Genome and Genetic Engineering of the House Cricket (*Acheta domesticus*): A Resource for Sustainable Agriculture

Aaron T. Dossey [1,*], Brenda Oppert [2], Fu-Chyun Chu [1], Marcé D. Lorenzen [3], Brian Scheffler [4], Sheron Simpson [4], Sergey Koren [5], J. Spencer Johnston [6], Kosuke Kataoka [7] and Keigo Ide [7]

1   All Things Bugs LLC, Invertebrate Studies Institute, Inc., 2211 Snapper Ln., Oklahoma City, OK 73130, USA
2   USDA Agricultural Research Service, Center for Grain and Animal Health Research, 1515 College, Ave., Manhattan, KS 66502, USA
3   Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC 27695, USA
4   USDA Agricultural Research Service, Jamie Whitten Delta States Research Center, 141 Experiment Station Road, Stoneville, MS 38776, USA
5   Genome Informatics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20894, USA
6   Department of Entomology, Texas A&M University, College Station, TX 77843, USA
7   Faculty of Science and Engineering, Waseda University, 2-2 TWIns #02C214, Wakamatsu-cho, Shinjuku-ku, Tokyo 162-8480, Japan
*   Correspondence: aaron.t.dossey@allthingsbugs.com

Abstract: Background: The house cricket, *Acheta domesticus*, is one of the most farmed insects worldwide and the foundation of an emerging industry using insects as a sustainable food source. Edible insects present a promising alternative for protein production amid a plethora of reports on climate change and biodiversity loss largely driven by agriculture. As with other crops, genetic resources are needed to improve crickets for food and other applications. **Methods:** We present the first high quality annotated genome assembly of *A. domesticus* from long read data and scaffolded to chromosome level, providing information needed for genetic manipulation. **Results:** Gene groups related to immunity were annotated and will be useful for improving value to insect farmers. Metagenome scaffolds in the *A. domesticus* assembly, including Invertebrate Iridescent Virus 6 (IIV6), were submitted as host-associated sequences. We demonstrate both CRISPR/Cas9-mediated knock-in and knock-out of *A. domesticus* and discuss implications for the food, pharmaceutical, and other industries. RNAi was demonstrated to disrupt the function of the vermilion eye-color gene producing a useful white-eye biomarker phenotype. **Conclusions:** We are utilizing these data to develop technologies for downstream commercial applications, including more nutritious and disease-resistant crickets, as well as lines producing valuable bioproducts, such as vaccines and antibiotics.

Keywords: food security; cricket; protein; genome; sustainability; agriculture; insects as food and feed; *Acheta domesticus*; insect genome; CRISPR-Cas9; genetic engineering

1. Introduction

Multiple reports demonstrate that the destruction of natural habitats and pollution from human activity, largely a result of land clearing for agriculture and climate change, have led to substantial global biodiversity loss and mass extinction [1–4]. Threats to biodiversity are also threats to all life on earth, including humans [2,5–7]. With historic levels of biodiversity loss and an increasing human population, it is critical to reduce the consumption of natural resources from earth and its ecosphere. The UN expects the human population to grow to nearly 10 billion by 2050 [8], and food demand is projected to increase up to 62% [9]. Climate change, reduced productivity of agricultural lands, overfishing, dwindling freshwater, pollution from fertilizers and pesticides, and a host of other factors resulting from population increase will place a disproportionate burden on

Earth's ecosphere. Population increases and scarcity of natural resources and food will lead to conflict around the world [10]. Nearly half of land on earth is used for agriculture, and at least 70% of agricultural land (30% of the land on earth) is used for livestock [10–13]. Global market demand for protein is projected to increase to $32.56 billion by 2028 [14]. Unfortunately, expanding the amount of land used for livestock production is neither feasible nor sustainable.

The value of insects to life on earth, including humans, cannot be overstated [15,16]. Insects provide critical ecological services, including pollination and nutrient cycling, which support practically all life on earth [16–20]. Not only do insects provide important functions for the natural world, they also are a nearly entirely untapped bioresource for humans. Insects contain a wealth of chemical biodiversity [16,21,22]. For example, insects are a promising source of high-quality protein, as well as oils, chitin nutrients, and other bioproducts with a substantially lower environmental footprint than other livestock [23,24]. Utilization of insects in food in place of less sustainable protein sources, such as vertebrate livestock, will significantly reduce the human impact on the environment, including biodiversity loss, climate change, and clean water depletion.

In the past decade, an insect-based food and feed industry has emerged [25,26]. This industry began based on the mass production of crickets, particularly *Acheta domesticus* (house cricket) and *Gryllodes sigillatus* (banded cricket) for human food [23,27]. As of 2012, at least 2000 tons of crickets were produced annually in the US [25,28]. While eating insects has been promoted in the past [29–35], this new industry offers long term-term solutions. By 2027, the edible insect market is projected to reach $4.63 billion, increasing 26.5% per year [36]. Major investments of as much as $400 million for a single insect farming company have been made [37–39]. These are some of the largest agricultural investments in the world.

Mass produced insects can contribute positively to sustainable human existence but require the development of genetic resources [40,41]. Insect genomes are critical to all fields of biological science, from medicine to agriculture and conservation [42]. *Drosophila melanogaster* was one of the first model organisms with a reference genome assembly [43]. Not only is sequencing more affordable, improvements in long read length and accuracy as well as scaffolding technology have resulted in near complete genome assemblies, often assembled to chromosomal level [44–46]. Food and agricultural systems have benefitted from sequencing, with genetic engineering of crops playing a critical role [47]. The short lifespan and high fecundity of insects offers an opportunity to incorporate genetic modifications rapidly with higher success rates than with larger animals and many plant crops, producing not only food but bioproduction of other valuable materials, such as vaccine antigens. Coupling genome research and CRISPR/Cas9-mediated genome editing offers an unprecedented potential to rapidly improve insect crops, such as increasing nutritional content, reducing allergenicity, providing resistance to diseases, biomanufacturing, improving feed conversion, increasing growth rate, reducing mortality, and other attributes desired by farmers [41,48]. More studies on insect biology in the context of genome sequences are critically important to not only investigate these species while we still can, but also to inspire the world to make greater strides in conserving and preserving biodiversity and natural habitats by demonstrating their value [15].

To provide genetic resources to the insect as food industry, we present a reference genome assembly for *A. domesticus* and demonstrate how genetics can be used to address limitations in these insects as a food resource. The assembly is scaffolded to chromosome level, providing information on promoters and genes needed for genetic transformation. We demonstrate both CRISPR/Cas9-mediated knock-in and knock-out of genes, and discuss implications for the food, biomedical, vaccine, and other bioproduction industries. Specific data includes: (1) metrics of a the scaffolded genome assembly, (2) results utilizing CRISPR/Cas9 technology to genetically modify (knock-in and knock-out) genes in this cricket species (to our knowledge the first published genetically engineered lines of this species), (3) demonstration of successful use of RNAi to reduce gene expression, and

(4) annotation of a select group of genes in our genomic data (including genes related to immunity, as well as scaffolded microbe and virus sequences). Subsequently, we provide a discussion on how this work utilizing modern cutting edge genomic research can play a critical role in utilizing the largest, most diverse yet almost entirely untapped biological resource on Earth: Class Insecta.

2. Results

Genome assembly. The size of the *A. domesticus* genome was measured by flow cytometry as described in [49]. The genome of female *A. domesticus* was larger than males (2378.8 ± 9.7 Mb vs. 2149.6 ± 11.2 Mb, respectively, Table S1). The size of the X-chromosome was estimated as 387 Mb, the difference between the 2C genome size measurement of the XX female and the XO male. The average size of autosomal chromosomes was 177 Mb.

Genomic DNA was extracted from a single *A. domesticus* adult male for long read sequencing, and nuclear DNA from the same adult male was used for long range data in scaffolding. The draft assembly of long read data contained 28,340 contigs with an N_{50} of 321,088 bases (Table 1). The final assembly with long range (Chicago and Hi-C) data contained 16,290 scaffolds originally with an N_{50} of 221.839 Mb and a BUSCO score of 94.8%. After contaminant scaffolds (metagenome, mitochondrial, duplicate, and vector) were removed, the final *A. domesticus* genome assembly contained 9964 scaffolds consisting of 2,346,604,983 bp. The final assembly had 11 larger chromosomes that totaled 2,137,583,504 bp, presumably representing the 10 autosomal chromosomes and single X sex chromosome predicted for males (Figure 1). The longest scaffold in the assembly was 304 Mb (scaffold_14394) and may represent a portion of the X chromosome.
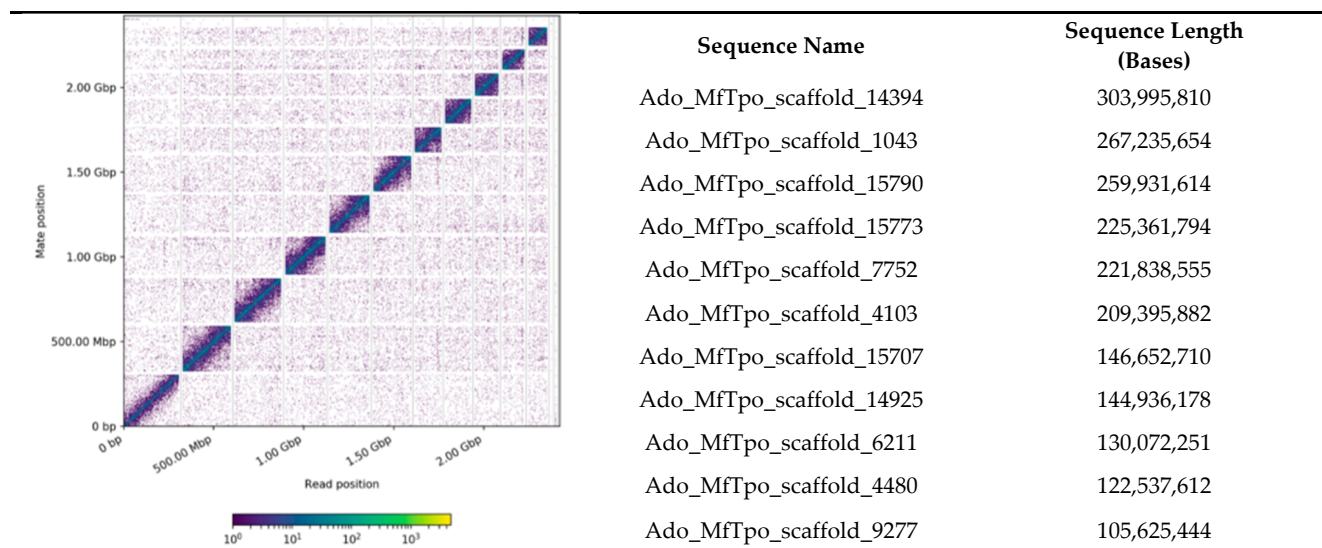
Table 1. Metrics of the *A. domesticus* genome assembly. Data included the CANU draft assembly from PacBio long read data, and the scaffolded assembly from Chicago and Hi-C long range data.

Metric	PacBio Draft Assembly	Chicago/Hi-C Scaffolded
PB coverage	27.3x (Sequel I)	-
#Contigs	28,340	-
ContigN_{50}	321,088 bp	-
Chicago/HiRise coverage	-	10.37x
HiC coverage	-	11,358.72x
Number of scaffolds	-	16,290
N_{50} (Mb)	-	221.839 Mb
Longest Scaffold	-	303,995,810 bp
L_{50} (#scaffolds)	-	5
Busco (Insecta)	-	94.8%

An in silico genome annotation revealed 29,304 predicted genes in 1064 scaffolds, concentrated in the 11 large scaffolds (84%). We compared the length of the BUSCO reference genes from *A. domesticus* to those from a model coleopteran genome, *Tribolium castaneum* (Figure S2). We found that 84% of the reference genes in *A. domesticus* were longer than those in *T. castaneum,* some more than 80-fold longer. The longer length of *A. domesticus* genes was due mostly to longer introns.

There were 6246 scaffolds in the *A. domesticus* genome assembly that were identified as metagenome sequences using multiple automated and manual annotations. Metagenomic scaffolds were from 1013 to 512,394 bp, and one larger scaffold, Ado_MfTpo_scaffold_8829, with 3,776,750 bases and 218 genes was identified as *Acinetobacter baumannii;* these scaffolds were submitted to NCBI as *A. domesticus*-associated metagenome data. Approximately 83% of the metagenome scaffolds had similarity to Invertebrate Iridescent Virus 6 (IIV6, File S3, "Ado_metagenome_Kraken2"). Bacterial scaffolds were mostly from the classes

Gammaproteobacteria, Flavobacteriia, Sphingobacteriia, Fusobacteriia, Betaproteobacteria, Alphaproteobacteria, Actinomycetia, and Bacilli. However, follow-up screening of unclassified reads using blastn [50] and NCBInr found an additional 913 potential metagenome scaffolds (File S3, "Blastn_unclassified"). The final scaffolds were identified as mostly IIV6 and bacteria, but also a smaller number of scaffolds were fungi, protozoa, and nematode (File S3, "Metagenome_summary"). Of note, there were 27 scaffolds from *Wolbachia* similar to those described from other insects. In our analysis, we did not find evidence of common foodborne pathogens, such as *Salmonella* spp., and the few scaffolds aligning to *Listeria* sp., *Escherichia coli* or *Staphylococcus* sp. were insufficient to confirm the presence of these pathogens.


Sequence Name	Sequence Length (Bases)
Ado_MfTpo_scaffold_14394	303,995,810
Ado_MfTpo_scaffold_1043	267,235,654
Ado_MfTpo_scaffold_15790	259,931,614
Ado_MfTpo_scaffold_15773	225,361,794
Ado_MfTpo_scaffold_7752	221,838,555
Ado_MfTpo_scaffold_4103	209,395,882
Ado_MfTpo_scaffold_15707	146,652,710
Ado_MfTpo_scaffold_14925	144,936,178
Ado_MfTpo_scaffold_6211	130,072,251
Ado_MfTpo_scaffold_4480	122,537,612
Ado_MfTpo_scaffold_9277	105,625,444

Figure 1. Link density histogram of the *A. domesticus* Hi-C scaffolded assembly (**left**) and scaffold name and length of the 11 largest scaffolds in the assembly (**right**).

