to manipulate the fate of wild populations. *Nature Reviews Genetics*, *17*(3), 146–159. https://doi.org/10.1038/nrg.2015.34

Champer, J., Wen, Z., Luthra, A., Reeves, R., Chung, J., Liu, C., Lee, Y. L., Liu, J., Yang, E., Messer, P. W., & Clark, A. G. (2019). CRISPR Gene Drive Efficiency and Resistance Rate Is Highly Heritable with No Common Genetic Loci of Large Effect. *Genetics*, *212*(1), 333–341. https://doi.org/10.1534/genetics.119.302037

Deredec, A., Burt, A., & Godfray, H. C. J. (2008). The Population Genetics of Using Homing Endonuclease Genes in Vector and Pest Management. *Genetics*, *179*(4), 2013–2026. https://doi.org/10.1534/genetics.108.089037

Edgington, M. P., Harvey-Samuel, T., & Alphey, L. (2020). Population-level multiplexing: A promising strategy to manage the evolution of resistance against gene drives targeting a neutral locus. *Evolutionary Applications*, *13*(8), 1939–1948. https://doi.org/10.1111/eva.12945

Elliott, B., Richardson, C., Winderbaum, J., Nickoloff, J. A., & Jasin, M. (1998). Gene Conversion Tracts from Double-Strand Break Repair in Mammalian Cells. *Molecular and Cellular Biology*, *18*(1), 93–101. https://doi.org/10.1128/MCB.18.1.93

Esvelt, K. M., Smidler, A. L., Catteruccia, F., & Church, G. M. (2014). Concerning RNA-guided gene drives for the alteration of wild populations. *ELife*, *3*. https://doi.org/10.7554/eLife.03401

Gantz, V. M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V. M., Bier, E., & James, A. A. (2015). Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*. *Proceedings of the National Academy of Sciences*, *112*(49). https://doi.org/10.1073/pnas.1521077112

Grunwald, H. A., Gantz, V. M., Poplawski, G., Xu, X.-R. S., Bier, E., & Cooper, K. L. (2019). Super-Mendelian inheritance mediated by CRISPR–Cas9 in the female mouse germline. *Nature*, *566*(7742), 105–109. https://doi.org/10.1038/s41586-019-0875-2

Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., Gribble, M., Baker, D., Marois, E., Russell, S., Burt, A., Windbichler, N., Crisanti, A., & Nolan, T. (2016). A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nature Biotechnology*, *34*(1), 78–83. https://doi.org/10.1038/nbt.3439

Hammond, A., Karlsson, X., Morianou, I., Kyrou, K., Beaghton, A., Gribble, M., Kranjc, N., Galizi, R., Burt, A., Crisanti, A., & Nolan, T. (2021). Regulating the expression of gene drives is key to increasing their invasive potential and the mitigation of resistance. *PLOS Genetics*, *17*(1), e1009321. https://doi.org/10.1371/journal.pgen.1009321

Harvey-Samuel, T., Feng, X., Okamoto, E. M., Purusothaman, D.-K., Leftwich, P. T., Alphey, L., & Gantz, V. M. (2023). CRISPR-based gene drives generate super-Mendelian inheritance in the disease vector Culex quinquefasciatus. *Nature Communications*, *14*(1), 7561. https://doi.org/10.1038/s41467-023-41834-1

Li, M., Yang, T., Kandul, N. P., Bui, M., Gamez, S., Raban, R., Bennett, J., Sánchez C, H. M., Lanzaro, G. C., Schmidt, H., Lee, Y., Marshall, J. M., & Akbari, O. S. (2020).

