Investigating The Potential of Population Specific Homing Drives in *Lolium Rigidum*

Patrick Gibbs 1083438

Abstract:

- Pests and invasive species pose a growing problem in a changing climate and increasingly globalised society.
- Allele specific homing drives have been proposed as a targeted means of pest control.
- In this paper I report the potential for allele specific homing drives for populations without barriers to gene flow utilizing 63 *Lolium rigidum* populations from South East Australia.
- I found in the absence of barriers to gene flow, genetic heterogeneity negates CRISPR/Cas9 annealing private to populations.
- I also found that Fst is a reliable indicator of the abundance of polymorphic CRISPR/Cas9 annealing sites between populations.
- These findings suggest that on a fine geographic scale in the absence of gene flow, allele specific homing drives do not present a robust solution for targeted pest control.

Introduction:

Pests and invasive species are responsible for substantial environmental, agricultural and health costs. In a changing climate and increasingly globalised society, the proliferation of pests and damage caused by them is predicted to grow (Diagne 2021).

Traditional methods of pest control such as poisons or pesticides are non specific, often causing collateral environmental damage. Moreover, growing xenobiotic resistance is requiring larger applications of pesticides further exacerbating costs (Pimentel and Burgess, 2014).

Gene drives are genetic elements which bias their inheritance above typical mendelian patterns. Gene drives can spread throughout a population, even if they bear a fitness cost (Unkless et al., 2015). Artificially constructed gene drives have been proposed as a means of pest management. Such pest management can be achieved through suppressor drives, which spread a fitness cost to reduce population size, or modification drives which spread a desired element to alter the nature of a population (Rode et al., 2019).

The recent development of the CRISPR/Cas9 homing drive provides an effective artificial gene drive which can be implemented in most species. Typically, homing drives are encoded in an allele such that in heterozygous, they can cleave their homolog. The construct can then convert to homozygous through homologous driven repair (HDR), thus ensuring its inheritance across generations. CRISPR/Cas9 homing gene drives can be placed specifically throughout the genome, and in *Drosophila melanogaster* have been shown to be as high as 95% effective in biasing inheritance (Gantz and Biar, 2015). Thus, CRISPR/Cas9 homing drives provide a promising tool for pest management through the suppression or alteration of natural populations. Indeed, modeling and controlled laboratory experiments have shown CRISPR/Cas9 suppressor drives have the potential to completely eliminate populations (Hammond et al., 2021; Unkless et al., 2015).

While gene drives offer enormous payoff through targeted pest and disease vector control, they have potential to reap catastrophic damage if their spread is unchecked. For example, suppression of a pest population may create a new niche for other invasive species, potentially exacerbating damage. Moreover, gene drives targeted at an invasive species could spread back to native populations harming off-target ecosystems (Rode et al,. 2019). Horizontal gene transfer between species is rare but presents risk to off-target species when introducing gene drives. Famously, the natural transposon P-element recently invaded and spread throughout *Drosophila melanogaster* populations (Clark & Kidwell, 1997).

Generally, the risks associated with unbounded gene drives render their application untenable. However, self limiting gene drives are designed such that their spread and/or duration is controlled, promising safer and specific implementation (Rode et al., 2019). It has been proposed that polymorphisms in a CRISPR/Cas9 gRNA annealing site between populations can allow for homing drives to be targeted specifically to one population (Willis and Burt 2021; Sudweeks et al., 2019). Under this design, alleles prevalent in the target population would be susceptible to a gene drive, while alleles abundant in the off-target populations are resistant. In particular, polymorphism in the protospacer adjacent motif (PAM site) would make an allele completely resistant to a CRISPR/Cas9 gene drive as its conservation is essential for Cas9 cleavage. Polymorphism in the gRNA also reduces cleavage efficiency depending on the number and position of the polymorphisms (Hsu et al., 2013).

Researchers Willis and Burt (2021) have investigated such applications in *Anopheles gambiae* populations across the African continent in the Ag100G dataset. They found loci at which PAM sites were private to certain island populations, meaning only those populations would be vulnerable to specifically targeted homing drives.

However, no previous literature has reported investigation of allele specific homing drives over a small geographic area with samples taken at high density. Use of the broader guide RNA annealing site for CRISPR/Cas9 homing drives, to which polymorphism also presents significant resistance, also remains an unreported research direction. Additionally, species other than *Anopheles gambiae* have not had their population genetics examined for the applicability of allele specific gene drives.

Research Aims

In this research I investigated the applicability of allele-specific gene drives to target individual populations at a subcontinental geographic scale. I utilized genomic sequence data from 63 *Lolium rigidum* populations sampled throughout Southeastern Australia. Such populations have limited barriers to gene flow and thus represent a low band of population specific gene drive applicability. That is, if findings show abundance of population specific targets, it would suggest wide applicability of the technology.

It is predicted that the local geography (subcontinental) of the data is less facilitating to population specific homing drives compared to the continental or island geography reported in past literature (Willis and Burt 2021). To test this hypothesis, each population was individually measured for unique CRISPR/annealing sites. The CRISPR/Cas9 annealing sites that showed high levels of differentiation were recorded, and are indicative of the viability of homing drives specific to given populations.

Following this, the underlying features which drive specificity of homing drives were investigated. In particular, I test the hypothesis that both Fst and geographic distance between any two populations is indicative of the number of polymorphic CRISPR/Cas9 annealing sites.

Finally, I investigate the potential of utilizing multiple polymorphisms in the gRNA annealing site for more robust allele specific homing drive design. This is achieved by computing the proportion of CRISPR/Cas9 annealing sites that contain multiple polymorphisms.

Hypothesis	Tested by Experiment
Compared to prior literature, smaller geographies are less facilitating to population specific homing drives.	1
Fst and geographic distance are reliable indicators of specific homing drive applicability between pairs of populations	2a
Multiple polymorphisms in the gRNA may be used for robust specificity of homing drives.	2b

Materials and Methods

Data collection and representation.