A closer examination of the predicted proteins from the virus scaffolds indicated that only seven were assigned as "high-quality" viral genomes, and 25 scaffolds were determined to be viral genomes of "medium-quality" as determined by CheckV (Ref. [51] File S4). Although many shorter scaffolds were not aligned to the IIV6 genome, many of the predicted genes in these scaffolds were annotated as viral. Therefore, these scaffolds are likely to be viral and quality may be affected by the database integrity, sequence read quality, or assembly artifacts. These sequences also may represent new viruses. Other sequences were classified as low-quality matches or undetermined genomes.

Interestingly, one medium quality scaffold (Scaffold_4233) was similar to a circular virus (DTR_035638) from the metagenome in the reduced digestive system of an oligochaete, *Olavius albidis*, a marine gutless worm related to another *Olavius* spp. with a symbiont metaproteome that evidently contributes to nutrient metabolism [52]. This scaffold was longer (8979 bp) than that of the circular virus (5878 bp).

Regarding immunity genes, conserved sequences involved in insect immunity were used to identify orthologs in the *A. domestica* genome assembly. Genes related to conserved cercropins and defensins were not found in the assembly and have not been identified in the literature from any orthopteran insect to date. However, there were two genes identified as peptidoglycan recognition protein (PGRP) (File S5). Three transcripts found in a previous transcriptome assembly [22] corresponded to these PGRP genes (Figure 2). Only one PGRP transcript appears to be expressed at low levels in the embryo. Expression of all PGRPs, especially ci_142522_1, ramps up during the 1st week of development, and gradually decreases throughout the nymphal stages. PGRPs are expressed at higher levels in female compared to male adult *A. domesticus*.
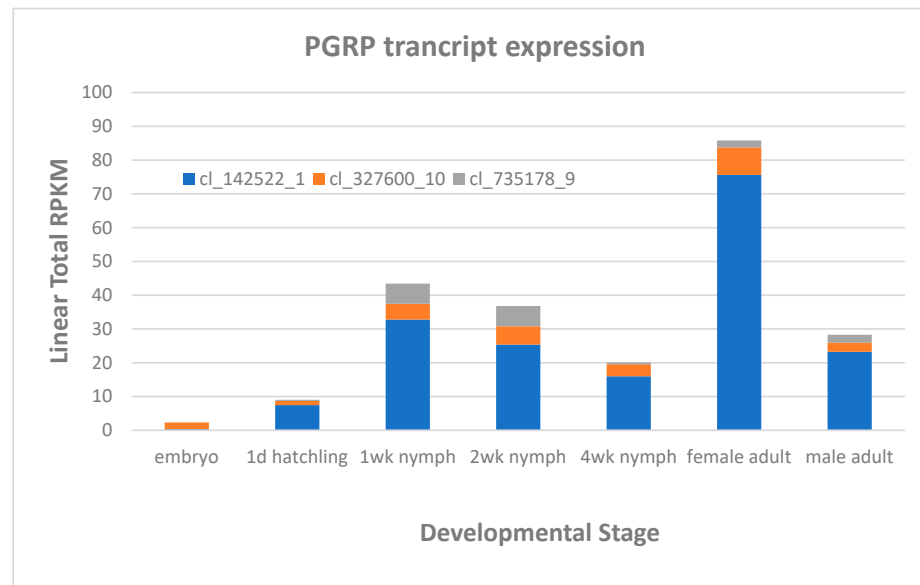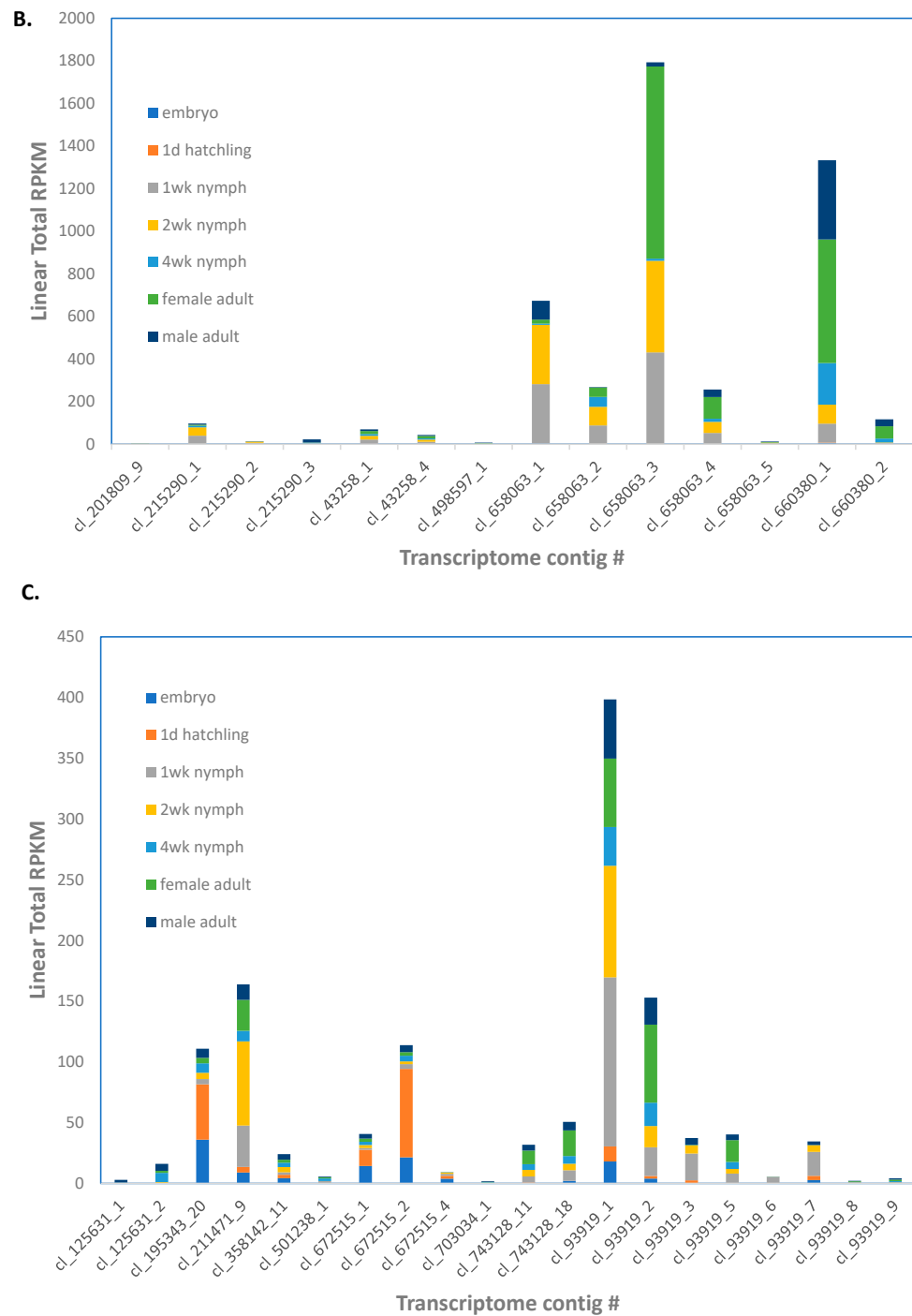
Figure 2. Expression profiles of three potential PGRP transcripts from *A. domesticus* developmental stages or male and female adults. The figure legend has the identification of transcripts from a previous transcriptome assembly [22].

Eleven genes were identified as encoding gram negative binding proteins (GNBP, File S5). These GNBP genes corresponded to 27 transcripts from the transcriptome assembly [22], some which were apparent isoforms of ANN05691-RA (5), ANN05692-RA (3), ANN05929-RA (4), and ANN16377-RA (3). Most of the GNBP transcripts were expressed at low levels in the *A. domesticus* adult (Figure 3A). However, cl_1105989_1 and _2 isoforms of ANN16377-RA were more highly expressed in all life stages except for embryos, as well as male and female adults.
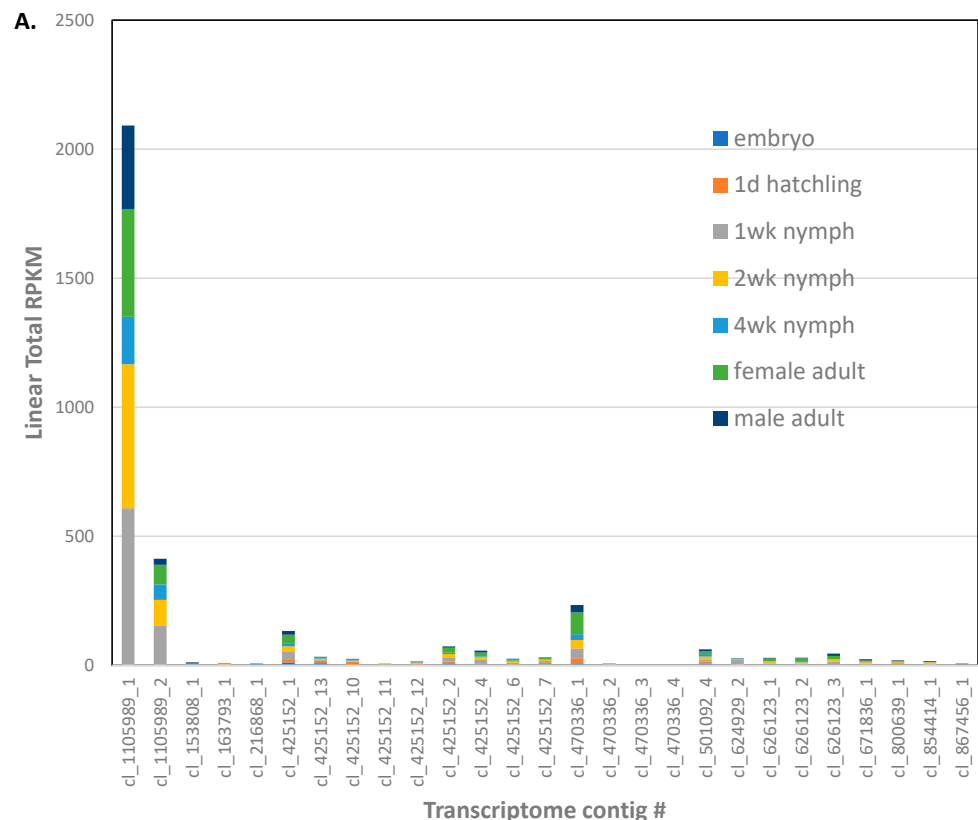
Figure 3. *Cont.*

Figure 3. Expression profiles of transcripts tentatively identified as immune-related genes in *A. domesticus* developmental stages or male and female adults (as indicated in the legend). Transcripts were identified in a previous transcriptome assembly [22]. (**A**) GNBP; (**B**) Lysozymes; (**C**) PPO.

There were four genes annotated as lysozymes in the *A. domesticus* assembly (File S5). Contigs cl_658063 (five isoforms) and cl_660380_1 (two isoforms) corresponding to ANN24139-RA and ANN24140-RA/ANN24139-RA, respectively, were the most highly expressed putative lysozyme transcripts in the transcriptome assembly [22]; ANN19980-RA transcripts were expressed at low levels (Figure 3B). Lysozyme transcripts were expressed higher in 1 and 2 week nymphs and male and female adults.

The *A. domesticus* genome assembly contained nine genes annotated as PPO (File S5). Contig cl_93919 (nine isoforms), corresponding to all nine genes, was more highly expressed throughout life stages (Figure 3C). However, cl_672515 (three isoforms) corresponding to ANN15897-RA, and cl_195343 corresponding to ANN15899-RA were more highly expressed in newly hatched larvae.

Regarding repetitive sequences, we annotated repetitive elements of the newly assembled genome of *A. domesticus* along with those in previously sequenced orthopteran genomes. We compared the repetitive elements in the genome assemblies of *A. domesticus* to those in other crickets (*Apteronemobius asahinai*, *Gryllus bimaculatus*, *Laupala kohalensis*, *Teleogryllus occipitalis*, *T. oceanicus*) and a locust, *Locusta migratoria*. This approach identified repetitive elements ranging from 33.00% of the *G. bimaculatus* genome to 58.13% of the *L. migratoria* genome (Table 2). Among the cricket genomes, *A. domesticus* had the highest overall content of repetitive elements at 49.42%. Among the major classes of the repetitive elements, DNA transposons were the most abundant element in the *A. domesticus* genome, accounting for 8.26% (Figure 4A). The major classes of repetitive elements in the cricket genomes examined were not uniformly present but were biased by species. For example, in the cricket genomes, *L. kohalensis* LINE had the highest abundance at 13.4% while *G. bimaculatus* and *T. oceanicus* had only 3.34% and 4.04% LINE, respectively.

Table 2. Comparison of the size of the genome of orthopteran species, repetitive content, and the mean length of introns.

