PNAS
www.pnas.org

Supplementary Information for

Analysis of off-target effects in CRISPR-based gene drives in the human malaria mosquito

William T Garrood, Nace Kranjc, Karl Petri, Daniel Y. Kim, Jimmy A. Guo, Andrew Hammond, Ioanna Morianou, Vikram Pattanayak, J. Keith Joung, Andrea Crisanti and Alekos Simoni

Alekos Simoni
Email: a.simoni@pologgb.com

This PDF file includes:

Figures S1 to S10
Tables S1 to S3
Legends for Datasets S1 to S6
SI References

Other supplementary materials for this manuscript include the following:

Dataset S1 to S6 provided as Excel datasheets

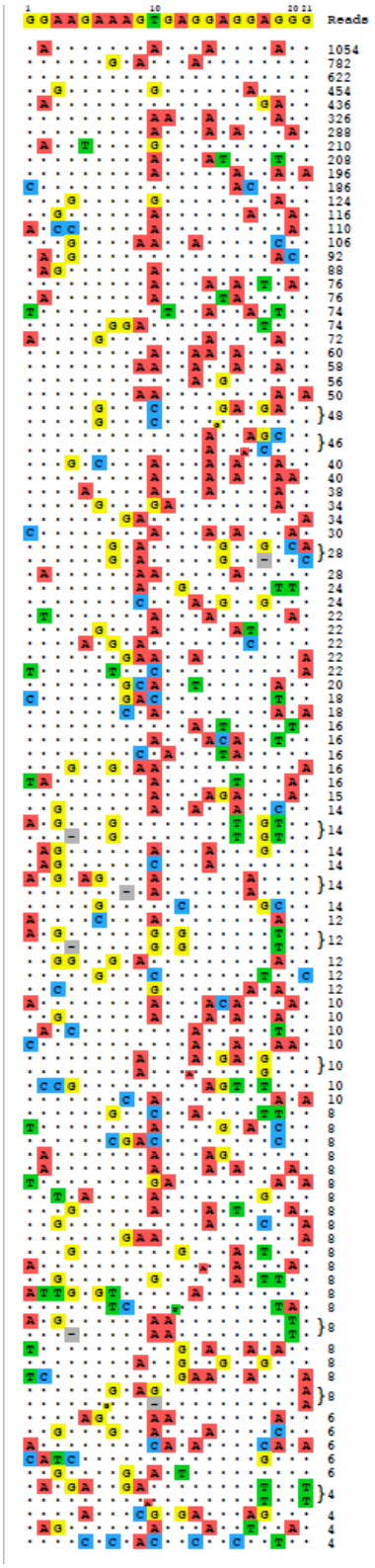

Fig. S1. CIRCLE-seq for gRNA-7280. The reference sequence is the on-target site within the *AGAP007280* gene. Mismatches to the on-target site are highlighted with colored bases, and CIRCLE-seq numbers for each site are displayed in descending order.