Sequence data of 63 *Lolium ridgum* populations sampled across South Eastern (SE) Australia were provided by The University of Melbourne Adaptive Evolution Lab (Paril 2020, ch. 4). This data set includes samples from both agricultural and wild settings. Leaf tissues from each of these populations were sampled and pooled. DNA was extracted from these pools and illumina sequencing was conducted (Toonen et al., 2013). A 2x restriction digest was applied in the illumina sequencing process, over-representing a subset of the genome in read coverage/depth. For this analysis the data was filtered only including positions to which more than 50 reads were mapped. Enforcing a minimum read depth reduces the probability of the results being an artifact of the sequencing process, rather than reflecting true populations. This process resulted in a genome coverage of 243,665 nucleotide positions, 0.03% of the 2.2 billion nucleotide Lolium rigidum genome. Thus, results of this study pertain to a sample of the genome which is assumed to be representative of the genetic structure pertaining to the full genome.

Sequence reads were aligned and piled up utilizing *samtools mpileup* (Li 2011). The pileup was summarized to a synchronized file (sync file) utilizing *mpileup2sync.jar* from Popoolation2 (Kofler et al. 2011). This resultant sync file was used for the analysis in this project.

Sync files specify the number of reads mapped to each locus of the reference genome, for each population in the data set. We note that additions are not recorded in the sync files. Haplotypic

information is also not recorded in sync files meaning it is not possible to know if closely linked polymorphisms pertain to the same allele or not. Addressing these issues was not practical given the time constraints of this project, thus it was chosen to only consider the SNPs (single nucleotide polymorphism) in CRISPR/Cas9 annealing sites. Not considering indels in the analysis is expected to lead to a minor underestimation of the true divergence in CRISPR annealing sites.

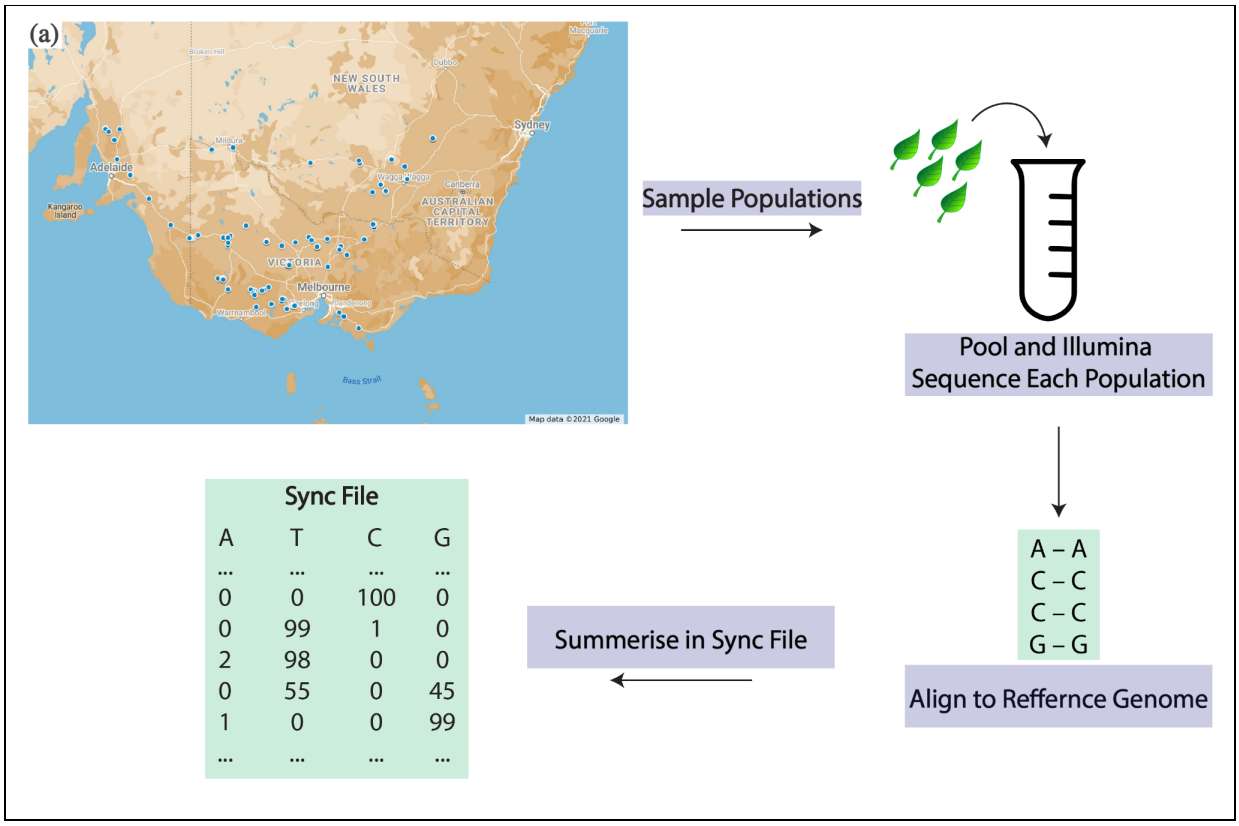

Figure 1: Summary of Data Collection Process. (a) Locations sampled across SE Australia.

Experiment 1: Test the abundance of allele specific homing drives throughout populations of the dataset.

A pipeline was developed to locate CRISPR/Cas9 annealing sites from the reads of sync file and measure their specificity to a chosen target population.

Firstly, a consensus sequence was constructed containing nucleotides with the highest number of reads for each position in the sync file. In the consensus sequences canonical NGG PAM sites were located, indicative of CRISPR/Cas9 annealing sites (Figure 2a). For each of these target sites, each nucleotide in the guide RNA binding site and in the two specific GG nucleotides of the PAM were measured for SNPs between the target population and all off-target populations. The presence of SNPs was measured using an originally formulated summary statistic *D*. *D* represents the average divergence between the target and all off-target populations at one nucleotide position (Figure 2b, 2c).

$$D = \frac{1}{n} \sum_{i=1}^{n} (\hat{p}_k - \hat{p}_i),\ D \in [-1, 1]$$

n is the total number of off-target populations. $\hat{p}_k$ is the allele frequency for a SNP in the target population, $\hat{p}_i$ is the frequency in each off-target population, of the allele measured by $\hat{p}_k$.