Species	Genome Size (bp)	Repetitive Content (%)	Median Length of Intron (bp)	Mean Length of Intron (bp)
Acheta domesticus	2,254,915,693	49.42	602	1720.0
Apteronemobius asahinai	1,676,217,857	39.68	356	1000.4
Gryllus bimaculatus	1,658,007,496	33.00	699	4124.6
Laupala kohalensis	1,595,214,429	43.16	802	2001.7
Locusta migratoria	6,476,667,755	58.13	6307	11,988.4
Teleogryllus occipitalis	1,933,897,352	45.90	864	2974.3
Teleogryllus oceanicus	2,045,067,382	38.91	719	2453.7

Because repetitive elements, in particular transposable elements (TEs), are well-known contributors to genome size expansion in a wide range of insect species [53,54], we further examined the relative contribution of repetitive elements to the genome sizes of the orthopteran species. There was a significant correlation between total contents of repetitive elements and genome sizes in Orthoptera species ($p = 0.024$ and $R^2 = 0.61$) (Figure S6a). Because the contribution of the large genome size of *L. migratoria* may be significant in this analysis, we excluded the data of *L. migratoria* in the analysis and found that there was no correlation ($p = 0.166$ and $R^2 = 0.27$ (Figure S6b). We then dissected the repetitive elements into the major classes of TEs (i.e., SINE, LINE, LTR, DNA, and rolling circle) and investigated their contributions to genome size. As a result, we found that LINE, LTR elements and DNA transposon may contribute to the expansion of genome size in Orthoptera (LINE, $p < 0.001$ and $R^2 = 0.63$; LTR, $p < 0.001$ and $R^2 = 0.58$; DNA, $p < 0.001$ and $R^2 = 0.94$) (Figure 4B). When *L. migratoria* data was excluded, we found that the contents of LTR elements and DNA transposon were positively correlated with genome sizes (LTR, $p < 0.001$ and $R^2 = 0.43$; DNA, $p < 0.001$ and $R^2 = 0.50$) (Figure S6a).
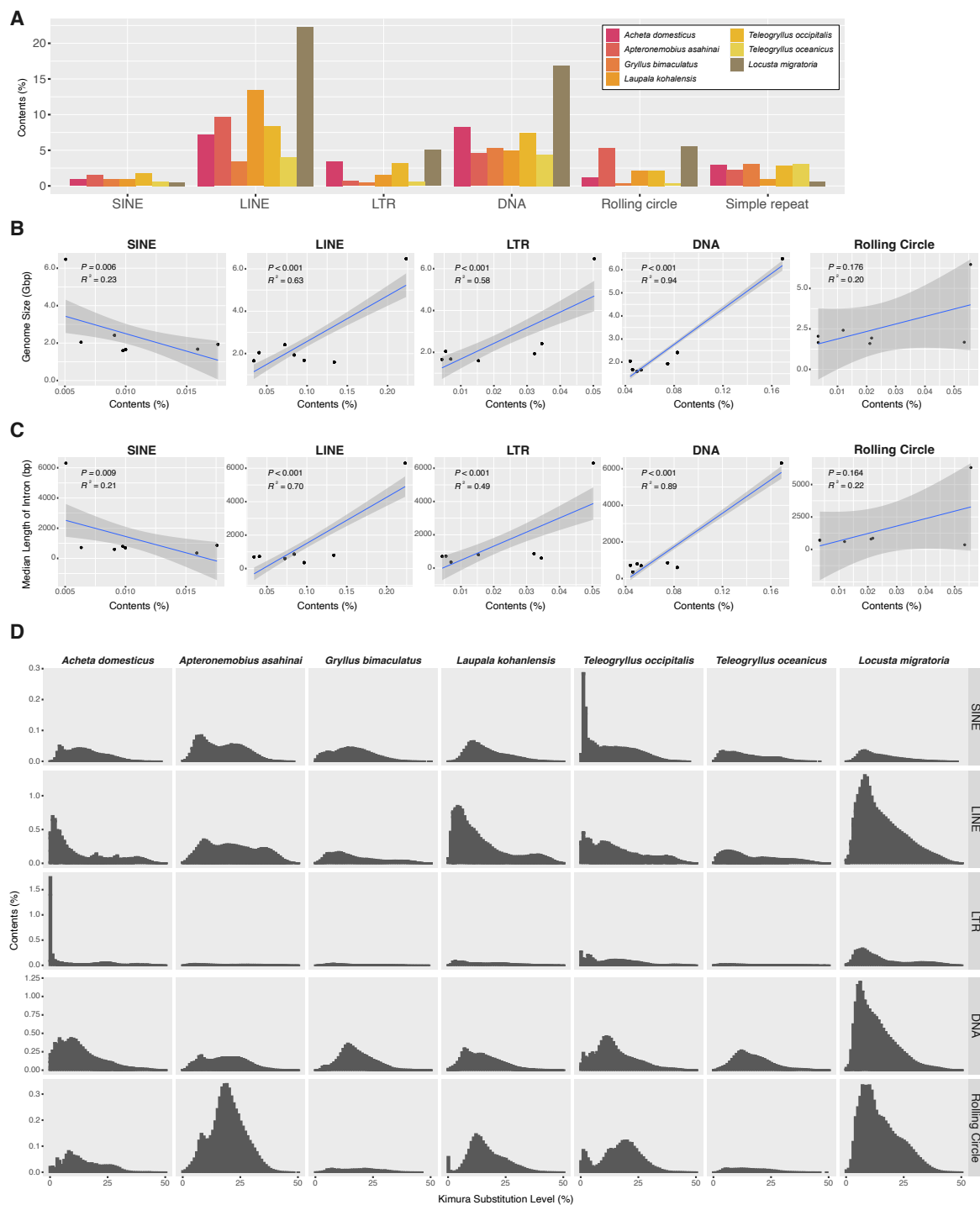
Figure 4. Transposable elements (TEs) in publicly available orthopteran genomes (six crickets, including *Acheta domesticus*, *Apteronemobius asahinai*, *Gryllus bimaculatus*, *Laupala kohalensis*, *Teleogryllus occipitalis*, *Teleogryllus oceanicus*, and one locust, *Locusta migratoria*). (**A**) Contents of representative repetitive elements, including TEs, in Orthoptera. The correlation between genome coverages of major TE classes (SINE, LINE, LTR, DNA, and rolling circle) and genome sizes; (**B**) median length of intron; (**C**) Adjusted by Pearson correlation *p*-values (P) and coefficients (R2); (**D**) Repeat landscape of the major classes of TE classes (SINE, LINE, LTR, DNA, and rolling circle) in the analyzed genomes. The Kimura substitution levels (*x*-axis) show sequence divergence, or "TE age", for the major classes. The classes that are skewed to the left (less sequence divergence) are estimated to have more recently diversified histories than the classes that are skewed to the right.

We next studied associations among intron length, genome size, and repeat content in Orthoptera, including *A. domesticus*, because they have been reported to be mutually correlated [55]. Median length of the *A. domesticus* gene model was 602 bp, which was about one-tenth of that of *L. migratoria* (Table 2). We found that total contents of repetitive elements and median length of introns in orthopteran species, including locust, were positively correlated ($p = 0.044$ and R2 = 0.51), but not within the cricket species ($p = 0.791$ and R2 < 0.01) (Figure S6b). Further, correlations between intron length and major TE contents were positive for LINE, LTR elements and DNA transposon in Orthoptera (LINE, $p < 0.001$ and R2 = 0.70; LTR, $p < 0.001$ and R2 = 0.49; DNA, $p < 0.001$ and R2 = 0.89) (Figure 4C), but not for SINE and rolling circle. These correlations were no longer established except for DNA transposons after limiting the analysis to crickets only (DNA, $p = 0.034$ and R2 = 0.05) (Figure S6b). There also was significant and strong positive correlation between the median length of introns and genome size in orthopteran species ($p < 0.001$ and R2 = 0.97) (Figure S6c). However, when limited to the analysis within crickets, genome size did not correlate with intron length ($p = 0.949$ and R2 < 0.01). Furthermore, a more detailed examination of the classes of TEs in Orthoptera revealed a positive correlation with intron length and genome size for LINE/CR1, LINE/L1, LINE/Penelope, LINE/RTE, DNA/hAT, DNA/Kolobok, and DNA/TcMar (Table S7). These results suggest that repetitive elements, especially several classes of TEs, such as LINE and DNA transposon, may contribute to variation in genome size and intron length in Orthoptera, and that while they are unlikely to be highly correlated within closely related species belonging to the same family, they are more likely highly correlated at the suborder level.

We also examined whether abundance pattern of TEs in orthopteran genomes is due to shared ancient proliferation events or recent/ongoing activity of TE. We analyzed the distribution of sequence divergence of the annotated TE to infer the timing of the change in TE composition. As a result, we found TE distribution patterns that were *A. domesticus*-specific or common to crickets (Figure 4D). In *A. domesticus*, distributions of LTR transposon, especially LTR/ERV1 and LTR/Gypsy, showed a high abundance of recently diverged TE sequences compared to the other species, indicating an ongoing proliferative activity (Figure S6d–f). Likewise, a similar pattern was seen in SINE of the *T. occipitalis* genome. On the other hand, the distribution pattern of DNA transposons in the orthopteran genomes examined here is relatively similar among the crickets, suggesting a possible shared activity of DNA transposons in ancient lineages of the crickets.

Regarding the promoters and genes annotated for genetic engineering, for genetic engineering experiments, we identified promoters of specific genes of interest (GOI). First, we identified and annotated the vermilion gene from *A. domesticus* (Ad vermilion) (Figure 5a) as our target for CRISPR editing and the muscle actin gene from *A. domesticus* (Ad muscle actin) (Figure 5b) for a promoter for GOI expression. The Ad vermilion mRNA was identified from our previously published *A. domesticus* transcriptome data [22] by tBlastn using the *T. castaneum* vermilion amino acid sequence [56]. An incomplete mRNA was identified as the Ad vermilion mRNA. We then used this mRNA sequence to annotate the Ad vermilion gene and map the intron–exon splicing sites to identify each exon. The Ad muscle actin gene and its promoter were identified the same way initially using the *T. castaneum* muscle actin gene. In addition, we were able to identify two very similar actin genes in the *A. domesticus* genome. Since muscle actin genes in many insect species are a single exon gene, we identified one likely muscle actin gene with the highest score from blast and having a single exon. Once the gene was identified, we searched the genomic sequence 1.5 kb upstream from the start codon for promoter regions (Figure 5b).

We targeted the Ad vermilion gene to also serve as a convenient visual marker (white/vermilion eye color phenotype) for preliminary identification of positive knock-out/knock-in G0 crickets. The Ad muscle actin gene's promoter was utilized due to its high level of continuous expression in all life stages, as well as its characteristic muscle anatomy expression phenotype, which is easily identified. For our downstream goals of expressing GOI in farm raised insects for food, feed, and bioproduct manufacturing, it is usually

critical to use promoters with high levels of GOI expression (e.g., obtain the most product of interest per kilogram of insect) as well as have visual biomarkers (e.g., white eye or EGFP phenotype maintained through breeding selection) for simple genotype maintenance in farms and laboratories.

Figure 5. Design of CRISPR target in *A. domesticus*. (**a**) Schematic of the annotated Ad vermilion gene from *A. domesticus* genome data. Red box: ATG starting code; Yellow boxes: exons; Purple arrow: sgRNA site. Note: in this figure intron sizes are not proportional to exons; (**b**) Ad muscle actin gene with promoter areas. Red box: Sequences used for promoter prediction; Green sequences: predicted promoters; Red sequences: start codon; Yellow sequences: protein coding area; Gary blue area: promoter sequence used in CRISPR knock-in construct.

CRISPR knock-in/out experiments. Three sgRNA sites were identified as targets for our CRISPR knock-in/knock-out constructs (Table S8). The promoter of Ad muscle actin has many predicted promoters (Figure 5b). To keep the knock-in DNA construct size small, we used 455 bp of the putative promoter sequences upstream from the starting code as the promoter for the marker gene.

For knock-outs, we targeted Ad vermilion using CRISPR/Cas9. Using this target sequence, we designed and used either (1) a combination of 3 different sgRNAs or (2) a single sgRNA (sgRNA#1) (Table 3). The resulting G_0 hatchling crickets that received all three sgRNAs had 68% eye color knock-out phenotype while 32% of those receiving only sgRNA#1 had the phenotype. This phenotype resulted in eye color change from black (wild type) to vermilion or white eye color (Figure 6A,B). Our results show the knock-out color appears white in freshly hatched crickets and gradually changes to vermilion color after a few hours. In each set of injection experiments, both partial and complete knock-out of eye color phenotypes were found (Figure 6B). Partial eye color knock-out was mosaic. In comparison, the eye color in wild-type crickets were always black even in freshly hatched individuals (Figure 6A). To generate stable lines, groups of positive G_0's were subsequently self-crossed. For self-crosses, all 16 crosses and 9 out of 10 crosses from three

and one sgRNA treatments, respectively, provided vermilion eye color G_1s, indicating the CRISPR/Cas9 knock-out rate was particularly high in *A. domesticus*. To check the knock-out sequences, we tested five of the knock-out strains (G_2) from the single sgRNA treatment. The sequencing results indicated deletions ranging from just a few base pairs to more than 240 bp among these five strains (Table 4).

Table 3. CRISPR knock-out microinjections and knock-out rate for Ad vermilion gene in *A. domesticus*.

sgRNA	Eggs Injected	Hatched	Hatch Rate	KO G_0s *	# of Crosses	KO G_1 **
#1,2,3	1737	222	13%	150	16	16
#1	2065	433	21%	139	10	9

* Knock-out positive G_0s. ** Number of crosses provide Knock-out positive G_1s.
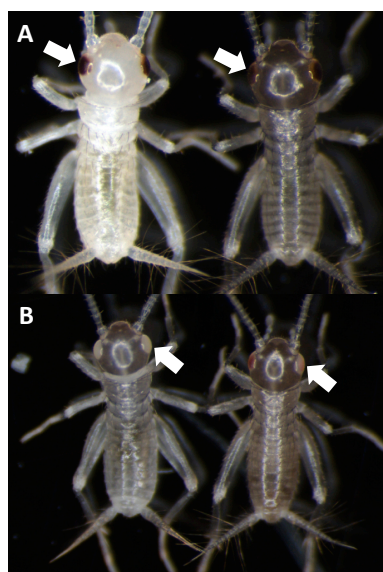
Figure 6. Vermilion eye color phenotype in *A. domesticus*. (**A**) wild-type eye color in freshly hatched cricket (left) and few hours old cricket (right). (**B**) Eye color of completely knocked-out Ad vermilion gene (left) and cricket with partial knock-out phenotype (right). The white arrows point out the location of the cricket eyes where the respective phenotypes can be observed.