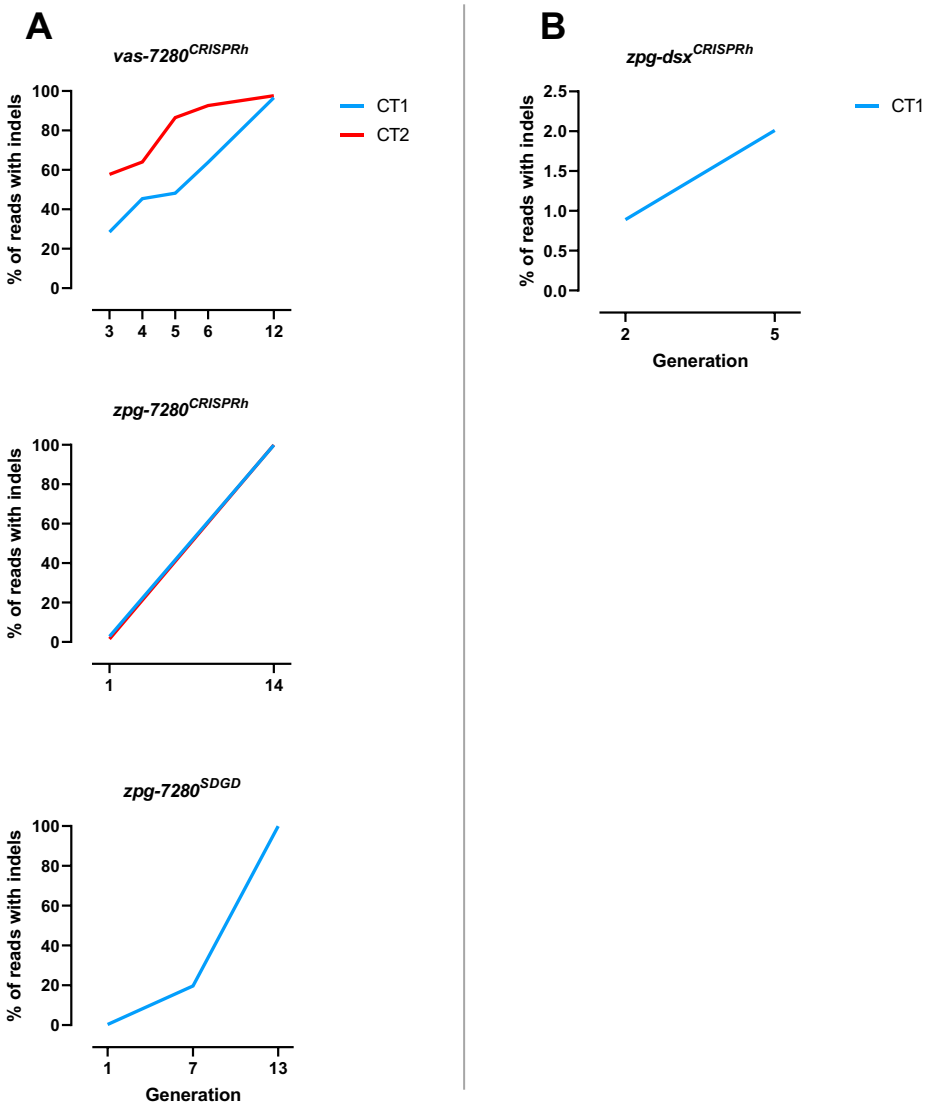

Fig. S2. Indel frequency at on-target sites within mosquito gene drive populations. (A) The frequencies (in percentage) of reads containing indels are displayed for the three different gene drive populations *vas-7280*^{CRISPRh}, *zpg-7280*^{CRISPRh} and zpg-7280^{SDGD} that targeted the *AGAP007280* gene. The indel frequency increases generation upon generation due to the intense selection pressure for mutations that block the activity of the gene drive, but are able to restore function to the gene. Therefore each biological replicate is dominated by one or two in-frame indels, with the near total depletion of the wild-type allele. (B) The frequency (in percentage) of reads containing indels is displayed for the zpg-*dsx*^{CRISPRh} gene drive population. For this population the indel frequency is far lower than seen for gRNA-7280, as indels introduced at this target site are unable to restore function to this gene, and therefore the majority of the alleles remain wild-type. CT1 (cage 1) and CT2 (cage 2) represent biological replicates.