As applied here, *D* measures the proportion of individuals in the target population susceptible to a gene drive allele minus the proportion of all other off-target individuals susceptible. The higher *D* the more specific a gene drive is to a targeted population. We note that *D* is directional $\in [-1, 1]$. In the most extreme case where *D* is 1, the susceptible allele is private to the target population meaning all individuals are susceptible and no individuals in the off-target population are susceptible. When D is negative, it means that the target population is less vulnerable to a gene drive compared to off-target populations.

Differentiated positions were queried to annotation files to see if they lay in known genes. Gene ontology is of interest because simply designed gene drives may be targeted at conserved regions such that resistance through mutation may be fatal to the organism.

The exact location of the SNPs in the CRISPR annealing sites were recorded as the effect of mutation on CRISPR target sites varies for different positions. Particularly, conservation of the PAM site is

essential for CRISPR/Cas9 cleavage, meaning polymorphisms in the PAM site are ideal for targeted gene drives (Hsu et al., 2013).

Thus, for every CRISPR annealing site in each population, each nucleotide was assigned a D value, as well as having its position and gene annotation recorded.

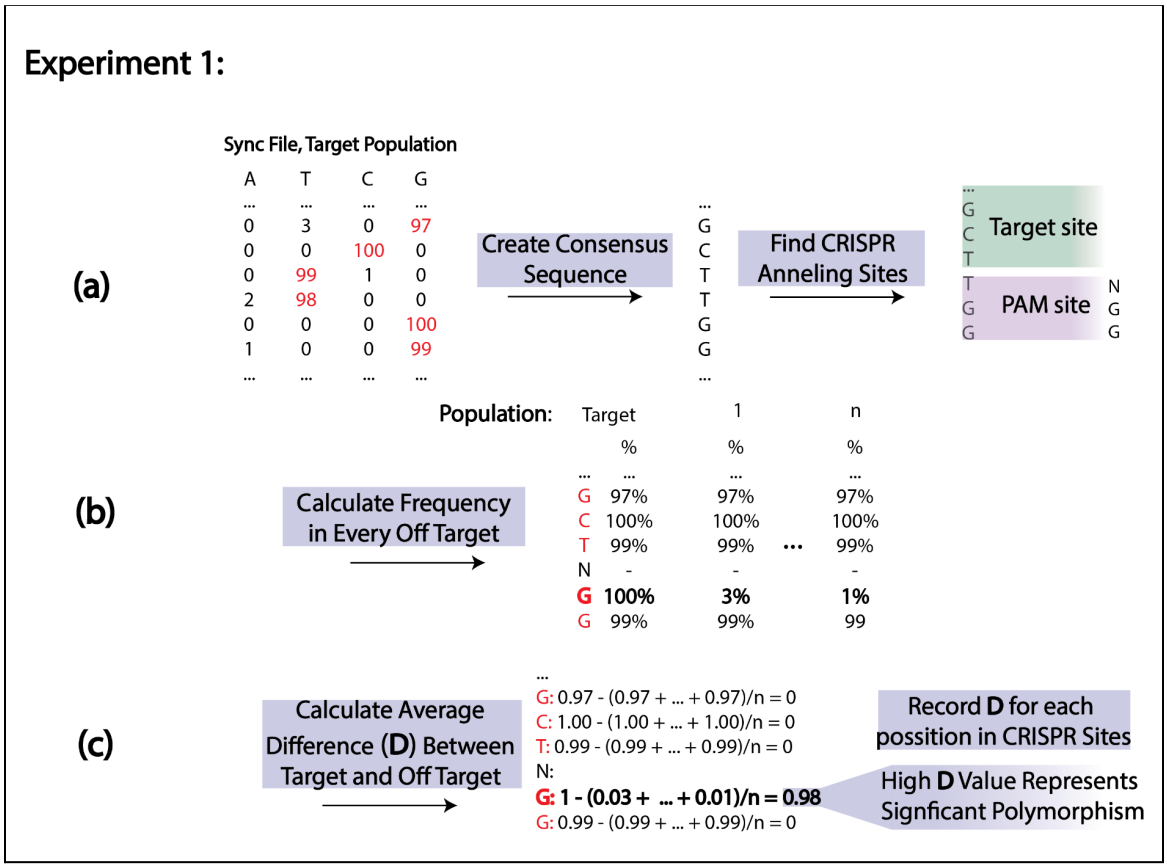

Figure 2: Experiment 1 summary, **find gene drives specific to each population in the dataset.** (a) Create consensus sequence for the target population locus then locate CRISPR/cas9 annealing sites. (b) Calculate the frequency of each nucleotide in the target population in every population. (c) Calculate *D* for each nucleotide position in CRISPR/cas9 annealing sites.

Experiment 2a: Investigate the reliability of Fst and geographic distance as a marker for the abundance of polymorphic CRISPR/Cas9 annealing sites.

Next I sought to identify underlying features that are correlated with the presence of differentiated CRISPR annealing sites between populations. In particular, I test the hypothesis that geographic distance and Fst are indicative of the abundance of polymorphic CRISPR/Cas9 annealing sites between populations.

This hypothesis was tested through pairwise comparisons between each population in the dataset (figure 3). For each pair of populations, one was declared the target population and the other the off-target population. A consensus sequence was created for each of these populations, consisting of nucleotides with the highest number of reads for each genomic position. All the CRISPR/Cas9 annealing sites were then identified in the target population consensus sequence. The number of the CRISPR/Cas9 annealing sites polymorphic to the consensus sequences of the off-target population were recorded. The total number and location of each SNP pertaining to each polymorphic annealing site were also recorded.