Development of a confinable gene drive system in the human disease vector Aedes aegypti. *ELife*, *9*. https://doi.org/10.7554/eLife.51701

López Del Amo, V., Bishop, A. L., Sánchez C., H. M., Bennett, J. B., Feng, X., Marshall, J. M., Bier, E., & Gantz, V. M. (2020a). A transcomplementing gene drive provides a flexible platform for laboratory investigation and potential field deployment. *Nature Communications*, *11*(1), 352. https://doi.org/10.1038/s41467-019-13977-7

López Del Amo, V., Bishop, A. L., Sánchez C., H. M., Bennett, J. B., Feng, X., Marshall, J. M., Bier, E., & Gantz, V. M. (2020b). A transcomplementing gene drive provides a flexible platform for laboratory investigation and potential field deployment. *Nature Communications*, *11*(1), 352. https://doi.org/10.1038/s41467-019-13977-7

Mansai, S. P., Kado, T., & Innan, H. (2011). The Rate and Tract Length of Gene Conversion between Duplicated Genes. *Genes*, *2*(2), 313–331. https://doi.org/10.3390/genes2020313

Meccariello, A., Hou, S., Davydova, S., Fawcett, J. D., Siddall, A., Leftwich, P. T., Krsticevic, F., Papathanos, P. A., & Windbichler, N. (2024). Gene drive and genetic sex conversion in the global agricultural pest Ceratitis capitata. *Nature Communications*, *15*(1), 372. https://doi.org/10.1038/s41467-023-44399-1

Pescod, P., Bevivino, G., Anthousi, A., Shelton, R., Shepherd, J., Lombardo, F., & Nolan, T. (2023). Measuring the Impact of Genetic Heterogeneity and Chromosomal Inversions on the Efficacy of CRISPR-Cas9 Gene Drives in Different Strains of *Anopheles gambiae*. *The CRISPR Journal*, *6*(5), 419–429. https://doi.org/10.1089/crispr.2023.0029

Pescod, P., Bevivino, G., Anthousi, A., Shepherd, J., Shelton, R., Lombardo, F., & Nolan, T. (2024). Homing gene drives can transfer rapidly between Anopheles gambiae strains with minimal carryover of flanking sequences. *Nature Communications*, *15*(1), 6846. https://doi.org/10.1038/s41467-024-51225-9

Purusothaman, D.-K., Shackleford, L., Anderson, M. A. E., Harvey-Samuel, T., & Alphey, L. (2021). CRISPR/Cas-9 mediated knock-in by homology dependent repair in the West Nile Virus vector Culex quinquefasciatus Say. *Scientific Reports*, *11*(1), 14964. https://doi.org/10.1038/s41598-021-94065-z

Sinkins, S. P., & Gould, F. (2006). Gene drive systems for insect disease vectors. *Nature Reviews Genetics*, *7*(6), 427–435. https://doi.org/10.1038/nrg1870

Terradas, G., Buchman, A. B., Bennett, J. B., Shriner, I., Marshall, J. M., Akbari, O. S., & Bier, E. (2021). Inherently confinable split-drive systems in Drosophila. *Nature Communications*, *12*(1), 1480. https://doi.org/10.1038/s41467-021-21771-7

Yadav, A. K., Butler, C., Yamamoto, A., Patil, A. A., Lloyd, A. L., & Scott, M. J. (2023). CRISPR/Cas9-based split homing gene drive targeting *doublesex* for population suppression of the global fruit pest *Drosophila suzukii*. *Proceedings of the National Academy of Sciences*, *120*(25). https://doi.org/10.1073/pnas.2301525120

Figure Legends

Figure 1: Schematic representations of CRISPR-Cas9 based homing gene-drive technologies. A) A 'global' homing gene-drive in which all the molecular components required to initiate the homing reaction are linked and integrated at a single locus in the genome which conforms to the target site of the integrated gRNA. Expression of the Cas9 enzyme and gRNA thus mediates the creation of a double-stranded break (DSB) at the insertion locus on the homologous chromosome. This DSB can be repaired through the Homology-Dependent Repair (HDR) pathway which results in the copying of the gene-drive transgene onto the homologous chromosome. If this process occurs in germline cells, it will result in super-mendelian inheritance of the gene-drive. The efficiency of HDR is dependent on homology between the sequence used as the repair template and the sequences flanking the DSB. These sequences, however, are not always perfectly matched, as shown in the schematic (specifically, γ and θ flanking sequences show heterology). B) Schematic representation of a 'split' homing gene-drive, here where the gRNA-cassette is located within the homing locus (Locus B) as with a global gene-drive design, but the Cas9-cassette is now located at an unlinked, non-homing, location (Locus A). C) The mosquito line crossing scheme for combining the Cas9-transgene and the gRNA-transgene in a split-drive system. In the 'grandparental' F0 generation, the two lines are crossed. In the 'parental' F1 generation, individuals carrying both transgenes (trans-heterozygotes) are isolated and crossed to wild type, producing the F2 generation where transgene fluorescence inheritance ratios are recorded in order to estimate homing rates. In this representation, Locus B (i.e. that containing the gRNA cassette, aka the homing element) is coloured in blue, while Locus A (i.e. that containing the Cas9 cassette) is coloured in green.