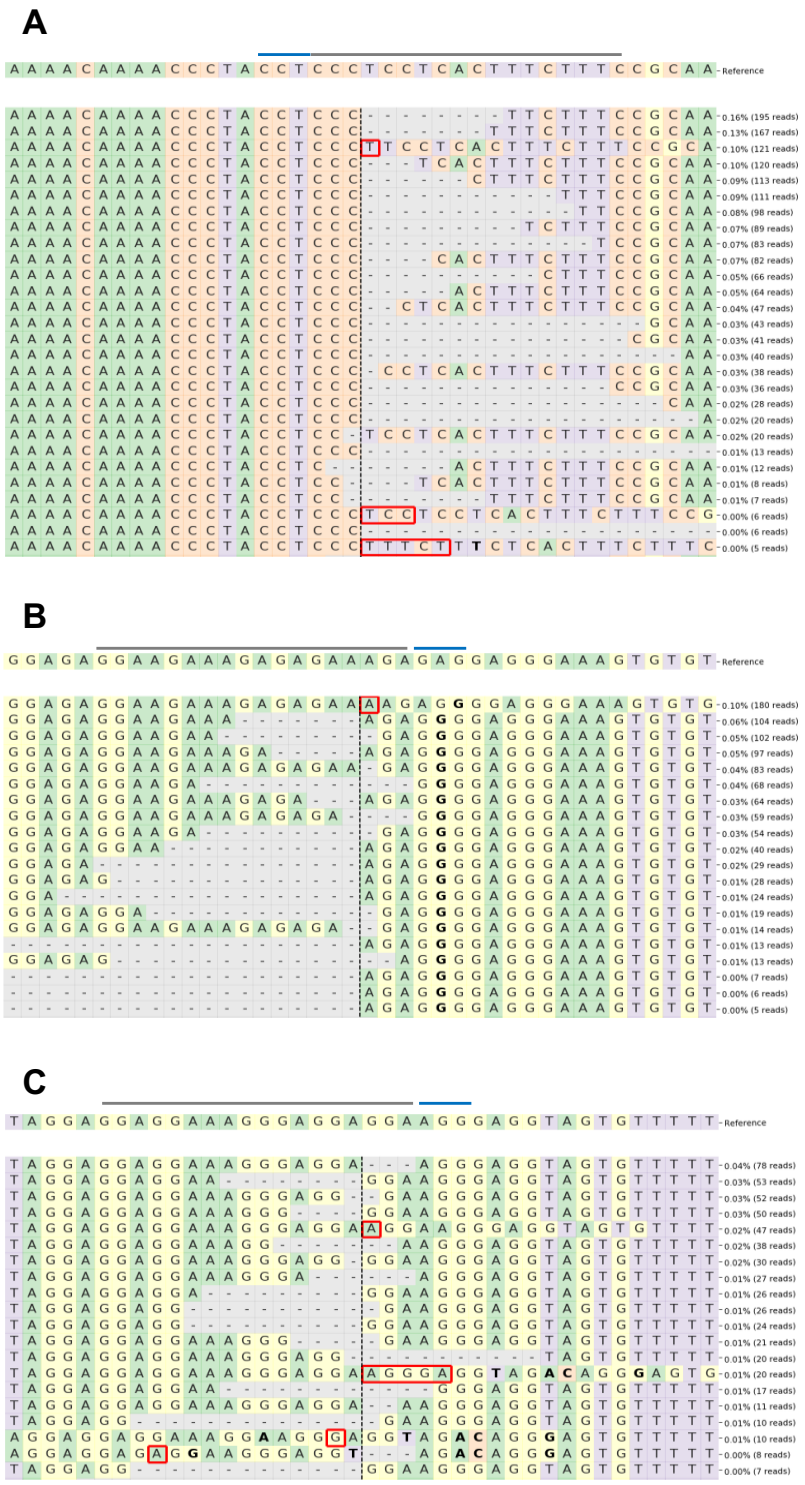

Fig. S3. Example of indels found at three off-target sites. Population displayed here is *vas-7280*^{CRISPRh} at G3 from cage 1. (A) Off-4 (B) Off-6 (C) Off-11. Top most sequence for each section displays the reference sequence for each off-target site, respectively. Read counts for each allele are displayed. Dashes show deletions, while red boxes show insertions. Bases in bold are substitutions, and the vertical dashed line indicates predicted cleavage site. Horizontal grey bar shows predicted gRNA binding site, and the blue bar shows the PAM.

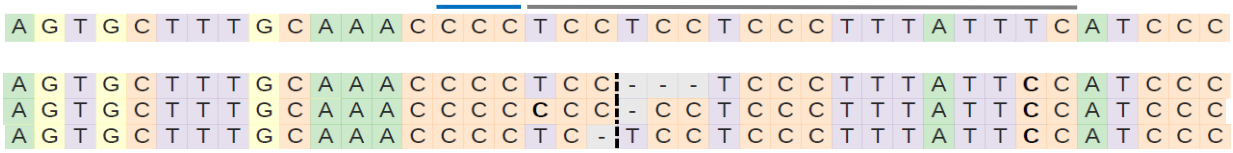

Fig. S4 Example of indels witnessed that did not fit pattern of Cas9 cleavage (off-7). Dashes represent deletions, while bases in bold display substitutions. Horizontal grey bar shows predicted gRNA binding site, and blue bar shows the PAM. These three indels were seen in all populations sequenced, including the wild-type (negative control), and are not indicative of mutations introduced via end joining after Cas9 cleavage.

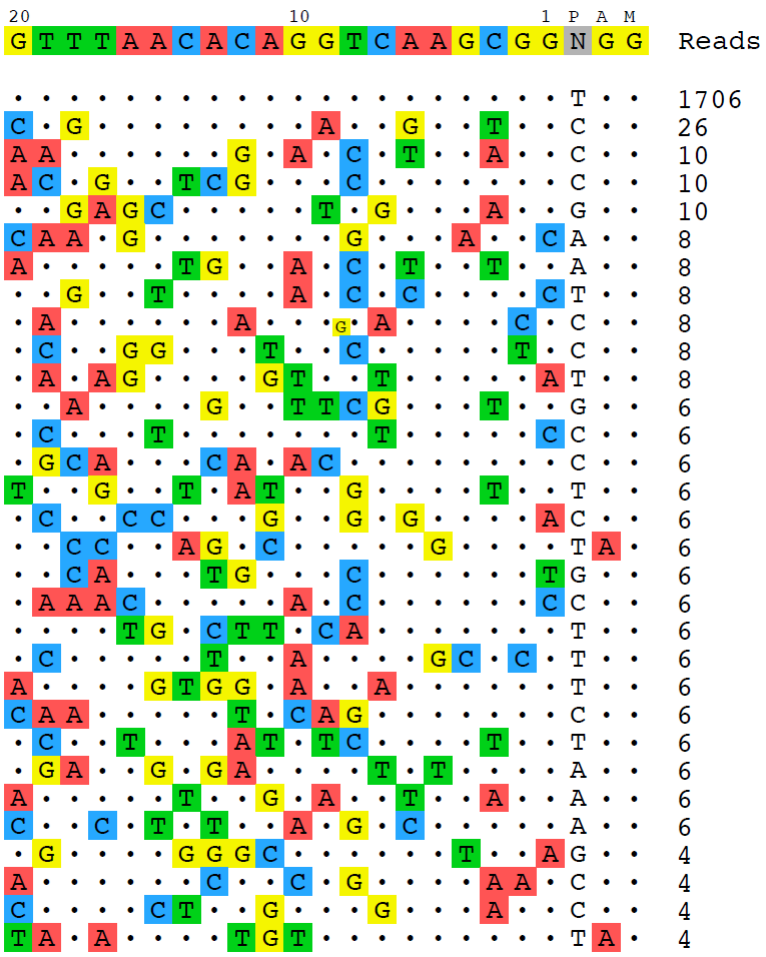

Fig. S5 CIRCLE of gRNA-*dsx*. The reference sequence is the on-target site within the *dsx* gene. Mismatches to the on-target site are highlighted with colored bases, and CIRCLE-seq numbers for each site are displayed in descending order.