The number of polymorphic CRISPR/Cas9 annealing sites between pairs of populations was then quantified against the geographic distance between them. Geographic distance was calculated as the direct path between points along the surface of a sphere scaled to the estimated size of the earth, utilizing the latitude and longitude coordinates of population samples. This measure is also known as orthodromic distance.

Pairwise Fst calculations were conducted for every pair of populations. Fst is variation within populations relative to variation between populations and is an indicator of genetic structure. Here Fst was computed with a sliding window approach utilizing software popoolation, which utilises SNPs between pooled populations for its calculation (Kofler et al., 2011) (Supplementary data 1).

One population showed significantly higher Fst compared to other pairs in the dataset. This population was identified not to be *Lolium Rigdium* and was subsequently removed from the analysis.

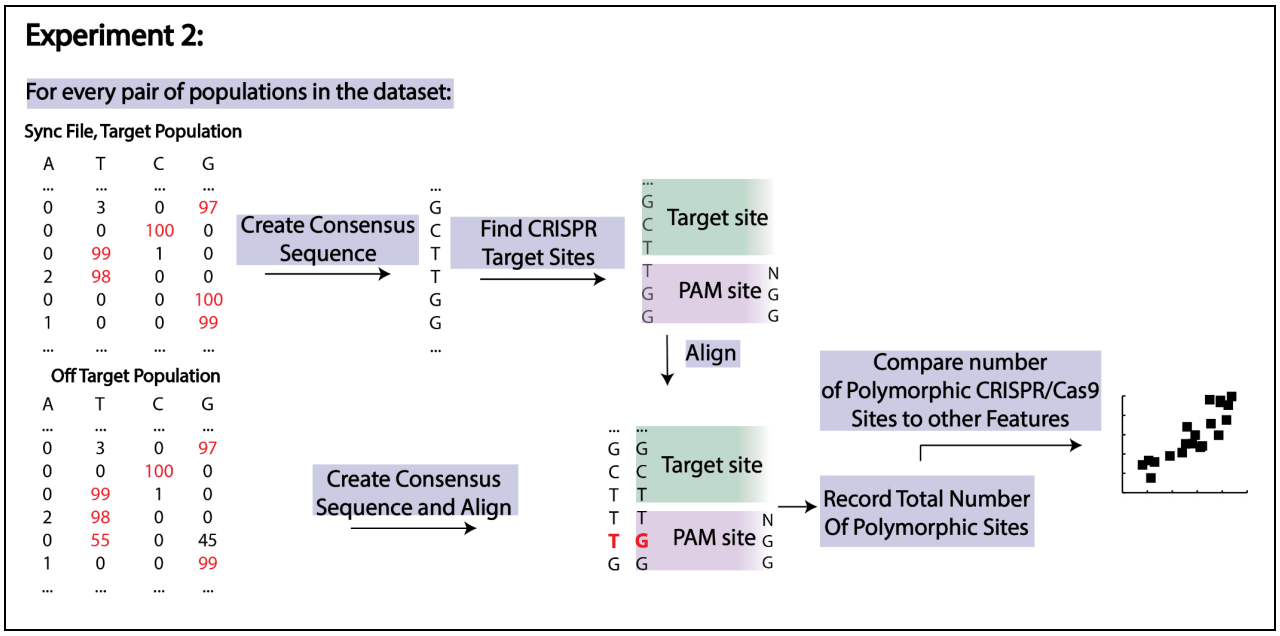

Figure 3: Summary of Experiment 2. Find the number of polymorphic CRISPR/Cas9 sites in the consensus sequence of every pair of populations. Then compare the number of polymorphic CRISPR/Cas9 sites to the graphic distance and Fst between the same pairs of populations.

Experiment 2b: Test the hypothesis that multiple SNPs may be used for specific homing drives.

Utilizing the data from experiment 2a, the number of SNPs pertaining to polymorphic CRISPR/Cas9 annealing sites between pairs of population were recorded.

Computational Packages

Python scripts utilized for this project have been published on github:

https://github.com/Patrick-Gibbs/SCI30001

GNU parallel, a general parallelizer to run multiple serial command line programs, was widely used for timely computation of results (Tange 2021).

Results:

Experiment 1: Limited differentiation was observed within CRISPR/Cas9 annealing sites.

From the 243,665 nucleotide positions in the dataset, 800,948 CRISPR annealing sites were found over all 63 populations. Of these annealing sites, the *D* score was calculated for each nucleotide position. The mean median and standard deviation of these *D* scores was

1.24×10^{-4}, 3.49×10^{-4}, 9.45×10^{-3} respectively. Thus, the majority of CRISPR/Cas9 annealing sites show very little polymorphism (Distribution shown figure 4a).

However some populations had significant differentiation within their observed CRISPR/Cas9 annealing sites. In particular, 5 populations had a *D* value above 0.6, representing 72 highly differentiated CRISPR/Cas9 annealing sites between them (peak of figure 4b).