Figure 2: Details of the *kmo* locus sequence targeted and genetic background of lines crossed in Assay 1 (allele-by-allele approach). A) Representations of the four *kmo* alleles identified through Sanger sequencing. Allele A was the only allele observed in the 'TPRI' genetic background and showed 100% homology over the sequenced region to the *kmo*-gRNA integrated homing element (0 mismatches over 776bp). Alleles B, C and D were the only alleles observed in the 'CA' genetic background and showed varying levels of heterology over the sequenced region to the *kmo*-gRNA integrated homing element. B) Schematic showing how introgression of the two genetic backgrounds allowed investigation of the influence of target allele identity on homing rates. In brief, in the F0 generation, individuals carrying the *Vasa*-Cas9 transgene (which resides in the CA genetic background – represented in orange) were first crossed to wild-type individuals possessing a TPRI genetic background (represented in blue). *Vasa*-Cas9 heterozygotes, now in a hybrid CA/TPRI genetic background were then crossed to individuals from the *kmo*-gRNA line (which possesses a TPRI genetic background) (F1 generation). This produced the F2 generation where the *kmo-gRNA* cassette was present opposite all four identified *kmo* alleles, in proportion to their

frequencies in the F1 *Vasa*-Cas9 heterozygotes. This allowed the homing reaction to take place into four separate target alleles, at the same locus, in a controlled genetic background (shown by curved arrows). In this generation, trans-heterozygotes were crossed to individuals from a *kmo* -/- strain (CA genetic background). This produced the F3 generation where, for each egg raft, fluorescence ratios and eye phenotypes were recorded and the target *kmo* allele present In the F2 trans-heterozygote which produced that egg raft identified through PCR.

Figure 3: Results from Assay 1 homing crosses: A) 'Cutting rates' estimated for the four identified *kmo* target alleles in the F2 *Vasa*-Cas9/*kmo*-gRNA trans-heterozygotes. Here cutting was calculated as the sum of NHEJ (signified by white-eye phenotype individuals) and HDR (derived from *kmo*-gRNA inheritance >50%) events observed in the F3 progeny. B) 'Homing rates' estimated for the four identified *kmo* target alleles in the F2 *Vasa*-Cas9/*kmo*-gRNA trans-heterozygotes. Here homing rate was calculated as the % of wild-type alleles converted to bear the homing-drive cassette in the trans-heterozygote germline (derived from level of *kmo*-gRNA inheritance >50%). As such, a homing rate of 50% would represent half the wild-type alleles in the F2 trans-heterozygote germline being converted to homing-element alleles, and an overall inheritance rate of the *kmo*-gRNA transgene of 75%. For both A) and B) Large symbols and error bars (vertical lines) represent mean and 95% confidence intervals calculated by a generalized linear mixed model with a binomial ('logit' link) error distribution, individual points to the right of each estimated mean represent pools of offspring derived from a single female parent. Relative size of the small points is in proportion to the number of individuals recorded for that data point (batch size). C) Raw data values of recorded 'Homing rates' as a percentage of the pooled total of 'Cutting rates' for each of the four identified *kmo* target alleles. Raw data proportions (C) represent unadjusted observations, while model-estimated means (A&B) account for covariates and random effects.