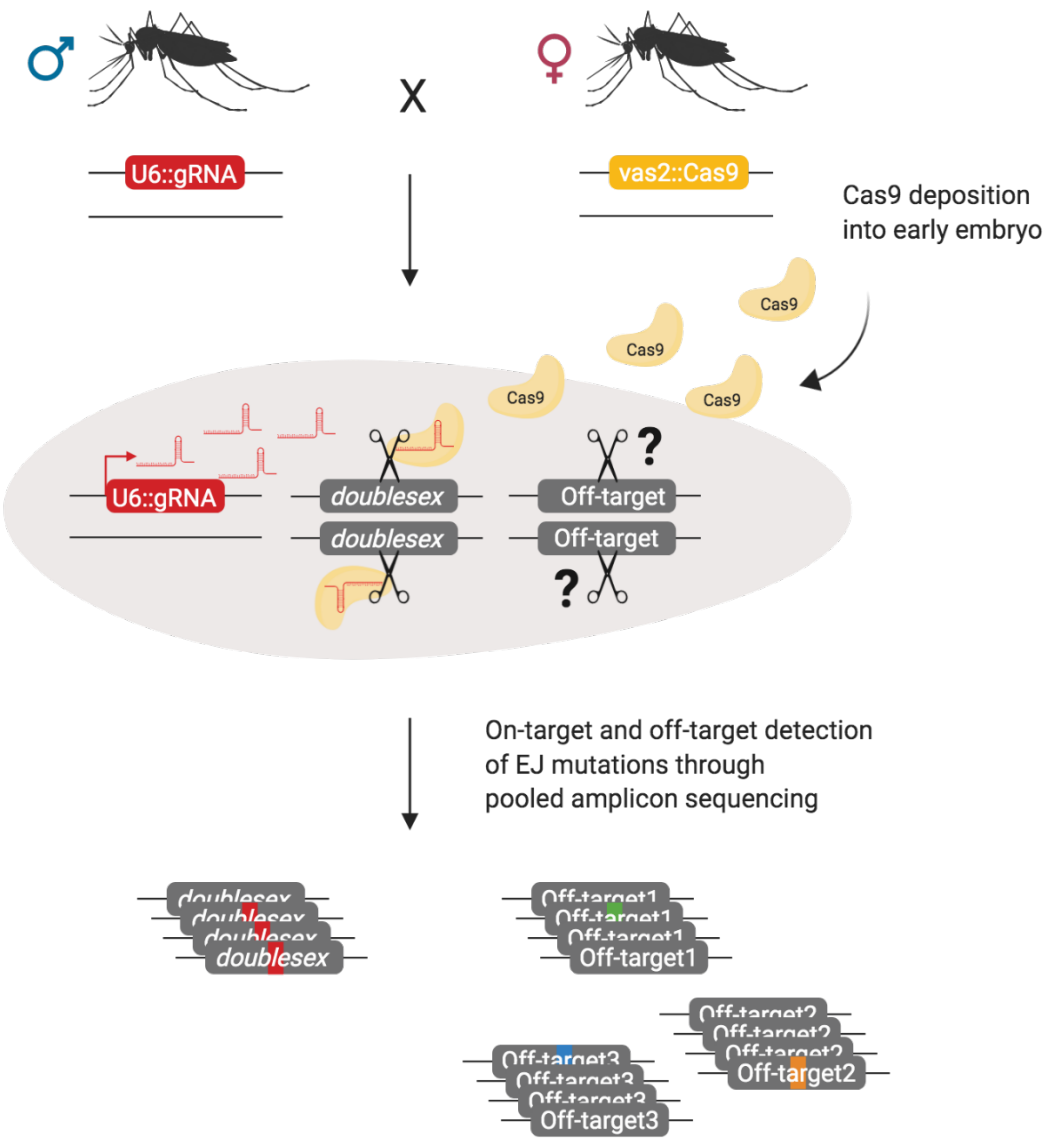

Fig. S6 Crossing scheme to analyze the *dsx* gRNA in combination with Cas9 under control of the *vasa* promoter. Male mosquitoes containing the gRNA-*dsx* were crossed to female mosquitoes that express Cas9, using the *vasa* promoter to drive expression. This Cas9 is deposited into the early embryo and is able to introduce mutations at the *dsx* on-target site. The progeny was analyzed for mutations at CIRCLE-seq predicted off-target sites.