Table 4. Deletions from CRISPR knock-out strains in *A. domesticus*.

| Strain	Sequence	Deletion (bp)
AA seq	G G A G R S G R R G A G R	
Wild Type	ATGCTGGCGTGTCGCAGGGAGGCGCCGGGCGCTCAGGA CGG CGAGGTGCTGGACGG	-
V2-1	--------------GGCGCTCAGGACGGCGAGGTGCTGGACGG	243
V2-3	------------------GGACGGCGAGGTGCTGGACGG	238
V2-6	ATGCTGGCGTGTCGCAGGGAGGCGCCGGGCGCT----CGAGGTGCTGGACGG	8
V2-7	ATGCTGGCGTGTCGCAGGGAGG-------GGACGGCGAGGTGCTGGACGG	13
V2-7-2	ATGCTGGCGTGTCGCAGGGAGG-GC GA GG T GCTCA C GACGGCGAGGTGCTGGACGG	14 *

* Miss match sequence for 14 bp.

For knock-in experiments, sgRNA#1 and the knock-in construct containing the enhanced green fluorescent protein (EGFP) marker gene coding region (Figure 7) were co-injected with Cas9 protein into young *A. domesticus* eggs. We injected 939 eggs, and 255 crickets successfully hatched (hatch rate of 27%). From all hatched crickets, we obtained six G_0 crickets showing EGFP expression (2%) (Figure 8). The knock-in positive phenotype selected was based on EGFP expression in muscle tissue due to use of the Ad muscle actin promoter. EGFP expression varied from a small group of muscles to more than half of the muscle in the body (Figure 8D,E). The enlarged picture shows a

more detailed view of the muscle structure with GFP expression in part of the cricket leg (Figure 8F). These results indicate that the promoter is functional and successfully expressing the marker gene in the appropriate tissue. From the G_0 EGFP positives, we were able to set up three out-crosses (with wild-type crickets) and one G_0 EGFP negative cross (using injected hatchlings that did not show EGFP phenotype) to screen G_1 offspring. From these experiments, we received EGFP positive G1s from two of the positive G0 out-crosses (#1 and #3) and were able to establish two muscle EGFP expression colonies (Figure 9). In Figure 9A,B, we can see that the EGFP expression patterns from #1 and #3 colonies are different. Crickets from colony #1 showed much lower EGFP expression and were only easy to screen from hind leg muscle, but most muscles in the body of crickets from colony #3 were expressing EGFP. There were no EGFP positive G_1 offspring from the negative G0 crosses.
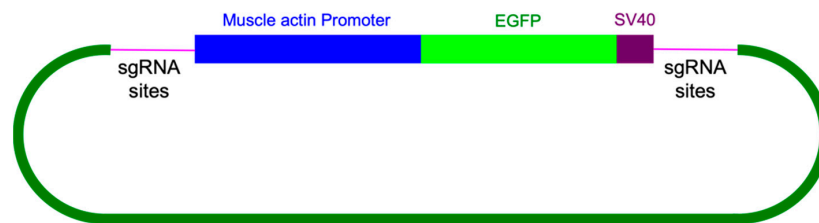
Figure 7. CRISPR knock-in construct in *A. domesticus*. Ad muscle actin was used to identify the promoter driving EGFP as the marker gene. Two short sequences contain the Ad vermilion sgRNA sites flanking both side of the marker gene.
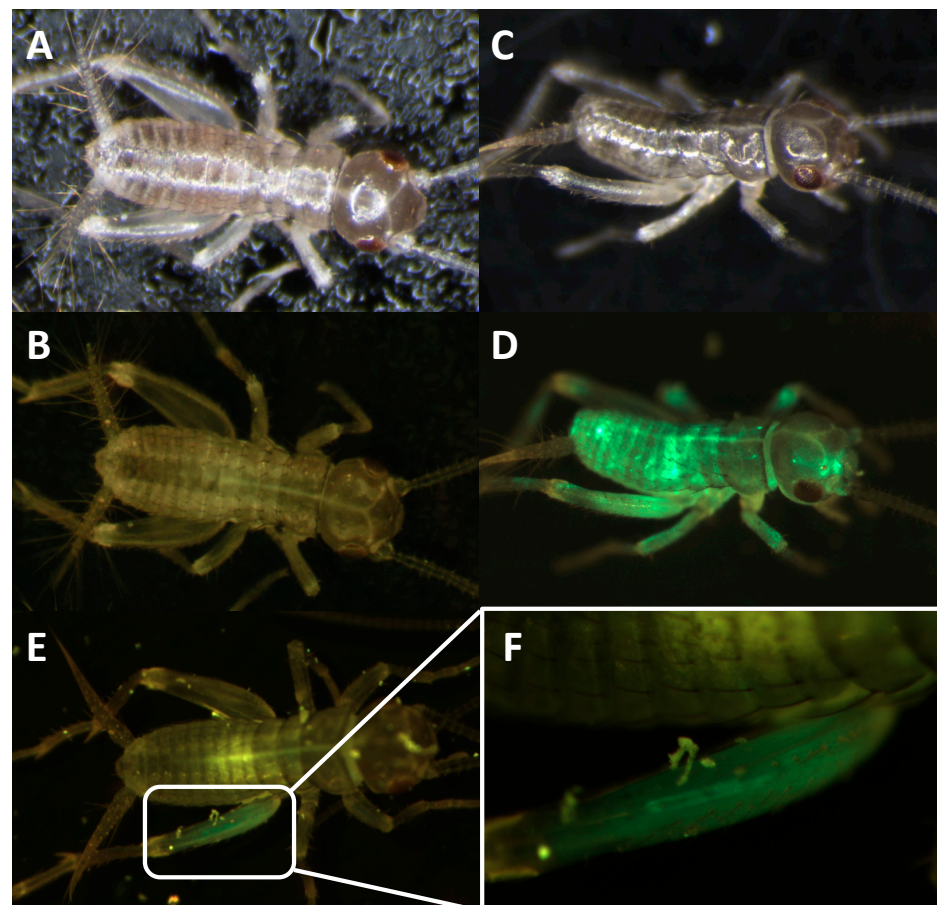
Figure 8. Somatic knock-in EGFP in *A. domesticus* by CRISPR/Cas9. Wild-type cricket under white light (**A**) and florescent light with GFP filter (**B**). Successful knock-in cricket with large area of somatic knock-in under white light (**C**) and florescent light with GFP filter (**D**). Cricket with small area GFP knock-in phenotype, (**E**) and enlarged picture of muscle structure GFP expression (**F**).
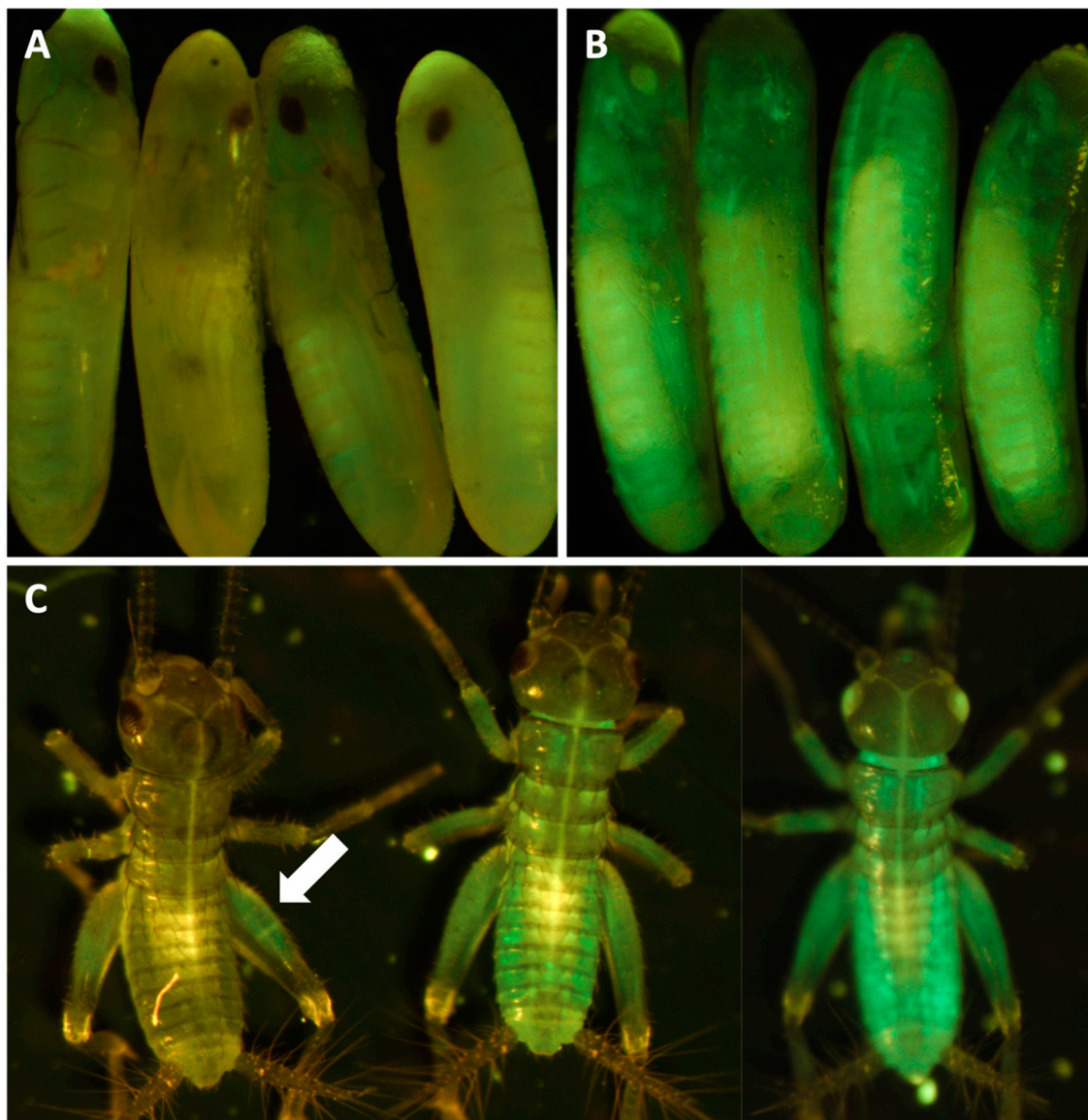
Figure 9. G_1 *A. domesticus* with GFP expression. (**A**) #1 G_1 eggs with some GFP expression. (**B**) #3 G_2 eggs with GFP expression. (**C**) GFP crickets from #1 (left) and #3 heterozygous and homozygous (middle and right) crickets. The arrow indicates EGFP expression in the cricket leg muscle. Note: Homozygous cricket with white eye phenotype.

RNAi experiments. In addition to demonstrating CRISPR-Cas9 knock-out and knock-in efficacy in *A. domesticus* based on our genomic data, we also demonstrated efficacy of RNAi in this species using the same data. Two different target genes were used: (1) Ad vermilion gene (in wild type crickets) and (2) the EGFP gene from CRISPR-Cas9 stable EGFP expressing lines. For each of the two strains, we injected three sets of crickets with each dsRNA (Table 5). The results had survival rates four weeks post injection of approximately 70% for RNAi constructs and 51% for controls. Based on phenotype screening, the average KD (Knock Down) rate of either white eye or reduced EGFP expression was approximately 86% (e.g., 70 successful knock down phenotypes divided by 81 total injected survivors) (Table 5). The observed phenotypes were visible at least two weeks post injection. The eye color knock-down phenotype was not as strong as the CRISPR knock-out but was sufficient for screening (Figure 10A). The GFP RNAi knock-down experiments were similar (Figure 10B). We continued screening both phenotypes for five weeks after injection and

documented continuation of a reduction of both eye color pigmentation and GFP expression between two- and four-weeks post injection.

Table 5. RNAi results in *A. domesticus*.

Name	Injected	Survival *	Survival Rate	KD	KD Rate
dsAdV-WT #1	19	16	84%	13	81%
dsAdV-WT #2	19	17	89%	13	76%
dsAdV-WT #3	20	10	50%	8	80%
dsEGFP-WT #1	25	13	52%	0	0%
dsEGFP-WT #2	18	9	50%	0	0%
dsEGFP-AdVELV1-3 #1	20	12	60%	12	100%
dsEGFP-AdVELV1-3 #2	20	17	85%	15	88%
dsEGFP-AdVELV1-3 #2	17	9	53%	9	100%

* Survival rate is count at four weeks post microinjection.
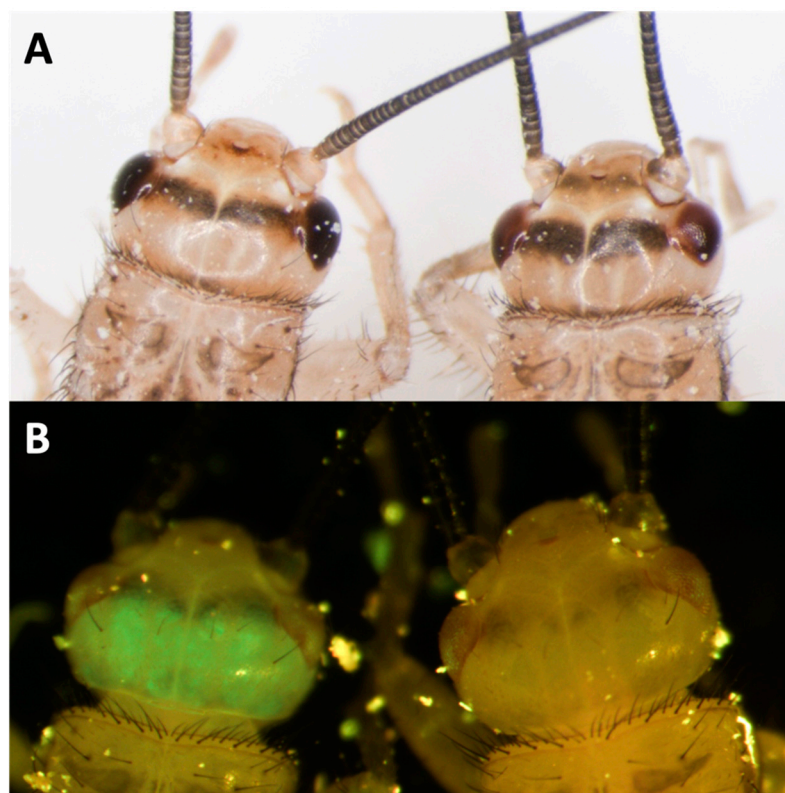
Figure 10. *A. domesticus* RNAi phenotypes. (**A**) Ad vermilion gene knock-down by RNAi in house cricket. Wild type eye color (left) compared with RNAi eye color (right). (**B**) GFP knock-down by RNAi in our EGFP expressing house cricket strain, AdVELV1-3. Regular EGFP expression (left) compared with dsEGFP RNAi cricket (right).