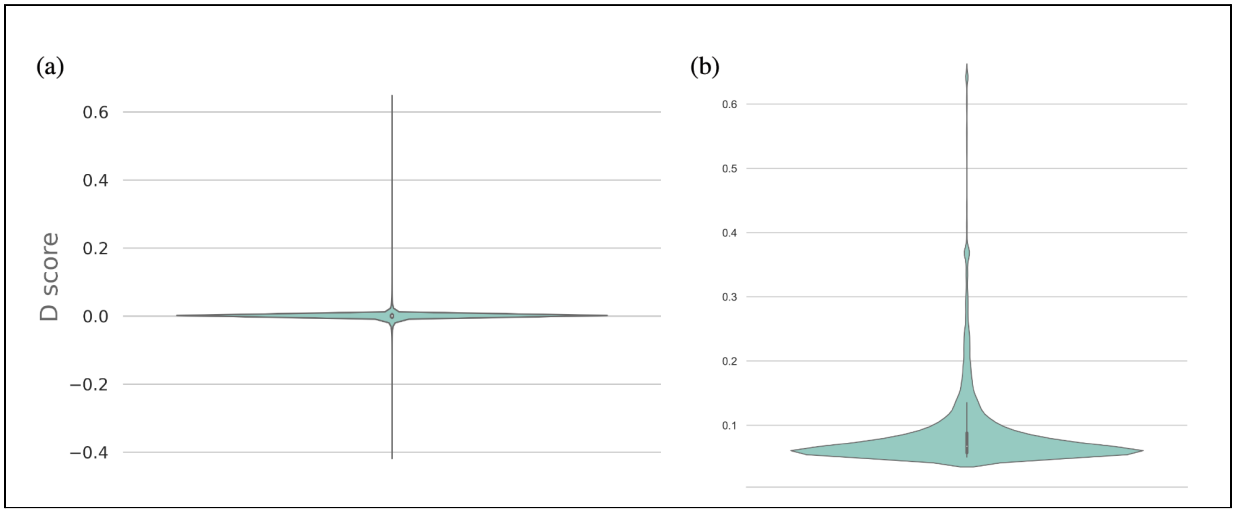

Figure 4: Distribution of D score pertaining to CRISP/Cas9 annealing sites across all populations. (a) Distribution of D score for all nucleotides is CRISPR annealing sites, illustrating low standard deviation. (b) Distribution of D score above 0.05, showing high outliers.

To summarize the applicability of allele specific homing drive, I took the highest D value for each of the populations. Of the maximum values I then took the mean, which is described in table 1.

For polymorphic CRISPR/Cas9 annealing sites, the median of the maximum *D* values for each population was 0.2. Thus, if a targeted gene drive were created for an arbitrary population, there is a 50% chance that the most differentiated target site will be less than 20% more prevalent in the target population than the off-target populations. If the design of homing drives required placement in a PAM site or a gene, differentiated loci are even fewer (table 1).

Modeling has indicated that such differentiation between the target and off-target does not reliably negate defusion of the gene drive through the off-target population (Sudweeks et al 2019; Willis and Burt 2021). Thus for an arbitrarily chosen population, it is unlikely a specific homing drive can be applied effectively.

Median of Maximum *D* Values Across All populations.	**D value**
Whole Genome	0.20
In Genes	0.17
PAM site that lays in genes	0.11

Table 1: The median of the combined maximum *D* values for each population pertaining to nucleotides located in CRISPR/Cas9 annealing sites.

However, select populations had substantial outliers for differentiation in CRISPR annealing sites, more tenable to an allele specific drive. One outlying population contains a CRISPR annealing site which is fixed and at an average rate of 0.29 in other populations. Such outliers present potential gene drive targets (table 2).

Highest *D* value	**D value**	**Proportion Susceptible in the Target Population %.**	**Average Susceptibility in Off-Target populations (%)**
Across all populations	0.69	100	29%
Within Gene	0.64	100	36%
PAM site within Gene	0.64	100	36%

Table 2: Maximum D values of all nucleotides in CRISPR/Cas9 annealing sites, across all populations in the dataset.

Experiment 2 Lolium rigidum is genetically Heterogeneous Across Victoria, and Fst is a strong indicator of abundance of polymorphic CRISPR sites between populations.

Combining the results for every pairwise comparison, yielded a total of 6119958 polymorphic annealing sites. For pairs of populations, high correlation was observed between Fst and the number of polymorphic annealing sites (Figure 5, $r^2 = 0.692$). This is expected as broadly Fst and number of polymorphic CRISPR/Cas9 annealing sites are a measure of population structure. This correlation thus validates that the framework developed for experiment 2 functions as intended. Importantly this analysis confirms that the Fst between two populations is a strong predictor of the number of polymorphic annealing sites between them.