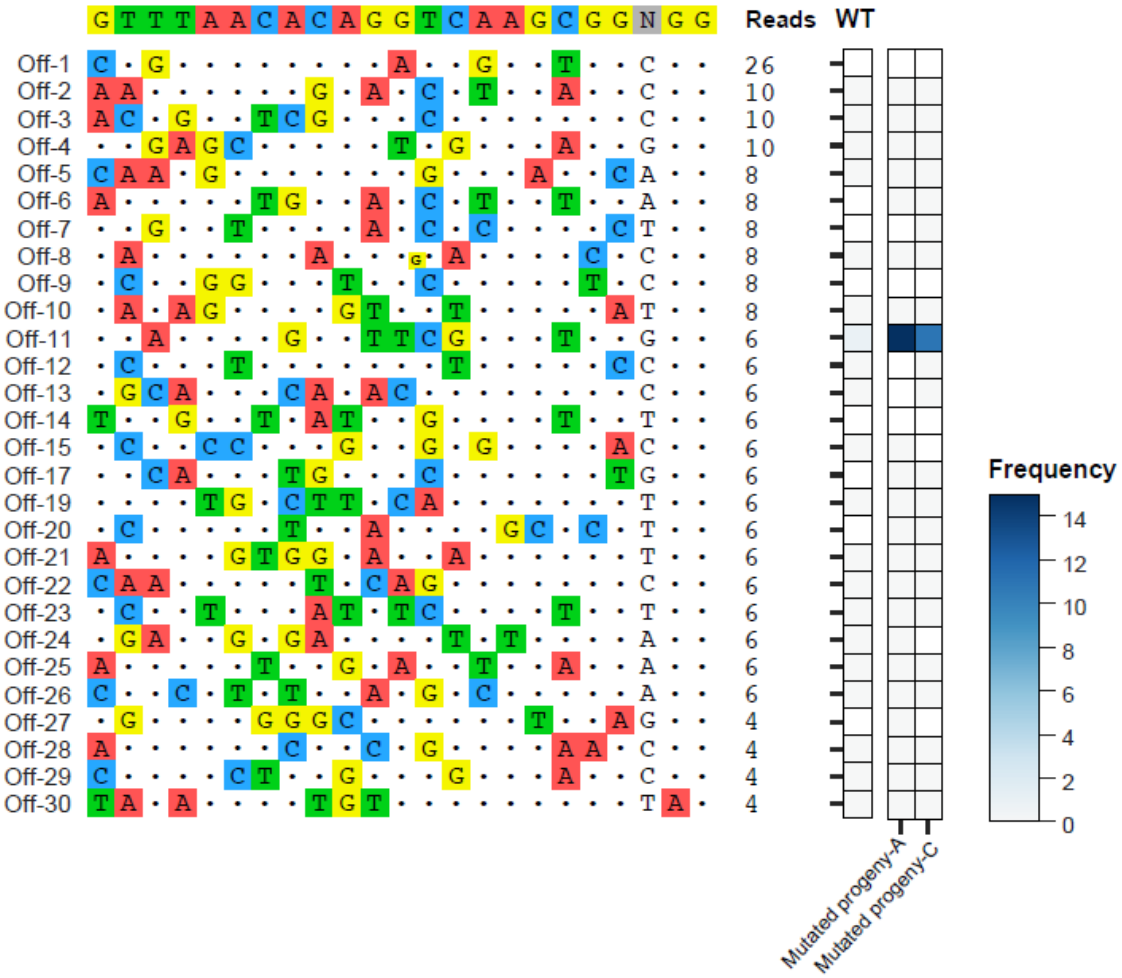

Figure S7. Assessment of off-targets indels within a *vasa-dsx* mosquito population. Amplicon sequencing was conducted for 30 off-target sites predicted from CIRCLE-seq for the *dsx* gRNA. The percentage of reads containing indels is displayed. A wild-type mosquito population served as control. Two different sets of progeny were assessed for evidence of indels (Mutated progeny-A and Mutated progeny-C). There was no evidence of off-target modifications at any of these sites. The indel seen at dsx-off-11 is a genetic polymorphism and not characteristic of Cas9 cleavage.