3. Discussion

Insects are a dense source of dietary nutrients [57,58]. Research demonstrates other health benefits of consuming insects including biomarkers for improved gut health in humans [59]. Zoonotic diseases, such as viruses, are much less likely to jump from farm-raised insects to humans than from mammals or bird livestock due to large genetic differences between humans and arthropods [23,35,60]. Additionally, studies show that foodborne pathogen loads in farmed insects, such as *Salmonella* spp. and *Listeria monocytogenes*, are low and often absent [61,62]. The European Union recently approved the focus species

of this paper, *A. domesticus*, as well as *Tenebrio molitor* (yellow mealworm) as a safe novel food [63].

Given the excellent nutritional value of insects and their ability to be grown efficiently in small and confined spaces with minimal resource inputs, they will be a critically important solution for sustainable food/protein production [23,24,32,64,65]. Entomophagy, consumption of insects as food, is accepted and practiced by over 2 billion people worldwide [34,66,67]. Some attractive features offered by insects include low land, water, and natural resource utilization [64], reduced greenhouse gas emissions [65,68], highly prolific (1500–3000 eggs per female) [64,69,70] and short life cycles for rapid scale-up to support food security. Insects are amenable to modularized vertical farming as well as automated and urban farming. They also can be grown closer to cities and/or processing facilities. Due to their small size and growth efficiencies, insects are likely the only animal feasible for production in space exploration, such as at stations on the moon and Mars [71,72].

Diversification of our food supply is critical for food security. According to the FAO, 75% of food comes from only 12 plant and 5 animal species [10,73]. Class Insecta are the largest, most diverse group of organisms on Earth, with approximately 950,000 species described and 4–30 million estimated [16,74,75]. At least 2000 species have been identified as eaten around the world, and many can be farmed [23,24,33,34,67]. One of the food security benefits of the biodiversity offered by insects is the ability to rapidly switch species in response to crop loss from disease [23,28].

In addition to being a food source for animals and humans, crickets have also been utilized as a valuable model system for studying biological processes such as neurobiology, developmental biology, animal behavior, and others [76]. With the recent interest in industrial farming crickets for human consumption, even greater potential exists for studying these creatures via expanded access to all life stages from farms, as well as the potential impact of research on cricket biology and genetics.

3.1. Genome Assembly

The *A. domesticus* genome assembly represents a contiguous assembly for downstream applications. The assembly contains 11 large scaffolds, presumably representing most of the autosomal and X chromosomes, in 2.138 Gb, close to the flow-cytometry predicted genome size of 2.150 Gb for male *A. domesticus*. There were 29,304 predicted genes, mostly concentrated on the large chromosomes. These data will be important for the insect food industries, but also for orthopteran studies in biology and evolution.

Immediately after obtaining the genome assembly, we looked for promoters for genes for use in our transgenic experiments but encountered genes with very long sequences that were distinctly longer than the coleopteran genomes we have studied. A quick comparison of the length of the BUSCO reference gene sequences estimated that the *A. domesticus* genes were on average 80% longer than those in *T castaneum*. We also noticed that, in most cases, the longer genes were due to longer introns that may be due to TEs as has been observed in other Orthoptera [53,54]. Therefore, we compared repetitive elements in the *A. domesticus* genome assembly with those available for other orthopterans and found that the *A. domesticus* genome assembly contains the highest content of TEs (almost half of the assembly) compared to other cricket genomes. Repetitive elements and genome size in Orthoperta were significantly correlated if we included the very large genome of *L. migratoria* ($p = 0.024$). When *L. migratoria* data was excluded, LTR elements and DNA transposon were positively correlated to the expansion of genome size in crickets ($p < 0.001$). As we suspected, repetitive element and median length of intron were positively correlated ($p = 0.044$) but only when *L. migratoria* was included, and especially for elements LINE/CR1, LINE/L1, LINE/Penelope, LINE/RTE, DNA/hAT, DNA/Kolobok, and DNA/TcMar. Our data suggest that LINE and DNA TEs contribute to variation in genome size and corresponding intron length in Orthoptera, but this is likely found only at the suborder level. Furthermore, the TE distribution pattern of LTR transposons shows a recent or

ongoing burst in the *A. domesticus* genome and likely contribute to the high abundance of repetitive elements in the genome of this species.

3.2. A. domesticus Metagenome

Of the 16,290 original scaffolds in the *A. domesticus* assembly, 6246 scaffolds were removed and submitted to NCBI as metagenome data. Most of these metagenome scaffolds had similarity to IIV6, but we were unable to assemble these scaffolds into a complete IIV6 sequence. Closer examination of the predicted proteins from the virus scaffolds indicated that only 32 scaffolds were high to medium quality viral genome sequences. Therefore, scaffolds with similarity to virus sequences at the nucleotide level may be sequencing artifacts, or some may represent novel viruses that have not been described in current databases. Importantly, our analysis did not identify common foodborne pathogens, such as *Salmonella* sp., *Listeria* sp., *E. coli*, or *Staphylococcus* sp.

The question of crickets harboring covert virus infections was recently addressed [77]. In that study, Densovirus sequences were found in sick and healthy insects, and the authors speculated that difference in immunity may prevent active diseases in healthy colonies. Our insects came from an insect farm with no reports of insect disease. We did find scaffolds similar to *Wolbachia* symbionts in other insects. *Wolbachia* has been reported to have a protective effect against viruses in some insects [78]. Further, *Wolbachia* was reported in a previous study of microbial communities in *A. domesticus* [79]. More research is needed to understand the prevalence of iridoviruses in farmed insects, as well as the mechanisms that prevent active disease in infected but healthy individuals. Our goal is to engineer virus resistant crickets through CRISPR genetic engineering technology to improve the health of farmed insects. This is an important reason we seek to understand genes related to immunity in these insects.

3.3. Immunity-Related Genes and Antimicrobial Peptides

We annotated genes related to immunity in *A. domesticus* to better understand how these crickets respond to infection, important in insects being reared in close confines for animal feed and human consumption. Insects depend almost exclusively on an innate peptide and antimicrobial secondary metabolite forms of immunity against infections/pathogens [16,80]. The current study did not identify analogs of classical canonical small insect antimicrobial peptides (AMPs), such as cecropins or defensins, in the *A. domesticus* genome. AMPs are found more in holometabolous than hemimetabolous species [81], perhaps due to sampling (i.e., originally identified in insect pupae/larvae). However, this may also be due to the poor mobility and frequent terrestrial lifestyle of insect larvae vs. nymphs (e.g., grubs/caterpillars vs. cricket/grasshopper nymphs), thus exposing holometabolous insects to a wider variety of microbial pathogens for longer periods of their life [81]. Insect immune-related proteins have activity against microbes, including *Plasmodium* sp. [82]. However, as with much of Class Insecta, insufficient research is available to identify biological activities and potential practical applications for these proteins [16]. Given their abundance in insects and the growing scale of the insect production industry, insect antibacterial, antifungal, antiparasitic, and antiviral proteins also could be a valuable and low-cost bioresource or byproduct for future biomedical applications. Many studies have identified numerous AMPs from insects with a wide variety of efficacies against various pathogens and other potentially valuable biological activities [83,84]. Thus, these proteins represent an untapped resource for potential therapeutic applications.

Instead of AMPs, we found other conserved genes in the *A. domesticus* genome assembly that may protect it from pathogens, including PGRP, GNBP, lysozymes, and phenoloxidase. We reverted to our previous transcriptome data [22] to understand how these genes are expressed in different life stages or in male vs. female adults. Two PGRP genes are increasingly expressed during early nymphal development and are more highly expressed in female than male *A. domesticus*. Eleven genes were annotated as GNBP, but only one gene, ANN16377-RA, was more highly expressed in nymphs and adults. Lysozymes and

PPO are important in resistance to disease in crickets [85]. Overall, lysozyme genes were expressed higher in 1- and 2-week-old nymphs, and male and female adults, and PPO genes were expressed mostly throughout life stages. However, two PPO genes, ANN15897-RA and ANN15899-RA, were more highly expressed in newly hatched larvae and thus at a time the cricket would be upregulating immune defenses. These data will be important in developing protocols for healthy farm-reared insects, both in diagnostics and prophylatics for disease control.

Additionally, when considering the impact of insect genetic engineering, such as work derived from this publication, insects may be an ideal host for expression and mass production of exogenous antimicrobial peptides from various species (e.g.,: crickets or mealworms may be useful as bioreactors to produce a particularly active/valuable peptide compared to other insects that may be more challenging or costly to mass produce). As such, applications are identified; given large insect biological diversity, as well as an established insect mass production industry, a massive and low-cost resource is now available for production of compounds for specific applications.

3.4. Genetic Engineering and Gene Editing via CRISPR

The primary purpose of this work was to sequence the *A. domesticus* genome to obtain necessary sequence data for genetic engineering to improve these animals as a sustainable food resource and for other forms of bioproduction. Thus, we identified and annotated two genes, Ad vermilion and Ad muscle actin, which were used to create knock-out and knock-in strains of *A. domesticus* as an initial proof of concept to develop the tools required for our downstream technologies. The analysis of Ad vermilion was our first indication that the genome of this species contains quite large introns. The exon sizes were similar to those of other orthologous insect genes such as vermilion from *T. castaneum*, but due to the extended intron length, this gene was more than 100 times longer in *A. domesticus*. The *T. castaneum* vermilion gene is around 2 kb but is around 90 kb in Ad vermilion. On the other hand, the annotation of the Ad muscle actin gene was more straightforward. As with *Tribolium* sp. and *Drosophila* sp., Ad muscle actin has only one exon and, therefore, is similar in size to that of other insect species. Once we had these genes annotated, we were then able to use the Ad vermilion gene to identify CRISPR target sites, and the promoter region from Ad muscle actin gene was harnessed for high level expression of knock-in marker genes. The Ad vermilion gene, when disrupted as a target for knock-in experiments, also functions as a convenient biomarker, as reducing or eliminating function of this gene results in a white-eye (or vermilion color eye) phenotype with less eye pigmentation than wild type. Thus, for the CRISPR genetic engineering work, we showed that the CRISPR/Cas9 gene editing tool works in *A. domesticus*. With up to a 68% CRISPR knock-out rate in G_0, the efficiency was sufficient to produce and establish the modified strains for research and future commercial applications. After multiple generations in the laboratory, the vermilion mutation in *A. domesticus* colonies does not demonstrate any obvious decrease in phenotype or fitness based on survival and growth rate (data not shown). For our knock-in CRISPR experiments, we used the same sgRNA target site as for the knock-outs. As such, we expected the Cas9-sgRNA-created DNA double break in both the insect genome and the knock-in plasmid. Therefore, the EGFP marker gene was likely incorporated into the genome through the non-homologous end joining (NHEJ) DNA repair process. The Ad muscle actin promoter was successfully utilized to express exogenous EGFP in muscle tissue. The G_0 somatic knock-in rate was around 2%, and two out of three outcrosses provided EGFP positive G_1, indicating the knock-in is efficient in those G_0. As this was our first attempt, an optimized protocol will be created in the future to further improve this knock-in rate. Interestingly, the EGFP colonies had different muscle EGFP expression patterns, possibly as a result of a damaged promoter during knock-in since we used NHEJ to create the knock-in or due to epigenetic factors. Based on these results, we successfully showed CRISPR gene editing works to knock-in genes and function in house cricket and creates possibilities for valuable applications going forward.

3.5. RNAi in A. domesticus

For the RNAi experiments, abdominal microinjections of young *A. domesticus* with target gene dsRNAs resulted in either eye color reduced phenotype or whole body GFP knock-down, depending on the construct and strain injected. Therefore, *A. domesticus* has the necessary genes for systematic RNAi, and they are functional, as was previously described in this insect [86] and in the *Allonemobius* socius complex of crickets [87]. From experience with other insect species, we expected phenotypic changes around 48 h post injection, but since crickets regain their eye color every time they molt, we did not see the eye color change until their next molt after dsRNA microinjections. Depending on the timing, the eye color was sometimes more reduced after the next two to three molts. However, since RNAi can only reduce existing gene expression, it is likely eye color will gradually be restored through time after each molt. The RNAi eye color phenotype was as obvious as CRISPR knock-out. EGFP knock-down also showed visible phenotype (reduced EGFP expression) after two weeks. We believe this slow response compared with other insects may be because the EGFP protein is more stable, so it takes longer for the pre-experiment protein to be degraded. For RNAi experiments, we observed a lower survival rate in some of the microinjection groups, likely due to non-optimal rearing conditions such as small observation containers that were not large enough for more than 2 molts/live stage, which will be optimized in future RNAi experiments.

4. Conclusions

This work provides the foundation for the potential of genome editing applications in farm raised *A. domesticus*, such as disease resistance and alteration of the nutrient profile for production of numerous bioproducts, using the low cost, low-tech farm raised insect bioreactor system. Insects are already highly efficient and sustainable compared to other livestock and protein production systems. Gene editing technology will produce insects with higher growth rate, lower mortality due to disease and other factors, and more nutrient dense that will in turn make insects more efficient and sustainable.

The assembly also will be a resource for basic research, as crickets are a valuable model system for studying biological processes such as neurobiology, developmental biology, animal behavior, and others [76]. With the recent interest in industrial farming crickets for human consumption, even greater potential exists for studying these creatures via expanded access to all life stages from farms, as well as the potential impact of research on cricket biology and genetics.