Figure 4: Comparison of estimated homing rates from (Harvey-Samuel et al., 2023) **'Original' (i.e. Harvey-Samuel & Feng, 2023) and Assay 2 (i.e. Generalised approach - this study).** In each case, a full factorial split-drive cross setup was applied (i.e. testing all combinations of grandparental and parental sex on estimated homing rates). Facets separate offspring according to the grandparent and parent from which they inherited the *Vasa*-Cas9 transgene. Large symbols and error bars represent mean and 95% confidence intervals, small points represent raw data. Relative size of the small points is in proportion to the number of individuals recorded for this data point (batch size). Annotated asterisks represent pairwise significant difference tests at (n.s. = non-significant, * <0.05, **<0.01, ***<0.001).

This data sheet shows the scored larval phenotypes for the two fluorophores and white eyes in the resul

Cas9 inherited from	Grandparent sex	transgenic parent sex	WT
M		M	11
M		M	22
M		M	27
M		M	13
M		M	24
M		M	5
M		M	14
M		M	20
M		M	12
M		M	2
M		F	3
M		F	7
M		F	35
M		F	5
M		F	20
M		F	0
M		F	23
M		F	1
M		F	13
M		F	5
F		F	7
F		F	15
F		F	9
F		F	18
F		F	21
F		F	18
F		F	9
F		F	16
F		F	9
F		F	8
F		M	14
F		M	18
F		M	7
F		M	18
F		M	11
F		M	10
F		M	10
F		M	21
F		M	4

ts generation of the 'Generalised approach' experiment.

Homing scoring

Homing element	Vasa-cas9	Vasa-cas9 + homing element
16	13	16
27	17	25
48	17	37
28	9	12
40	26	39
14	6	12
19	14	20
29	20	42
34	11	33
3	3	6
2	6	2
16	14	15
39	25	43
43	7	36
43	21	65
4	0	1
45	16	56
5	2	4
22	24	23
8	3	6
31	10	34
42	22	44
47	8	31
35	9	33
33	15	32
44	10	45
55	12	46
55	25	64
49	10	43
36	15	23
28	14	38
43	21	46
17	8	17
31	16	28
45	15	51
17	11	23
13	7	9
36	15	35
27	6	17

Analysis

sum 'have homing element'	sum total	sum	sum	%
32	56			
52	91			
85	129			
40	62			
79	129			
26	37			
39	67			
71	111			
67	90			
9	14	500	786	0.636132
4	13			
31	52			
82	142			
79	91			
108	149			
5	5			
101	140			
9	12			
45	82			
14	22	478	708	0.675141
65	82			
86	123			
78	95			
68	95			
65	101			
89	117			
101	122			
119	160			
92	111			
59	82	822	1088	0.755515
66	94			
89	128			
34	49			
59	93			
96	122			
40	61			
22	39			
71	107			
44	54	521	747	0.697456

White eye data

Cas9 Grandparent sex	transgenic parent sex	WT	Vasa-cas9	sum
M	M	6	2	
M	M	1	0	
M	M	6	3	
M	M	3	1	
M	M	11	7	
M	M	2	0	
M	M	2	5	
M	M	5	5	
M	M	3	4	
M	M	0	0	39
M	F	0	1	
M	F	5	4	
M	F	10	8	
M	F	5	3	
M	F	7	9	
M	F	0	0	
M	F	8	5	
M	F	1	1	
M	F	7	13	
M	F	3	1	46
F	F	2	3	
F	F	2	6	
F	F	6	5	
F	F	8	4	
F	F	5	4	
F	F	9	5	
F	F	7	6	
F	F	6	8	
F	F	4	10	
F	F	4	10	53
F	M	3	4	
F	M	3	3	
F	M	3	5	
F	M	8	6	
F	M	5	7	
F	M	5	0	
F	M	1	1	
F	M	8	6	
F	M	1	1	37

Final cutting analysis

sum	As % WT	As%Cas9
27	0.26	0.198529
45	0.410714	0.381356
61	0.407692	0.448529
33	0.327434	0.257813