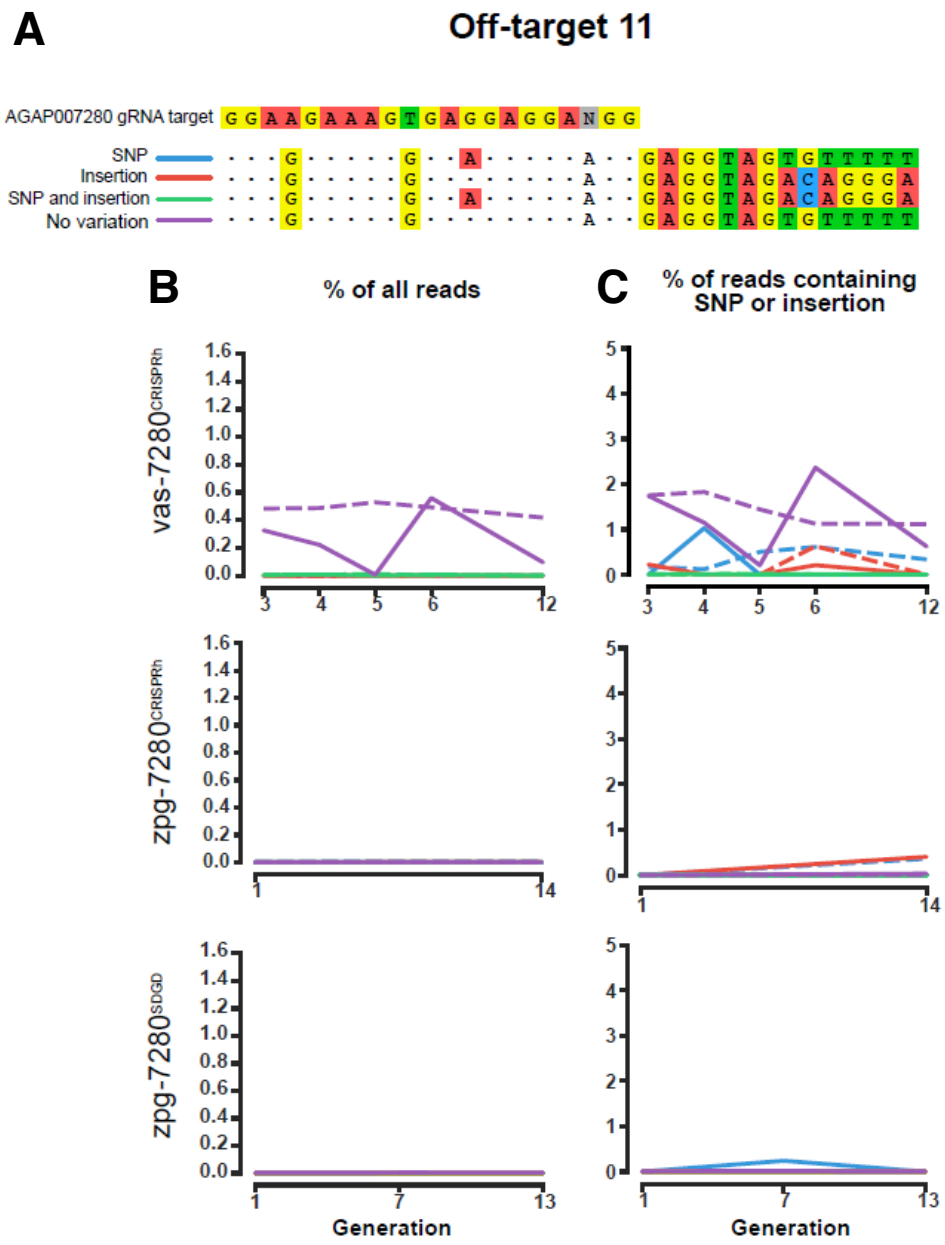

Fig. S8. The impact of genetic variation upon CRISPR/Cas9 off-target cleavage. (A) Off-target site 11 contained four alleles. (B) Indels were detected in one allele (No variation) within *vas-7280^{CRISPRh}* population, with no significant indels detected for *zpg-7280^{CRISPRh}* or *zpg-7280^{SDGD}*. This represents editing as a % of all reads that align to the off-target site. (C) Levels of indels detected are displayed as a % of reads that align to each allele respectively. Indels witnessed for SNP in vas-7280^{CRISPRh} are not evidence of genuine off-targeting editing, but artefacts of very low number of reads aligning to this allele. Cage 1 (solid line) and cage 2 (dash line), with zpg-7280^{CRISPRh} containing one biological replicate only (due to the other cage population crashing at G8).

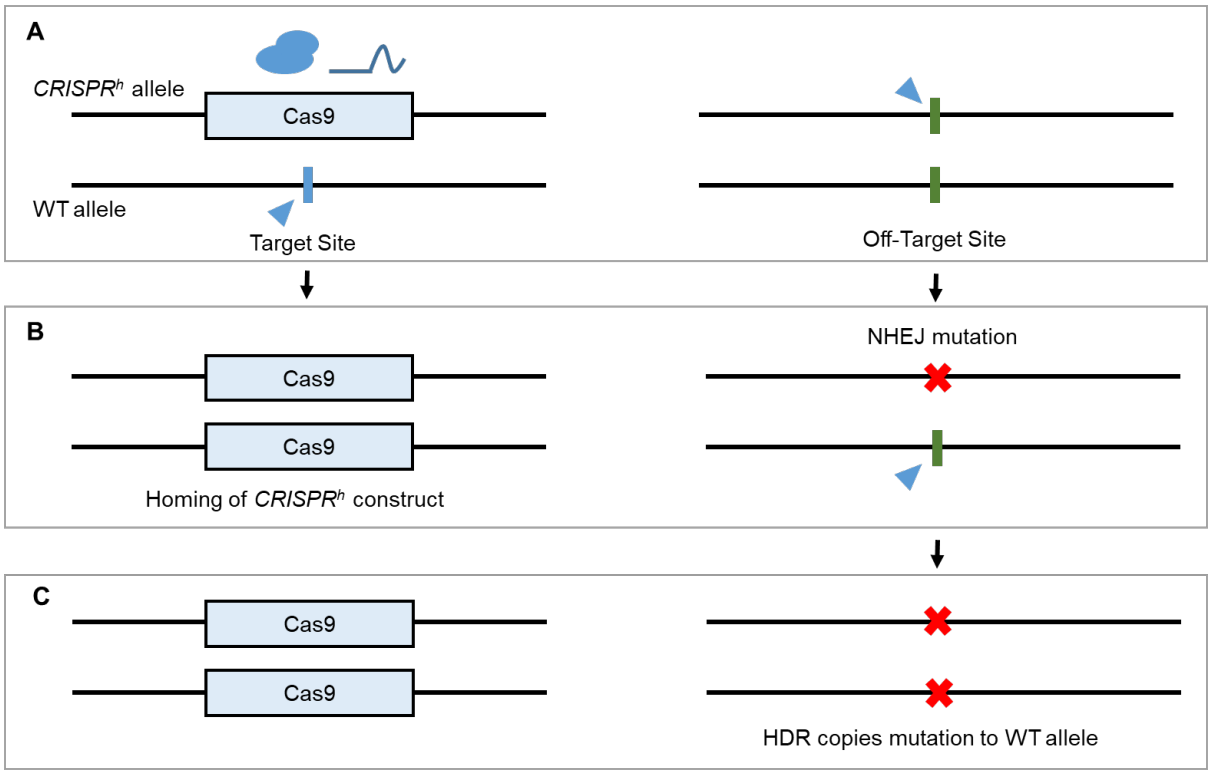

Fig. S9. A schematic of the hypothesized 'Drag-along' drive of mutations. (A) CRISPR/Cas9 cleaves at its on-target site and a genomic off-target location. (B) The on-target site cleavage is repaired by HDR and leads to homing of the CRISPR construct. At the off-target site the cleavage can be repaired by either HDR or NHEJ. If repair is by NHEJ this could cause the introduction of a mutation into this allele. (C) The CRISPR/Cas9 is still able to cleave the wild-type allele at the off-target location. However, if repair of the double strand break (DSB) was to occur via HDR rather than NHEJ then the mutation formerly introduced by NHEJ would then be copied to the other allele, causing an increase in the frequency of that mutation.