Many opportunities exist for insects as an untapped resource for applications like pharmaceuticals, including antimicrobials and antiviral substances, drug lead compounds, material for bioprospecting, enzymes, bioactive peptides, oils, biocompatable materials for wound healing and other medical applications, food waste mitigation, nutrient cycling, biodegradable plastics and packaging, new materials from chitin with novel properties, and durability [16,23]. However, little bioprospecting of insects has been done to date compared with other taxa.

Developing the tools for insect genetic engineering, including high quality reference genomes, provides an open-ended opportunity to use insects for purposes besides mere sources of food, protein, and dietary nutrients. Beyond the substances that insects contain naturally, genetic engineering provides additional potential for efficient, low-cost bioproduction of vaccines, enzymes, antibiotics, antimicrobial peptides, antibodies, color pigments and dyes, flavors, fragrances, functional ingredients, and many others. We believe this foundational research will play a critical role in reducing human environmental impact by utilizing the largest, most diverse, and yet almost entirely untapped biological resource on Earth: Class Insecta.

5. Materials and Methods

A. domesticus colony maintenance. *A domesticus* were originally purchased from a US insect farm in 2018 and maintained as a lab colony for multiple generations. They were fed

a diet of specially formulated cricket feed from LoneStar (TFP Nutrition) (Nacogdoches, TX, USA) (http://lonestarfeed.com), which was the same feed used by the insect farm. Feed and water were provided to each cricket cage twice weekly. Eggs were collected twice per week from wild type adults in a petri dish (150 mm diameter × 10 mm deep) filled with hydrated (deionized water saturated) polyacrylamide water crystals, as crickets prefer to lay eggs in moist areas. Each week, the old egg-lay dish was replaced with a new dish on Monday and Friday. Eggs were collected each Tuesday for colony maintenance. For egg collection, the petri dish with water crystals containing eggs was washed into a glass beaker using deionized water. Additional water was added to the beaker and stirred a few times, and eggs were allowed to settle to the bottom for approximately 10–15 s. Most of the excess water and crystals were slowly removed to retain eggs. Washing was repeated 2–3 times, and a plastic Pasteur pipette was used to aspirate eggs from the bottom of the beaker and transfer to a piece of damp/moist paper towel in a petri dish lid. The lid was placed into an 8 oz Deli container (S-21216, Uline, Pleasant Prairie, WI, USA) with a lid until the eggs hatched (approximately 8–10 days). After hatching, crickets were transferred into cricket cages. Cricket cages consisted of 20 qt plastic storage containers (Gasket Box, Sterilite, Townsend, MA, USA) with four 100 mm circles cut on the side and two on the lid that were covered by screen mesh material.

Cytometric genome size estimation of *A. domesticus*. The 1C (haploid) genome size of males and females of *A. domesticus* was estimated as described in Johnston et al. (2019) [49]. In brief, a single *A. domesticus* head, a single *female* head (1C = 328 Mbp), and a portion of brain from a *Periplaneta americana* male (1C = 3300) were placed together into 1 mL of ice-cold Galbraith buffer in a 2 mL Dounce tissue grinder. Nuclei were released by grinding with 10 strokes of an A (loose) pestle, then filtered through nylon mesh into a 1.5 mL microfuge tube. The DNA in the released nuclei were stained by adding 25 μL of propidium iodide (1 mg/mL) and allowed to stain 3 h in the dark at 4 °C. The mean red PI fluorescence of the stained DNA in the nuclei of the sample and the standards was quantified using a CytoFlex flow cytometer (Beckman Coulter). Haploid (1C) DNA quantity was calculated as (2C sample mean fluorescence/2C standard mean fluorescence) times 328 Mbp for the *D. virilis* standard and times 3300 Mbp for the *P. americana* standard. The estimates based on the two standards were averaged to produce a 1C estimate for each sample.

Genome sequencing and assembly. A single adult male *A. domesticus* was shipped to Dovetail Genomics (now Cantata Bio, Scotts Valley, CA, USA). Genomic DNA was extracted and shipped to North Carolina State University and USDA ARS, Stoneville, MS for long read sequencing, and extracted nuclei were used for Chicago and Hi-C libraries.

Long read data was obtained using Pacific Biosciences technologies (PacBio, Menlo Park, CA, USA). Libraries were prepared using SMRTbell Express Template Prep Kit, and Sequel Binding Kit 2.1 was used for polymerase binding. The binding reaction was loaded at 8 pmol concentration with a 15 h run time. The gDNA was sequenced on eight LR SMRTCells 1M v3 for the Sequel I for a total of 58.7 Gb of long read data. The genome was assembled from PacBio reads using CANU v1.8 [88]. The two sets of PacBio CLR reads were corrected separately and co-assembled. One set was iteratively corrected with the command canu -correct -p asm -d set1_round1 genomeSize = 2.5 g corMinCoverage = 0 corMhapSensitivity = high corOutCoverage = 100 -pacbio-raw set1/*.subreads.fastq.gz with the output of each round being the input for the next round. The reads were trimmed after three rounds of correction with the command canu trim -p asm -d set1_round3_trim genomeSize = 2.5 g correctedErrorRate = 0.105 -pacbio-corrected set1_round3/asm.correctedReads.fasta.gz. The second set of reads was corrected with the command canu -correct -p asm -d set2_round1 genomeSize = 2.5 g corMinCoverage = 0 corMhapSensitivity = normal corOutCoverage = 100 -pacbio-raw set2/*subreads.fastq.gz and trimmed as set1. Lastly, the combined corrected reads were assembled with the command canu -assemble -p asm -d set1+set2 'genomeSize = 2.5 g' 'correctedErrorRate = 0.105' 'batOptions = -dg 3 -db 3 -dr 1 -ca 500 -cp 50' -pacbio-corrected set2_round1_trim/asm.trimmedReads.fasta.gz -pacbio-corrected set1_round3_trim/

asm.trimmedReads.fasta.gz. This final assembly was polished with long read data in Arrow (SMRT Link 3.1.1, PacBio).

A Chicago library was prepared as described previously [89]. Briefly, ~500 ng of HMW gDNA was reconstituted into chromatin in vitro and fixed with formaldehyde. Fixed chromatin was digested with DpnII, the 5′ overhangs filled in with biotinylated nucleotides, and then, free blunt ends were ligated. After ligation, crosslinks were reversed and the DNA purified from protein. Purified DNA was treated to remove biotin that was not internal to ligated fragments. The DNA was then sheared to ~350 bp mean fragment size and sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The libraries were sequenced on an Illumina HiSeq X to produce 208 million 2 × 150 bp paired end reads. A Dovetail HiC library was prepared in a similar manner as described previously [90]. Briefly, for each library, chromatin was fixed in place with formaldehyde in the nucleus and then extracted Fixed chromatin was digested with DpnII, the 5′ overhangs filled in with biotinylated nucleotides, and then free blunt ends were ligated. After ligation, crosslinks were reversed, and the DNA was purified from protein. Purified DNA was treated to remove biotin that was not internal to ligated fragments. The DNA was then sheared to ~350 bp mean fragment size and sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The libraries were sequenced on an Illumina HiSeq X to produce 143 million 2 × 150 bp paired end reads.

The input de novo assembly, Chicago library reads, and Dovetail Hi-C library reads were used as input data for HiRise, a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies [89]. An iterative analysis was conducted. First, Chicago library sequences were aligned to the draft input assembly using a modified SNAP read mapper [91]. The separations of Chicago read pairs mapped within draft scaffolds were analyzed with HiRise to produce a likelihood model for genomic distance between read pairs, and the model was used to identify and break putative misjoins to score prospective joins and make joins above a threshold. After aligning and scaffolding Chicago data, Dovetail HiC library sequences were aligned and scaffolded following the same method. Scaffolds that were determined to be "contaminants" (mitochondrial, duplicate, vector, or microbial as described in the next section) were removed prior to submission, resulting in 9961 final scaffolds and a total length of 2,346,604,983 bp. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JAHLJT000000000. The version described in this paper is version JAHLJT010000000.

To determine scaffolds that were from the metagenome, we used multiple tools to screen the scaffolds, including BBSketch and NCBI Refseq (https://www.biostars.org/p/234837/) and Kraken2 [92] in Omicsboxes (Biobam, Valencia, Spain, version 2.1.2, database RefSeq 2021.04, 2022.02). The scaffolds also were used in beta testing for GX, a screening tool for microbial contamination at NCBI (https://github.com/ncbi/fcs). We combined results from all three datasets (BBtools, Kraken2, GX) and manually annotated unclassified scaffolds with blastn [50], further using read depth to assign as insect or microbial. After this process, 80 scaffolds were removed as a result of NCBI screening tools to identify vector or duplicates, and 6245 metagenomic scaffolds were submitted to NCBI MIMS as a metagenome/environmental, host-associated sample (accession number PRJNA908265).

Annotation and analysis. Annotation of the *A. domesticus* genome assembly was by Dovetail Genomics. Repeat families found in the genome assemblies of *A. domesticus* were identified de novo and classified using RepeatModeler (version 2.0.1) [93]. Repeat-Modeler depends on RECON (version 1.08) and RepeatScout (version 1.0.6) for the de novo identification of repeats within the genome. The custom repeat library obtained from RepeatModeler was used to discover, identify, and mask the repeats in the assembly file using RepeatMasker (version 4.1.0) [94]. Coding sequences from *Locusta migratoria*, *Teleogryllus occipitalis*, *Laupala kohalensis*, and *Xenocatantops brachycerus* were used to train

the initial ab initio model for *A. domesticus* using AUGUSTUS (version 2.5.5) [95], with six rounds of prediction optimization. The same coding sequences also were used to train a separate ab initio model for *A. domesticus* using SNAP (version 2006-07-28) [89]. RNAseq reads were mapped to the genome using STAR aligner (version 2.7) [96]. and intron hints were generated with the bam2hints tools within AUGUSTUS. MAKER and SNAP (intron-exon boundary hints provided from RNA-Seq) were then used to predict genes in the repeat-masked reference genome. Swiss-Prot peptide sequences from the UniProt database were downloaded and used in conjunction with the protein sequences from the same training species to generate peptide evidence in Maker [97]. Only genes that were predicted by both SNAP and AUGUSTUS were retained in the final gene sets. AED scores were generated for each of the predicted genes as part of the MAKER pipeline to assess the quality of the gene prediction. Genes were further characterized for putative function by performing a BLAST [50] search of the peptide sequences against the UniProt database. tRNA were predicted using the software tRNAscan-SE (version 2.05) [98]. Predicted genes were analyzed by BUSCO (v.2.0) [99], using the lineage dataset insecta_odb9 (creation date: 2016-10-21, number of species: 42, number of BUSCOs: 1658). Predicted genes and sequence information are in supplemental files (Files S9 and S10).

We annotated repetitive elements against the publicly available genomes consisting of six crickets and one locust. The sources of the genomes used in the analysis were: *Apteronemobius asahinai*, GCA_019974035.1; *Gryllus bimaculatus*, GCA_017312745.1; *Laupala kohalensis*, GCA_002313205.1; *Teleogryllus oceanicus*, www.chirpbase.org, *Teleogryllus occipitalis*, GCA_011170035.1; *Locusta migratoria*, GCA_000516895.1. Libraries of repetitive elements were compiled by RepeatModeler-2.0.3 [93] for each species. Resulting libraries were classified using reference-based similarity search in RepBase (20170127 edition). RepeatMasker-4.1.2-p1 [94] was used with alignment option '-a' to annotate the repetitive elements in the genome assemblies. The resulting alignment files were processed by the scripts with RepeatMasker pipeline and our in-house scripts (https://github.com/kataksk/repeatlandscape_parser.git). To analyze correlation between the repetitive elements and intron lengths, R packages GenomicFeatures [100] was used to obtain gene structures of gene models.

To extract virus sequences from the genome, blastn of potential viral scaffolds were compared to *Chilo iridescent* virus (GenBank: AF303741.1). Quality assessment of the extracted virus-like scaffolds was performed by CheckV (v1.0.1) [51].

CRISPR target gene for knock-out/in. We chose the 5′ of the Ad vermilion gene as our CRISPR target site for knock-in experiments. We identified the Ad vermilion gene from the *A. domesticus* genome assembly using a transcriptome mRNA sequence (not complete, missing 3′) by Blastation (TM Software, Inc., Arcadia, CA, USA). We were able to identify eight exons, including the first exon (Figure 5a). We then used the second exon (around 150 bp) to select three target CRISPR targeting sites for designing sgRNAs (Table S8) using GGP sgRNA designer (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design). The target sgRNA sequences were ordered from Synthego (Redwood City, CA, USA). To verify the knock-out sequence in *A. domesticus*, genomic DNA extraction was using a ZymoResearch Quick-DNA insect miniprep kit (Irvine, CA, USA), forward primer "GAGCAGTAGGCGAGAAAG" and reverse primer "TCCAAACGCAGAAGAGACCA" to amplify gDNA. After PCR, we used primer "TGTAGTGAGTGTTATCGCCA" to sequence the PCR product using the Oklahoma Medical Research Foundation sequencing service facility (OMRF DNA sequencing, https://omrf.org/). Using these data, we compared the gRNA sequence to wild type sequence to determine the indel(s) generated by our knock-out and knock-in experiments (Table 3).

Knock-in DNA construct design. For knock-in construct design, we used the Ad muscle actin gene promoter to express the marker gene, enhanced green florescent protein (EGPF) via CRISPR/Cas9. We utilized the muscle actin gene from *T. castaneum* and *D. melanogaster* to conduct a BLAST search of the *A. domesticus* transcriptome data and used the mRNA sequence from that search to identify the Ad muscle actin gene in the genome assembly. We searched for a promoter region 1.5 kb upstream from the start codon of the

Ad muscle actin genomic sequence using the Neural Network Promoter Prediction tool (https://fruitfly.org/seq_tools/promoter.html).

To design the DNA construct, 455 bp of the Ad muscle actin promoter sequence with its 5′UTR were placed upstream of the EGFP coding sequence along with a sv40 polyadenylation signal 3′ of EGFP as the marker gene. A section of 86 bp of gDNA included CRISPR knock-in site(s) placed on both sides of the marker gene as the final knock-in DNA construct, Ad-actin-EGFP-KI. This DNA sequence was synthesized by GeneScript (Piscataway, NJ, USA) in their pUC57 vector plasmid as the final product.