% A element WE from original experitment due to deposit

MF	0.054347826
FF	0.038793103

overall cutting rate = homing + cutting, controlled for dep

	%Wt alleles homed	%WT alleles cut
MM	28	18.72
MF	36	20.85093168
FF	52	15.68992042
FM	40	19.6460177

:ion

osition white eyes (original experiment)

Total % WT alleles edited
46.72
56.85093168
67.68992042
59.6460177

This datasheet shows the results from the individual by individal approach experiment. It calcultates

Individual no.	Cas9+homing element	Cas9	homing element	WT
45	4	2	0	0
65	13	8	13	8
42	54	30	40	30
5	4	3	1	0
48	9	7	16	6
46	26	8	22	13
1	10	10	8	4
4	14	13	17	13
17	6	2	0	3
5	44	15	35	16
6	23	14	29	8
12	12	12	16	12
8	10	7	12	7
20	8	1	9	1
9	13	9	10	8
22	16	11	19	6
16	5	5	5	6
11	37	20	33	20
23	13	9	14	6
52	28	4	22	7
37	24	11	32	11
47	18	17	12	12
41	26	11	20	7
36	13	13	4	12
17	12	8	12	11
20	34	14	34	11
19	30	15	33	17
39	23	29	36	25
27	31	13	20	15
24	32	26	41	23
13	40	18	34	22
1	30	14	37	11
16	33	21	44	21
18	40	27	30	22
6	22	30	25	16
48	21	13	18	11
3	15	6	8	6
15	21	6	16	6
12	19	13	22	10
29	42	9	25	13
2	37	16	31	15
6	28	8	28	14
9	53	30	36	23
18	5	5	6	2
14	16	13	9	14

	17	11	7	20	8
	27	21	13	26	6
	20	32	23	32	21
	21	39	20	35	20
	23	18	8	19	9
	28	43	28	36	35
	30	42	24	35	26
	37	20	25	22	12
	36	33	13	29	15
	8	53	11	55	18
y		17	3	9	4
x		16	11	8	7
	41	42	27	46	26
	13	35	7	44	8
	33	35	12	26	19
	40	31	30	31	21
	24	17	16	28	10
	31	30	23	31	21
	35	4	9	9	9
	10	31	26	28	20
	7	29	27	36	24
	43	14	8	16	8

homing rates for individual egg rafts (mating events) and connects these with the KMO genotype of

homing rate	Allele type (NA = sequencing failed)
0.666666667	NA
0.619047619	C
0.61038961	C
0.625	C
0.657894737	C
0.695652174	A
0.5625	B
0.543859649	C
0.545454545	C
0.718181818	A
0.702702703	C
0.538461538	A
0.611111111	C
0.894736842	NA
0.575	A
0.673076923	C
0.476190476	A
0.636363636	B
0.642857143	C
0.819672131	NA
0.717948718	A
0.508474576	NA
0.71875	A
0.404761905	A
0.558139535	NA
0.731182796	C
0.663157895	A
0.522123894	B
0.64556962	A
0.598360656	C
0.649122807	D
0.72826087	A
0.647058824	NA
0.588235294	D
0.505376344	C
0.619047619	A
0.657142857	A
0.755102041	A
0.640625	B
0.752808989	A
0.686868687	A
0.717948718	A
0.626760563	D
0.611111111	C
0.480769231	C

	average ho
	A
average	0.667919
SE	0.019102
std dev	0.097402
sample size	26
CI	0.037439

0.673913043	A
0.712121212	A
0.592592593	C
0.649122807	NA
0.685185185	C
0.556338028	D
0.606299213	C
0.53164557	C
0.688888889	C
0.788321168	A
0.787878788	A
0.571428571	A
0.624113475	A
0.840425532	A
0.663043478	A
0.548672566	C
0.633802817	A
0.580952381	D
0.419354839	D
0.561904762	D
0.560344828	B
0.652173913	NA

the target allele in that cross by sequencing

ming

B	C	D
0.584391	0.61007	0.568953
0.023236	0.014586	0.027977
0.051957	0.066842	0.074019
5	21	7
0.045541	0.028588	0.054833