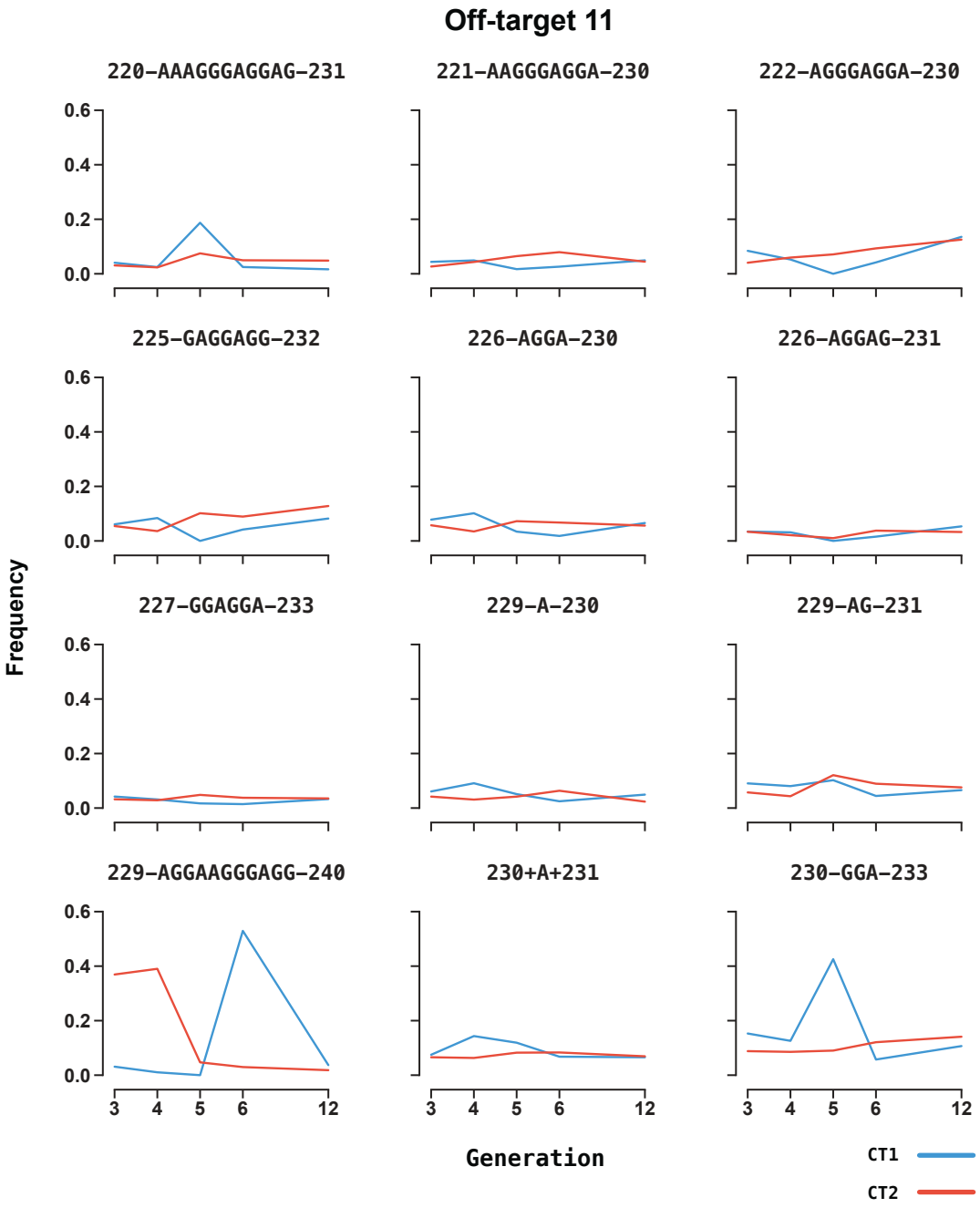

Fig. S10 Indel frequency at Off-11 for *vas-7280-*^{CRISPRh}. The top 10 indels found in the G3 population were individually followed through subsequent generations to assess if frequency changed over time. The frequencies of the indels were normalized to the total number of reads containing indels for that sample. There are 12 indels displayed, as the top 10 indels in CT1 (cage trial 1) were not identical to those witnessed for CT2 (cage trial 2).

Table S1. A summary of the experimental mosquito strains utilized in this study.