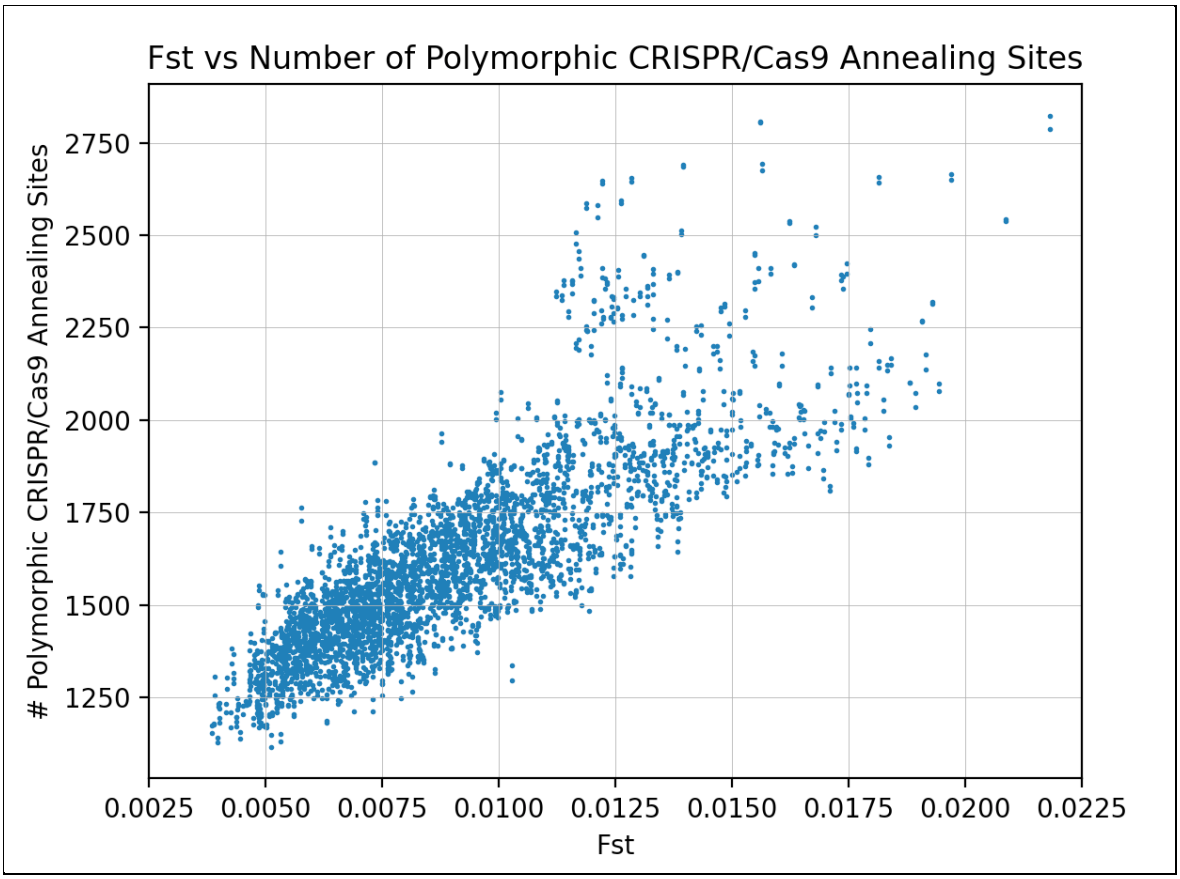

Figure 5: Fst is correlated with the number of polymorphic annealing sites between pairs of population. Figure depicts Fst against the number of polymorphic CRISPR/Cas9 annealing sites for every pair of populations in the dataset.

No correlation was observed between the geographic distance and the number of polymorphic sites between two populations, as well as geographic distance and Fst (Figure 6). This implies genetic heterogeneity between geographic regions, likely due to a mixing between populations across SE Australia. Thus, this dataset suggests that gene drives targeted to a population will likely spread beyond the local geography observed in this analysis. The heterogeneity observed here explains the results of experiment 1. That is, over the dataset analysed here genetic heterogeneity negates the possibility of population specific gene drives.

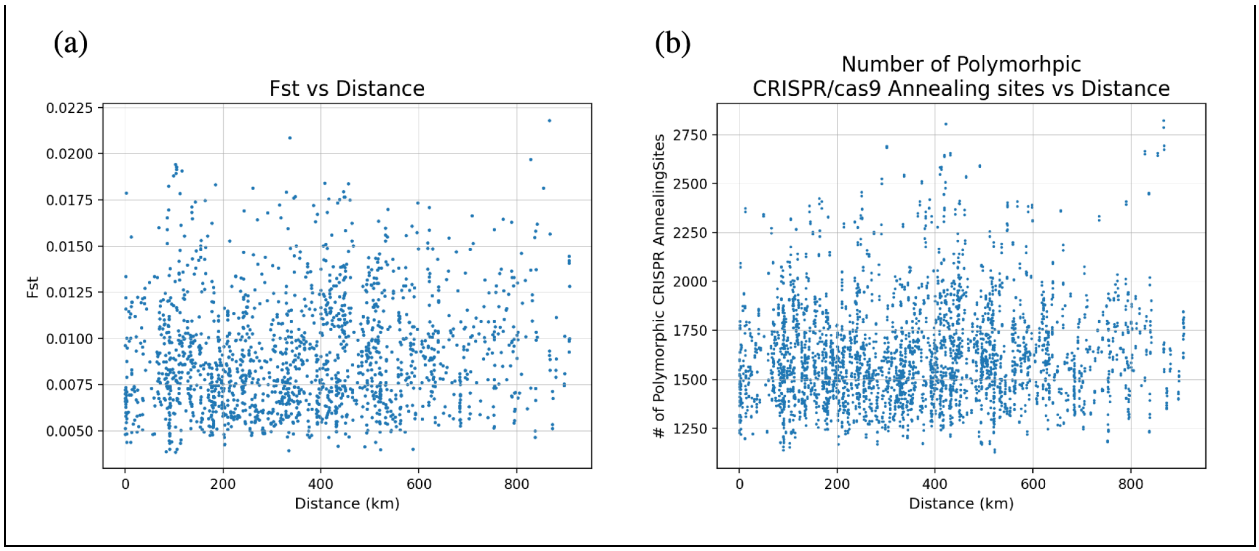

Figure 6: Distance is not a reliable indicator for genetic structure between sampled *Lolium Rigidum* populations. (a) The number of polymorphic CRISPR/Cas9 annealing sites against geographic distance for every pair of populations in the dataset. (b) Fst vs geographic distance for every pair of populations in the dataset.

Experiment 2b The data set shows possibility of using multiple SNPs for robust specificity of population specific homing drives:

It was found that between pairs of populations 762494 of the 6,119,958 (12.78%) polymorphic CRISPR/Cas9 annealing sites had more than one SNP between their aligned consensus sequences. Alleles containing multiple polymorphisms are likely to be highly resistant to off-target Cas9 cleavage (figure 7).

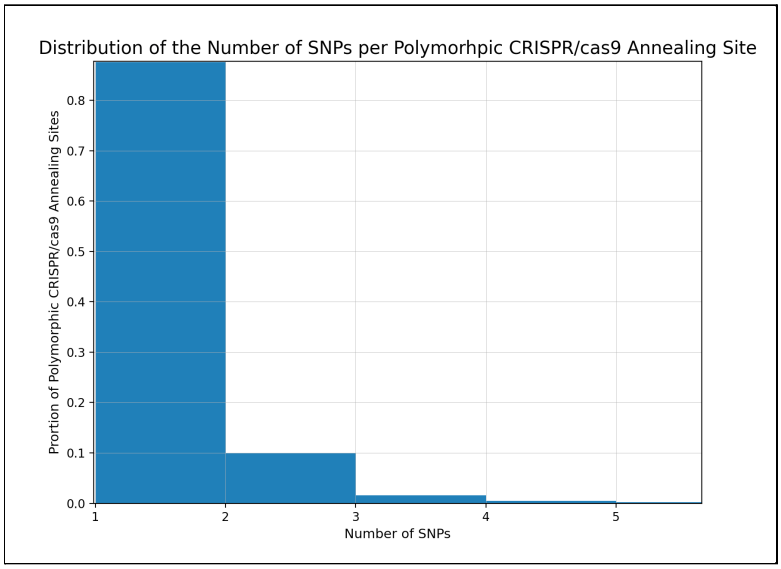

Figure 7: Distribution of SNP number pertaining to polymorphic CRISPR/Cas9 annealing sites.