Microinjection solution preparation. The concentrations of the components of knock-out or knock-in microinjection solutions were: 100 ng/µL of DNA construct Ad-actin-EGFP-KI (only for knock-in treatment), 10 pmol/µL of sgRNA, 1 µg/µL of TrueCut Cas9 protein V2 (ThermoFisher, Waltham, MA, USA), and 20% phenol red buffer (Sigma Aldrich, St. Louis, MO, USA). The solution was mixed and incubated at room temperature for 5–10 min for Cas9-sgRNA binding, and the solution was kept on ice during microinjection. For knock-out experiments, we used either three sgRNAs or sgRNA#1 alone for microinjections (Table S8, sgRNA target sequences). For knock-in experiments, we only used sgRNA#1 in the microinjection solution.

A. domesticus egg microinjection protocol for CRISPR. An egg lay dish with hydrated water crystals was placed into an adult wild type *A. domesticus* colony for 4 h to receive fresh and non-developed eggs. Eggs were collected and washed using the protocol described above in the "*A. domesticus* Colony Maintenance" section. After washing the eggs from the water crystals, a solution of 70% ethanol (ethanol/deionized water) was used to briefly rinse the eggs for 10–15 s and quickly rinsed with deionized water at least 3 times. The eggs were placed on microinjection slides (Figure S12). To prepare the microinjection slide, a piece of black filter paper was cut into 50×10 mm pieces and laid on a standard clean glass microscope slide (75×26 mm). Another two pieces of black filter paper strips, 50×2 mm, were cut and laid one on the top of the other placed in the middle of the larger piece of black filter paper. Deionized water was added to the filter paper, and eggs were placed on the edge of two sides of the filter paper strips for microinjection. The wet filter prevented the eggs from drying out during microinjection and allowed a backstop to keep the eggs from moving during the procedure. The black paper made visualizing the eggs easier compared to using white filter paper. The needle for microinjection was pulled from Standard Glass Capillaries (1B100-4, World Precision Instruments) using a P-2000 needle puller (Sutter instruments). The following puller apparatus settings were used: Heat: 335 Fil: 4 Vel: 40 Del: 240 Pul: 120. For microinjection, 1 µL of microinjection solution was loaded into the back of the needle with Femtotips (Eppendorf, Hamburg, Germany), and all air bubbles were removed from the liquid in the needle, particularly the tip. The filled needle was then applied to a FemtoJet 5247 microinjector (Eppendorf) and air pressure of 4–7 psi was used for holding the solution in place and 9–10 psi for injection. Before injection, forceps were used to gently break the needle tips to achieve a small opening. The injection site for eggs was 2/3 from the smaller end. After injection, the black filter paper containing the injected eggs was transferred from the glass slides to a piece of wet paper towel and kept in a 150×10 mm petri dish with lid. The dish was placed into a hatch chamber (Modular Incubator Chamber, MIC-101, Billups-Rothenberg, San Diego, CA, USA) with two wet paper towels for 10–12 d.

Screening for CRISPR knock-out/in and established colonies. To screen for either white eye (Ad vermilion knock-out) or EGFP (EGFP knock-in) expression phenotypes, freshly hatched G_0 *A. domesticus* were transferred from their egg incubation chamber (described above) into a standard petri dish with lid. The petri dish was then placed on ice to immobilize for 3 min. The crickets where then observed under a fluorescent dissecting microscope with digital camera (Leica M125, Leica Microsystems Inc., Deerfield, IL, USA) for either different eye coloration from wild type in knock-out treatments (using a standard white LED ring light) or using a blue florescent light and GFP filter to check for knock-in EGFP expression phenotype. Photos were taken using a QImaging Retiga R6 digital

camera. After screening, G_0 crickets with positive knock-out white/vermilion eye (full or partial) phenotype were separated and reared to adulthood in containers described in the house cricket colony maintenance section above. Crickets with wild type eye phenotype from knock-out experiments were not kept. G_0 crickets from the knock-in experiments with EGFP positive and negative phenotypes were reared in separate containers until adulthood. Once crickets became adults, self-cross experiments were set up (two males with four to six females) groups for positive G_0s from the knock-out experiments. For knock-in experiments, small out-cross groups were set up using EGFP positive G_0s with two G_0s out-cross to 2–3 wildtype crickets, and one self-cross group from all EGFP negative G_0s pooled together in the same cage. Cross groups were maintained, and eggs were collected as described in the house cricket colony maintenance section above. Once the G_1 generations began to hatch from eggs, we then screened and separated all white eye color phenotype knock-out G_1's from self-cross groups, and EGFP positive G_1s from out-cross groups to establish separate colonies from all crosses. No EGFP positive G_1 crickets were found from the pooled EGFP negative G0 injected cricket colony.

RNAi target genes and dsRNA sequences. Short portions of the Ad vermilion mRNA gene sequence (679 bp) were used to design 450 bp (51–500 bp) dsRNA (dsAdV) (Table S11) to target the Ad vermilion gene for knock-down via RNAi. An additional dsRNA (dsEGFP) also was made using 450 bp of EGFP coding sequence (region 96–545 bp) and was used both to knock-down EGFP in our knock-in cricket lines via RNAi and a negative control for the RNAi experiments for knocking down Ad vermilion for white eye color via RNAi. Both dsRNA sequences were from Genolution Inc. (Seoul, Republic of Korea).

A. domesticus microinjection and screening for RNAi. The concentrations of components of the cricket RNAi microinjection solution were: 2.5 µg/uL for dsRNA (dsAdV or dsEGFP) with 20% phenol red buffer. Two separate experiments were utilized to demonstration of RNAi efficacy in *A. domesticus*; one using wild type crickets and the other using our genetically modified strain. The dsAdV was injected into wildtype crickets to knock down eye color. The dsEGFP was injected into wild type crickets both as a negative control and also in separate experiments to knock down EGFP expression in our AdVELV1-3 EGFP expression strain. Two to three sets of microinjections of each RNAi experiment were conducted using young nymphs (following their first molt after hatching from eggs). Crickets were placed on ice for 5 min to immobilize them and were subsequently injected in the posterior abdomen (Figure S12). Around 25 to 30 crickets were injected using a total of 4 µL of injection solution (approximately 0.15 µL per injection). Following injection, crickets were maintained in a medium sized cage (Kritter Keeper®, 4.80 × 7.40 × 5.60 in, Lee's Aquarium & Pet Products, San Marcos, CA, USA) with food and water. After microinjection, crickets were screened based on eye color or EGFP expression every week and phenotype changes were documented.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/biom13040589/s1, Table S1. Genome size of *A. domesticus*. 1C = the amount of DNA in a gamete (1C is an average of the gametes with and without the X in the male). Figure S2. Comparison of the lengths of BUSCO reference genes (bp) in *Tribolium castaneum* and *Acheta domesticus*. Figure S3. Metagenomic scaffolds from the *A. domesticus* genome assembly. Spreadsheet "Ado_metagenome_Kraken2") details the output from files tentatively identified as non-insect through the first screening; "Blastn_unclassified" is the output from blastn of the unclassified scaffolds to NCBInr; "Metagenome_summary" are the scaffolds submitted to NCBI as A. domesticus-associated sequences. Figure S4. Screen of viral genome sequences through CheckV [51]. Data is sorted according to high, low, or medium quality, or not-determined based on matches to the CheckV database. Figure S5. Immune-related transcripts that are predicted from the annotation of the *A. domesticus* assembly. Predicted transcripts include: PGRP—peptidoglycan-recognition protein; GNBP—b-1-3-glucan-binding protein; lysozyme; PPO—prophenyloxidase. Figure S6. Supporting data for the comparison of intron length, genome size, and repeat content in orthopteran species (six crickets, including *Acheta domesticus*, *Apteronemobius asahinai*, *Gryllus bimaculatus*, *Laupala kohalensis*, *Teleogryllus occipitalis*, *Teleogryllus oceanicus* and one locust, *Locusta migratoria*). (A) LINE, (B) DNA, and (C) LTR transposon subclasses and correlation with intron length and genome size. Table S7.

Comparison of intron length, genome size, and repeat content in orthopteran species (six crickets, including *Acheta domesticus*, *Apteronemobius asahinai*, *Gryllus bimaculatus*, *Laupala kohalensis*, *Teleogryllus occipitalis*, *Teleogryllus oceanicus* and one locust, *Locusta migratoria*). (A) Correlation between total repeat content and genome size; (B) Correlation between total repeat content and median length of intron; (C) Correlation between genome size and median length of intron; and Comparison of (D) LTR, (E) LINE, and (F) DNA transposable elements. Table S8. CRISPR sgRNA target sequences for the Ad vermillion gene. Table S9. Double-stranded RNA sequences used for RNAi experiments. Figure S10. Name and annotation of predicted *A. domesticus* genes, with scaffold number and length. Figure S11. Transcript sequences of predicted *A. domesticus* genes. Figure S12. *A. domesticus* egg microinjection slide. (A) Slide set up with eggs positioned on each side of the filter paper strips; (B) Injected eggs showing microinjection site with phenol red color in eggs; and (C) Developed eggs after microinjection showing eye pigmentation.

References

1. Hallmann, C.A.; Sorg, M.; Jongejans, E.; Siepel, H.; Hofland, N.; Schwan, H.; Stenmans, W.; Müller, A.; Sumser, H.; Hörren, T.; et al. More than 75 percent decline over 27 years in total flying insect biomass in pro tected areas. *PLoS ONE* **2017**, *12*, e0185809. [CrossRef]
2. IPBES. Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. In *IPBES Secretariat*; Brondízio, E.S., Settele, J., Díaz, S., Ngo, H.T., Eds.; IPBES: Bonn, Germany, 2019.
3. Jarvis, B. The Insect Apocalypse Is Here. The New York Times. 2018. Available online: www.nytimes.com/2018/11/27/magazine/insect-apocalypse.html (accessed on 1 July 2022).
4. Vogel, G. Where have all the insects gone? *Science* **2017**, *356*, 576–579. [CrossRef] [PubMed]
5. IPCC. Climate Change 2021: The Physical Science Basis. In *Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., et al., Eds.; Cambridge University Press USA: Cambridge, MA, USA, 2021. [CrossRef]

6. Pörtner, H.O.; Scholes, R.J.; Agard, J.; Archer, E.; Arneth, A.; Bai, X.; Barnes, D.; Burrows, M.; Chan, L.; Cheung, W.L.W.; et al. *Scientific Outcome of the IPBES-IPCC Co-Sponsored Workshop on Biodiversity and Climate Change*; IPBES Secretariat: Bonn, Germany, 2021. [CrossRef]

7. United Nations. Sustainable Development Goals, 15, Life on Land, Article: Sustainably Manage Forests, Combat Desertification, Halt and Reverse Land Degradation, Halt Biodiversity Loss. 2021. Available online: www.un.org/sustainabledevelopment/biodiversity/ (accessed on 17 March 2022).

8. United Nations. Department of Economic and Social Affairs, Population Division. World Population Prospects. 2019. Highlights. ST/ESA/SER.A/423. Available online: https://population.un.org/wpp/publications/files/wpp2019_highlights.pdf (accessed on 17 March 2022).

9. van Dijk, M.; Morley, T.; Rau, M.L.; Saghai, Y. A meta-analysis of projected global food demand and population at risk of hunger for the period 2010–2050. *Nat. Food* **2021**, *2*, 494–501. [CrossRef]

10. FAO; IFAD; UNICEF; WFP; WHO. The state of food security and nutrition in the world 2017. In *Building Resilience for Peace and Food Security*; FAO: Rome, Italy, 2017.

11. Bland, A. Is the Livestock Industry Destroying the Planet? Smithsonian Magazine. 2012. Available online: www.smithsonianmag.com/travel/is-the-livestock-industry-destroying-the-planet-11308007 (accessed on 1 July 2022).

12. FAO; United Nations. Land Use in Agriculture by the Numbers. 7 May 2020. Available online: www.fao.org/sustainability/news/detail/en/c/1274219 (accessed on 1 July 2022).

13. Steinfeld, H.; Gerber, P.; Wassenaar, T.D.; Castel, V.; Rosales, M.M.; Haan, C.D. *Livestock's Long Shadow: Environmental Issues and Options*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2016; p. xxiv 390.

14. GlobeNewswire. Demand for Global Protein Supplements Market Size to Surpass USD 32.56 Billion by 2028, Exhibit a CAGR of 9.29% | Protein Supplements Industry Trends, Share, Value, Analysis & Forecast Report by Facts & Factors. N.p. Web 3 October 2022 16:03 ET. Available online: https://www.globenewswire.com/en/news-release/2022/10/03/2527296/0/en/Demand-for-Global-Protein-Supplements-Market-Size-to-Surpass-USD-32-56-Billion-by-2028-Exhibit-a-CAGR-of-9-29-Protein-Supplements-Industry-Trends-Share-Value-Analysis-Forecast-Repo.html (accessed on 1 July 2022).

15. Donkersley, P.; Ashton, L.; Lamarre, G.P.A.; Segar, S. Global insect decline is the result of willful political failure: A battle plan for entomology. *Ecol. Evol.* **2022**, *12*, e9417. [CrossRef] [PubMed]

16. Dossey, A.T. Insects and their chemical weaponry: New potential for drug discovery. *Nat. Prod. Rep.* **2010**, *27*, 1737–1757. [CrossRef]

17. Losey, J.E.; Vaughan, M. The economic value of ecological services provided by insects. *BioScience* **2006**, *56*, 311–323. [CrossRef]

18. Noriega, J.A.; Hortal, J.; Azcárate, F.M.; Berg, M.P.; Bonada, N.; Briones, M.J.; Del Toro, I.; Goulson, D.; Ibanez, S.; Landis, D.A.; et al. Research trends in ecosystem services provided by insects. *Basic Appl. Ecol.* **2018**, *26*, 8–23. [CrossRef]

19. Saunders, M.E.; Rader, R. Ecosystem services of insects. In *Encyclopedia of Social Insects*; Starr, C., Ed.; Springer: Cham, Switzerland, 2020. [CrossRef]

20. Scudder, G.G. The importance of insects. In *Insect Biodiversity*; Foottit, R.G., Adler, P.H., Eds.; John Wiley & Sons: Hoboken, NJ, USA, 2017; pp. 9–43. [CrossRef]

21. Hotaling, S.; Sproul, J.S.; Heckenhauer, J.; Powell, A.; Larracuente, A.M.; Pauls, S.U.; Kelley, J.L.; Frandsen, P.B. Long reads are revolutionizing 20 years of insect genome sequencing. *Genome Biol. Evol.* **2021**, *13*, evab138. [CrossRef]

22. Oppert, B.; Perkin, L.C.; Lorenzen, M.; Dossey, A.T. Transcriptome analysis of life stages of the house cricket, *Acheta domesticus*, to improve insect crop production. *Sci. Rep.* **2020**, *10*, 3471.