Strain name	Construct type	Target gene	Integration locus	Cas9 promoter	gRNA name	gRNA sequence (PAM underlined)	Reference
vas-7280^{CRISPRh}	Gene drive	AGAP007280	AGAP007280	Vas2	gRNA-7280	GGAAGAAAGTGAGGAGGAGGG	(1)
zpg-7280^{CRISPRh}	Gene drive	AGAP007280	AGAP007280	Zero population growth (zpg)	gRNA-7280	GGAAGAAAGTGAGGAGGAGGG	(2)
zpg-7280^{SDGD}	Sex-distorter gene drive	AGAP007280	AGAP007280	Zero population growth (zpg)	gRNA-7280	GGAAGAAAGTGAGGAGGAGGG	(3)
zpg-dsx^{CRISPRh}	Gene drive	AGAP004050	AGAP004050-Exon5	Zero population growth (zpg)	gRNA-dsx	GTTTAACACAGGTCAAGCGGTGG	(4)
vas-Cas9	Ex-locus nuclease source	AGAP004050	Unknown (piggyBac-mediated)	Vas2	gRNA-dsx	GTTTAACACAGGTCAAGCGGTGG	n.a.

Table S2. Predicted off-target sites in gene drive mosquitoes. Cas-OFFinder is the *in silico* prediction tool, whilst CIRCLE-seq is the *in vitro* annotation assay

		Canonical PAM (NGG)							
		Sites with mismatches to gRNA							
Gene	Method	0	1	2	3	4	5	6	7
Dsx	Cas-OFFinder	1	0	0	0	9	127	1142	7740
	CIRCLE-seq	1	0	0	0	1	1	7	19
7280	Cas-OFFinder	1	0	10	115	1139	8377	44763	190965
	CIRCLE-seq	1	0	3	16	30	7	2	3

		Non-canonical PAM (NRG)							
		Sites with mismatches to gRNA							
Gene	Method	0	1	2	3	4	5	6	7
Dsx	Cas-OFFinder	1	0	0	0	9	127	1142	7740
	CIRCLE-seq	1	0	0	0	1	1	7	21
7280	Cas-OFFinder	1	0	10	115	1139	8377	44763	190965
	CIRCLE-seq	1	0	3	16	35	14	6	3

		Non-canonical PAM (1MM in PAM)							
		Sites with mismatches to gRNA							
Gene	Method	0	1	2	3	4	5	6	7
Dsx	CIRCLE-seq	1	0	0	0	1	1	7	21
7280	CIRCLE-seq	1	0	3	24	41	17	8	5

Table S3. A summary of CRISPR/Cas9 specificity in gene drive mosquito populations.

Cas9 promoter	Germline restricted	Target gene	gRNA specificity	Significant Cas9 off-target editing
Vasa	No	*AGAP007280*	Low	Yes
Vasa	No	*doublesex*	Very high	No
Zpg	Yes	*AGAP007280*	Low	No
Zpg	Yes	*doublesex*	Very high	No

Dataset S1 (separate file). Off-target cleavage sites for gRNA-7280 identified by CIRCLE-seq. The chromosomal coordinates are provided, along with gene names and whether the site is located in an exon/intron/intergenic region. The CIRCLE-seq read counts are provided, with the off-target sequence and the number of mismatches to the on-target sequence. The CIRCLE-seq was performed on WT mosquitoes.

Dataset S2 (separate file). Predicted off-target sites for gRNA-7280 analyzed by amplicon sequencing of *vas-7280*^{CRISPRh}, *zpg-7280*^{CRISPRh} and *zpg-7280*^{SDGD}. The chromosomal coordinates are provided, followed by sample information including population, generation, treatment and replicate. The read counts for each sample are shown, along with the reads containing indels, with primers used displayed last.

Dataset S3 (separate file). Off-target cleavage sites for gRNA-*dsx* identified by CIRCLE-seq. The chromosomal coordinates are provided, along with gene names and whether the site is located in an exon/intron/intergenic region. The CIRCLE-seq read counts are provided, with the off-target sequence and the number of mismatches to the on-target sequence. The CIRCLE-seq was performed on WT mosquitoes.

Dataset S4 (separate file). Predicted off-target sites for gRNA-*dsx* analyzed by amplicon sequencing of *zpg-dsx*^{CRISPRh}. The chromosomal coordinates are provided along, followed by indel counts for each sample, and total reads aligned. The indel rate is provided and the statistical test result between the mosquito populations.

Dataset S5 (separate file). Predicted off-target sites for gRNA-*dsx* analyzed by amplicon sequencing from mosquitoes exposed to gRNA-*dsx* and *vas*-Cas9. The chromosomal coordinates are provided, followed by sample information including population, generation, treatment and replicate. The read counts for each sample are shown, along with the reads containing indels, with primers used displayed last.

Dataset S6 (separate file). The impact of genetic variation upon CRISPR/Cas9 off-target cleavage. The off-target site is provided along with sample information including population, generation, treatment and replicate. The total reads is displayed, along with the number edited

SI References

1. Hammond, *et al.*, A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. *Nat. Biotechnol.* **34**, 78–83 (2016)
2. A. Hammond, *et al.*, Improved CRISPR-based suppression gene drives mitigate resistance and impose a large reproductive load on laboratory-contained mosquito populations. *bioRxiv* (2018)
3. A. Simoni, *et al.,* A male-biased sex-distorter gene drive for the human malaria vector *Anopheles gambiae, Nat. Biotechnol. **38**, 1054–1060 (2020)*
4. K. Kyrou, *et al.*, A CRISPR–Cas9 gene drive targeting *doublesex* causes complete population suppression in caged *Anopheles gambiae* mosquitoes. *Nat. Biotechnol.* **36**, 1062–1066 (2018)