Discussion:

This research has shown that on a subcontinental geographic scale without barriers to gene flow, there are few CRISPR/Cas9 annealing sites with sufficient polymorphism for population targeted homing drives. The limited number of differentiated CRISPR/Cas9 annealing sites found can be explained as populations in the dataset were shown to be genetically heterogeneous. In other words, geographic distance between any two populations did not correlate with the genetic differentiation between them. It was also found that between two populations Fst is indicative of the abundance of polymorphic CRISPR/Cas9 annealing sites.

Geographically connected populations show limited amenability to population specific homing drives.

Specific gene drives are potentially applicable on a large geographic scale, such as the continent of Africa (Willis and Burt 2021). However, on a smaller geographic scale as shown in the research presented here, heterogeneity between populations negates specificity of CRISPR/Cas9 annealing sites.

Some outliers were identified which showed a high level of differentiation. It is thus plausible that in particular populations, a highly targeted gene drive could be created. However, at such a low proportion it is possible that such a phenomenon is an artifact of the sequencing process. The impact of sequencing errors is reduced utilizing a minimum depth of 50 reads for the analysis, however the possibility of such artifacts is not completely eliminated. It is also possible that some outliers could have been due to hybridisation of other species or mis-specification of samples. Fst measurements were used to remove one such outlier in the dataset, however there is the possibility that some remain.

Assuming the data is reflective of natural populations, a number of other factors in gene drive design may nevertheless render such outliers untenable for gene drive targets. These include off-targets in the genome (Zhang et al., 2015), the chromatin state in that region (Verkuijl and Rots, 2019) and the level

of conservation in a given target. Therefore using such outliers is not a good indication of the design potential for population specific gene drives.

A limitation of this study is results only pertain to 0.03% of the genome which had sufficient sequencing coverage. However in the sample analysed, few populations contained CRISPR/Cas9 annealing sites of significant differentiation and no population had completely private CRISPR/Cas9 annealing sites. Given this definitive lack of genetic variation, it is unlikely that regions of the genome not analysed will yield significantly different results. Therefore, the data suggests a targeted allele specific gene drive is not a robust solution to control Lolium rigidum in southeastern Australia. More broadly, these findings suggest that over small geographic areas, with few barriers to gene flow, population specific gene drives are not a discriminative means of population control.

Nevertheless, the findings of this research could be strengthened through further sequencing allowing for brouder genome coverage. Alternatively individual populations which were under-represented in genome coverage could be removed for the dataset. This would allow a wider analysis of the genome yet capture fewer populations.

***Lolium Rigidum* populations across south eastern Australia are genetically heterogeneous.**

It was found that geographic distance between *Lolium rigidum* populations does not indicate the degree of genetic divergence.

This finding is consistent with prior literature which shows wide gene flow of agricultural pests (Karn and Jasieniuk, 2019). It is conjectured that gene flow may be facilitated through human induced movement such as tractors and other agricultural machinery. Moreover, the lack of geographical barriers between populations likely allows natural gene flow between the studied regions. In cases such as the geography investigated here, distance does not serve as a sufficient barrier to facilitate the population structure required for specific gene drives. Therefore my research suggests any such applications of

specific gene drives for pest suppression pertain to situations of limited gene flow such as isolated island populations (Willis and Burt 2021).

Currently in Australia, invasive species such as mice, rabbits and feral cats cause substantial damage to native flora and fauna in island populations (Courchamp et al., 2003). Such situations present possible applications for allele specific gene drives and form a tenable continuation of this research direction (Leitschuh et al., 2019; Sudweeks et al 2019).

Fst is a reliable indicator of abundance of polymorphic CRISPR/Cas9 annealing sites between populations.

This investigation has shown that Fst may be used to assess the viability of allele specific drives between two populations. These findings are significant in that it allows evaluation of gene drive tenability before a genome wide search for specific annealing site must be completed. This is particularly relevant to pre-existing data sets which can be reassessed to gene drive viability, without the application of a specific framework.

The utilization of the gRNA annealing site for population specific homing drives remains an unexplored research direction.

In contrast to prior literature, research here investigates the presence of SNPs in CRISPR/Cas9 annealing sites in addition to canonical PAM sites (Willis and Burt 2021). In the studied dataset, we note that the most differentiated target allele pertained to a polymorphism in the gRNA annealing site. Indeed, the 20nt annealing site of the gRNA provides 10x more opportunities for polymorphism compared to the 2 conserved nucleotides of the Cas9 PAM site. Given this research has shown the rarity of outlier CRISPR/Cas9 annealing sites with sufficient differentiation for specific homing drives, investigating the tenability of using the broader gRNA annealing site is warranted. However, absolute complementarity of the guide RNA is not required for Cas9 cleavage – some polymorphism can be tolerated at the expense of cleavage efficiency. Here I have shown that 13% of CRISPR/annealing sites

contain more than one SNP, implying scope to utilize multiple mutations in a guide RNA to allow a homing drive to be highly specific (Hsu et al., 2013). However, this analysis utilized sync files which do not contain haplotypic information. It was therefore not possible to know at what rate each SNP in CRISPR/Cas9 annealing sites pertain to the same allele. Thus a continuation of the analysis conducted here would be to analyse individual reads to measure the frequencies of CRISPR/Cas9 alleles containing multiple polymorphisms. This future direction could include interpreting the effect of polymorphisms through models that have been developed to predict off-target Cas9 cleavage efficiency (Alkan et al., 2018). However, additional research would be required to understand how estimations of cleavage efficiency translate to conversion efficiency of the homing drive at a population level.

In addition to the off-target cleavage efficiency, on target cleavage efficiency varies based on gRNA annealing site sequence. Thus, consideration of on-target cleavage efficiency is also important when discriminating appropriate CRISPR/Cas9 annealing sites for population specific drives. Free energy between the guide RNA and annealing site is key in determining cleavage efficiency, and models have been developed to predict it (Xiang et al., 2021). Yet as stated, the relationship between cleavage efficiency and conversion efficiency through HDR, within a homing drive is not well understood. Thus, a limitation of this study and future direction would be to develop models to predict conversion efficiency such that it can be used for discrimination between annealing sites for a homing drive. Such descrimination is important as modeling shows threshold conversion efficiency required for the successful spread of a gene drive carrying a genetic load (Unckless et al., 2015).

Conclusion

In a changing climate it is likely that pest species will be a growing challenge. In the presence of gene flow my research has shown that population specific gene drives are not a robust means to tackle this issue. It is clear that further innovation is required to tackle pest species moving forward – be that gene drives or other.

Supplementary material:

Popoolation Fst parameters:

Using min-count	5
Using min-coverage	20
Using max-coverage	5%
Using min-covered-fraction	0
Using pool-size	375
Using window-size	10000
Using step-size	10000
Using asympt-unbiased	0

References

Alkan, F., Wenzel, A., Anthon, C., Havgaard, J. H., & Gorodkin, J. (2018). CRISPR-Cas9 off-targeting assessment with nucleic acid duplex energy parameters. *Genome biology*, *19*(1), 1-13.

Carballar-Lejarazú, R., Ogaugwu, C., Tushar, T., Kelsey, A., Pham, T. B., Murphy, J., ... & James, A. A. (2020). Next-generation gene drive for population modification of the malaria vector mosquito, Anopheles gambiae. *Proceedings of the National Academy of Sciences*, *117*(37), 22805-22814.

Clark, J. B., & Kidwell, M. G. (1997). A phylogenetic perspective on P transposable element evolution in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(21), 11428–11433. https://doi.org/10.1073/pnas.94.21.11428

Courchamp, F., Chapuis, J. L., & Pascal, M. (2003). Mammal invaders on islands: impact, control and control impact. *Biological Reviews*, *78*(3), 347-383.

Diagne, C., Leroy, B., Vaissière, A. C., Gozlan, R. E., Roiz, D., Jarić, I., ... & Courchamp, F. (2021). High and rising economic costs of biological invasions worldwide. *Nature*, *592*(7855), 571-576.

Drury, D. W., Dapper, A. L., Siniard, D. J., Zentner, G. E., & Wade, M. J. (2017). CRISPR/Cas9 gene drives in genetically variable and nonrandomly mating wild populations. *Science advances*, *3*(5), e1601910.

Gantz, V. M., & Bier, E. (2015). The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science*, *348*(6233), 442-444.

Hammond, A., Pollegioni, P., Persampieri, T., North, A., Minuz, R., Trusso, A., ... & Crisanti, A. (2021). Gene-drive suppression of mosquito populations in large cages as a bridge between lab and field. *Nature communications*, *12*(1), 1-9.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., ... & Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology*, *31*(9), 827-832.

Kofler, R., Orozco-terWengel, P., De Maio, N., Pandey, R. V., Nolte, V., Futschik, A., ... & Schlötterer, C. (2011). PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PloS one*, *6*(1), e15925.

Karn, E., & Jasieniuk, M. (2017). Genetic diversity and structure of Lolium perenne ssp. multiflorum in California vineyards and orchards indicate potential for spread of herbicide resistance via gene flow. *Evolutionary applications*, *10*(6), 616-629.

Leitschuh, C. M., Kanavy, D., Backus, G. A., Valdez, R. X., Serr, M., Pitts, E. A., ... & Godwin, J. (2018). Developing gene drive technologies to eradicate invasive rodents from islands. *Journal of Responsible innovation*, *5*(sup1), S121-S138.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987-2993.

Paril, J. (2020). *Adaptive evolution of Annual ryegrass (Lolium rigidum, L.) to herbicides in agroecosystems* (Doctoral dissertation).

Pimentel, D., & Burgess, M. (2014). Environmental and economic costs of the application of pesticides primarily in the United States. In *Integrated pest management* (pp. 47-71). Springer, Dordrecht.

Rode, N., Estoup, A., Bourguet, D., Courtier-Orgogozo, V., & Débarre, F. (2019). Population management using gene drive: molecular design, models of spread dynamics and assessment of ecological risks. *Conservation Genetics*, *20*(4), 671-690.

Sudweeks, J., Hollingsworth, B., Blondel, D. V., Campbell, K. J., Dhole, S., Eisemann, J. D., ... & Lloyd, A. L. (2019). Locally fixed alleles: a method to localize gene drive to island populations. *Scientific reports*, *9*(1), 1-10.

Tange, Ole. (2021). GNU Parallel 20210822 (\'Kabul\'). Zenodo. https://doi.org/10.5281/zenodo.5233953

Toonen, R. J., Puritz, J. B., Forsman, Z. H., Whitney, J. L., Fernandez-Silva, I., Andrews, K. R., & Bird, C. E. (2013). ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ*, *1*, e203.

Unckless, R. L., Messer, P. W., Connallon, T., & Clark, A. G. (2015). Modeling the manipulation of natural populations by the mutagenic chain reaction. *Genetics*, *201*(2), 425-431.

Verkuijl, S. A., & Rots, M. G. (2019). The influence of eukaryotic chromatin state on CRISPR–Cas9 editing efficiencies. *Current opinion in biotechnology*, *55*, 68-73.

Willis, K., & Burt, A. (2021). Double drives and private alleles for localised population genetic control. *PLoS genetics*, *17*(3), e1009333.

Xiang, X., Corsi, G. I., Anthon, C., Qu, K., Pan, X., Liang, X., ... & Luo, Y. (2021). Enhancing CRISPR- Cas9 gRNA efficiency prediction by data integration and deep learning. *Nature communications*, *12*(1), 1-9.

Zhang, X. H., Tee, L. Y., Wang, X. G., Huang, Q. S., & Yang, S. H. (2015). Off-target effects in CRISPR/Cas9-mediated genome engineering. *Molecular Therapy-Nucleic Acids*, *4*, e264.