23. Dossey, A.T.; Morales-Ramos, J.; Rojas, G. (Eds.) *Insects as Sustainable Food Ingredients: Production, Processing and Food Applications*; Academic Press: San Diego, CA, USA, 2016.

24. Dossey, A.T. Why insects should be in your diet. *Scientist* **2013**, *27*, 22–23.

25. Dossey, A.T.; Tatum, J.T.; McGill, W.L. Modern insect-based food industry: Current status, insect processing technology, and recommendations moving forward. In *Insects as Sustainable Food Ingredients: Production, Processing and Food Applications*; Dossey, A.T., Morales-Ramos, J.A., Guadalupe Rojas, M., Eds.; Academic Press: San Diego, CA, USA, 2016; pp. 113–152. [CrossRef]

26. Melgar-Lalanne, G.; Hernández-Álvarez, A.-J.; Salinas-Castro, A. Edible insects processing: Traditional and innovative technologies. *Compr. Rev. Food Sci. Food Saf.* **2019**, *18*, 1166–1191. [CrossRef] [PubMed]

27. Reverberi, M. Edible insects: Cricket farming and processing as an emerging market. *J. Insects Food Feed* **2020**, *6*, 211–220. [CrossRef]

28. Weissman, D.B.; Gray, D.A.; Pham, H.T.; Tijssen, P. Billions and billions sold: Pet-feeder crickets (Orthoptera: Gryllidae), commercial cricket farms, an epizootic densovirus, and government regulations make for a potential disaster. *Zootaxa* **2013**, *3504*, 67–88. [CrossRef]

29. DeFoliart, G.R. Insects as a source of protein. *Bull. Entomol. Soc. Amer.* **1975**, *21*, 161–164. [CrossRef]

30. Defoliart, G. Insects as human food. *Crop Prot.* **1992**, *11*, 395–399. [CrossRef]

31. Defoliart, G.R. Edible insects as minilivestock. *Biodivers. Conserv.* **1995**, *4*, 306–321.

32. Nakagaki, B.J.; Defoliart, G.R. Comparison of diets for mass-rearing *Acheta domesticus* (Orthoptera: Gryllidae) as a novelty food, and comparison of food Conversion efficiency with values reported for Livestock. *J. Econ. Entomol.* **1991**, *84*, 891–896.

33. Ramos-Elorduy, J. Insects: A sustainable source of food? *Ecol. Food Nutr.* **1997**, *36*, 247–276.

34. Ramos-Elorduy, J. Anthropo-entomophagy: Cultures, evolution and sustainability. *Entomol. Res.* **2009**, *39*, 271–288.

35. van Huis, A.; Van Itterbeeck, J.; Klunder, H.; Mertens, E.; Halloran, A.; Muir, G.; Vantomme, P. (Eds.) *Edible Insects: Future Prospects for Food and Feed Security*; FAO: Rome, Italy, 2013; p. 187.

36. Edible Insects Market to Reach $4.63 Billion by 2027, Growing at a CAGR of 26.5% from 2020 with COVID-19 Impact - Meticulous Research®Analysis. Available online: https://www.globenewswire.com/en/news-release/2021/01/12/2157080/0/en/Edible-Insects-Market-to-Reach-4-63-billion-by-2027-Growing-at-a-CAGR-of-26-5-from-2020-With-COVID-19-Impact-Meticulous-Research-Analysis.html (accessed on 3 June 2021).

37. Byrne, J. Investment in Insect Production Hits a New Level. 6 October 2020. Available online: www.feednavigator.com/Article/2020/10/06/Investment-in-insect-production-hits-a-new-level (accessed on 1 July 2022).

38. Manning, L. EXCLUSIVE: Beta Hatch Banks $9.3m Series A, Breaks Ground on US's Largest Insect Farm. AgFunderNews. 10 December 2020. Available online: https://agfundernews.com/exclusive-beta-hatch-raises-9-3m-series-a-funding-to-build-a-dozen-insect-ranches (accessed on 3 June 2021).

39. Ranken, T. Commentary: Insect Farming–Smart Investment in a World of Climate & Pandemic Challenges. CleanTech Alliance. 7 April 2020. Available online: www.cleantechalliance.org/2020/04/07/commentary-insect-farming-smart-investment-in-a-world-of-climate-pandemic-challenges/ (accessed on 1 July 2022).

40. Kataoka, K.; Minei, R.; Ide, K.; Ogura, A.; Takeyama, H.; Takeda, M.; Suzuki, T.; Yura, K.; Asahi, T. The draft genome dataset of the Asian cricket *Teleogryllus occipitalis* for molecular research toward entomophagy. *Front. Genet.* **2020**, *11*, 470. [CrossRef] [PubMed]

41. Ylla, G.; Nakamura, T.; Itoh, T.; Kajitani, R.; Toyoda, A.; Tomonari, S.; Bando, T.; Ishimaru, Y.; Watanabe, T.; Fuketa, M.; et al. Insights into the genomic evolution of insects from cricket genomes. *Commun. Biol.* **2021**, *4*, 7.

42. Li, F.; Zhao, X.; Li, M.; He, K.; Huang, C.; Zhou, Y.; Li, Z.; Walters, J.R. Insect genomes: Progress and challenges. *Insect Mol. Biol.* **2019**, *28*, 739–758. [CrossRef] [PubMed]

43. Adams, M.D.; Celniker, S.E.; Holt, R.A.; Evans, C.A.; Gocayne, J.D.; Amanatides, P.G.; Scherer, S.E.; Li, P.W.; Hoskins, R.A.; Galle, R.F.; et al. The genome sequence of *Drosophila melanogaster*. *Science* **2000**, *24*, 2185–2195. [CrossRef]

44. Lewin, H.A.; Robinson, G.E.; Kress, W.J.; Baker, W.J.; Coddington, J.; Crandall, K.A.; Durbin, R.; Edwards, S.V.; Forest, F.; Gilbert, M.T.P.; et al. Earth BioGenome Project: Sequencing life for the future of life. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 4325–4333. [CrossRef]

45. Mardis, E.R. DNA sequencing technologies: 2006–2016. *Nat. Protoc.* **2017**, *12*, 213.

46. Pareek, C.S.; Smoczynski, R.; Tretyn, A. Sequencing technologies and genome sequencing. *J. Appl. Genet.* **2011**, *52*, 413–435.

47. Takeyama, N.; Kiyono, H.; Yuki, Y. Plant-based vaccines for animals and humans: Recent advances in technology and clinical trials. *Ther. Adv. Vaccines* **2015**, *3*, 139–154.

48. Downs, M.; Johnson, P.; Zeece, M. Insects and their connection to food allergy. In *Insects as Sustainable Food Ingredients: Production, Processing and Food Applications*; Dossey, A.T., Morales-Ramos, J.A., Guadalupe Rojas, M., Eds.; Academic Press: San Diego, CA, USA, 2016; pp. 255–272. [CrossRef]

49. Johnston, J.S.; Bernardini, A.; Hjelmen, C.E. Genome size estimation and quantitative cytogenetics in insects. *Methods Mol. Biol.* **2019**, *1858*, 15–26. [CrossRef]

50. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410.

51. Nayfach, S.; Camargo, A.P.; Schulz, F.; Eloe-Fadrosh, E.; Roux, S.; Kyrpides, N.C. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **2021**, *39*, 578–585. [CrossRef]

52. Kleiner, M.; Wentrup, C.; Lott, C.; Teeling, H.; Wetzel, S.; Young, J.; Chang, Y.-J.; Shah, M.; VerBerkmoes, N.C.; Zarzycki, J.; et al. Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E1173–E1182. [CrossRef]

53. Kataoka, K.; Togawa, Y.; Sanno, R.; Asahi, T.; Yura, K. Dissecting cricket genomes for the advancement of entomology and entomophagy. *Biophys. Rev.* **2022**, *14*, 75–97. [CrossRef] [PubMed]

54. Petersen, M.; Armisén, D.; Gibbs, R.A.; Hering, L.; Khila, A.; Mayer, G.; Richards, S.; Niehuis, O.; Misof, B. Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. *BMC Evol. Biol.* **2019**, *19*, 11. [CrossRef] [PubMed]

55. Wang, X.; Fang, X.; Yang, P.; Jiang, X.; Jiang, F.; Zhao, D.; Li, B.; Cui, F.; Wei, J.; Ma, C.; et al. The locust genome provides insight into swarm formation and long-distance flight. *Nat. Commun.* **2014**, *5*, 2957. [CrossRef] [PubMed]

56. Lorenzen, M.D.; Brown, S.J.; Denell, R.E.; Beeman, R.W. Cloning and characterization of the *Tribolium castaneum* eye-color genes encoding tryptophan oxygenase and kynurenine 3-monooxygenase. *Genetics* **2002**, *160*, 225–234. [CrossRef] [PubMed]

57. Latunde-Dada, G.O.; Yang, W.; Aviles, M.V. In vitro iron availability from insects and sirloin beef. *J. Agric. Food Chem.* **2016**, *64*, 8420–8424. [CrossRef] [PubMed]

58. Williams, J.P.; Williams, J.R.; Kirabo, A.; Chester, D.; Peterson, M. Nutrient content and health benefits of insects. In *Insects as Sustainable Food Ingredients: Production, Processing and Food Applications*; Dossey, A.T., Morales-Ramos, J.A., Guadalupe Rojas, M., Eds.; Academic Press: San Diego, CA, USA, 2016; pp. 61–84.

59. Stull, V.J.; Finer, E.; Bergmans, R.S.; Febvre, H.P.; Longhurst, C.; Manter, D.K.; Patz, J.A.; Weir, T.L. Impact of edible cricket consumption on gut microbiota in healthy adults, a double-blind, randomized crossover trial. *Sci. Rep.* **2018**, *8*, 10762. [CrossRef]

60. Bertola, M.; Mutinelli, F.A. A systematic review on viruses in mass-reared edible insect species. *Viruses* **2021**, *13*, 2280. [CrossRef]

61. Giaccone, V. Hygiene and health features of mini livestock. In *Ecological Implications of Mini Livestock: Potential of Insects, Rodents, Frogs and Snails*; Paoletti, M.G., Ed.; Science Publisher: Enfield, UK, 2005; pp. 579–598.

62. Klunder, H.C.; Wolkers-Rooijackers, J.; Korpela, J.M.; Nout, M., Jr. Microbiological aspects of processing and storage of edible insects. *Food Control* **2012**, *26*, 628–631. [CrossRef]

63. European Commission. Approval of Third Insect as a Novel Food. 10 February 2022. Available online: ec.europa.eu/food/safety/novel-food/authorisations/approval-insect-novel-food_en (accessed on 17 March 2022).

64. Gahukar, R.T. Edible insects farming: Efficiency and impact on family livelihood, food security, and environment compared with livestock and crops. In *Insects as Sustainable Food Ingredients: Production, Processing and Food Applications*; Dossey, A.T., Morales-Ramos, J.A., Guadalupe Rojas, M., Eds.; Academic Press: San Diego, CA, USA, 2016; pp. 61–84. [CrossRef]

65. Oonincx, D.G.; van Itterbeeck, J.; Heetkamp, M.J.W.; Brand, H.V.D.; van Loon, J.J.A.; van Huis, A. An exploration on greenhouse gas and ammonia production by insect species suitable for animal or human consumption. *PLoS ONE* **2010**, *5*, e14445.

66. Cather, A. Two Billion People Eat Insects and You Can too. Food Tank. 19 September 2016. Available online: https://foodtank.com/news/2016/03/two-billion-people-eat-insects-and-you-can-too/ (accessed on 1 July 2022).

67. Costa-Neto, E.M.; Dunkel, F.V. Insects as food: History, culture, and modern use around the World. In *Insects as Sustainable Food Ingredients: Production, Processing and Food Applications*; Dossey, A.T., Morales-Ramos, J.A., Guadalupe Rojas, M., Eds.; Academic Press: San Diego, CA, USA, 2016; pp. 29–60. [CrossRef]

68. Oonincx, D.G.; De Boer, I.J.M. Environmental impact of the production of mealworms as a protein source for humans–A life cycle assessment. *PLoS ONE* **2012**, *7*, e51145. [CrossRef] [PubMed]

69. Capinera, J.L. *Encyclopedia of Entomology*; Kluwer Academic: Dordrecht, The Netherlands; London, UK, 2004.

70. Clifford, C.W. The Biology of Egg Production in the House Cricket, *Acheta domesticus* L. LSU Historical Dissertations and Theses. 1985; 4046. Available online: www.digitalcommons.lsu.edu/gradschool_disstheses/4046 (accessed on 17 March 2022).

71. Cannon, K.M.; Britt, D.T. Feeding one million people on Mars. *New Space* **2019**, *4*, 245–254. [CrossRef]

72. Katayama, N.; Ishikawa, Y.; Takaoki, M.; Yamashita, M.; Nakayama, S.; Kiguchi, K.; Kok, R.; Wada, H.; Mitsuhashi, J.; Space Agriculture Task Force. Entomophagy: A key to space agriculture. *Adv. Space Res.* **2008**, *41*, 701–705. [CrossRef]

73. FAO. *Women-Users, Preservers and Managers of Agro-Biodiversity*; FAO: Rome, Italy, 1998; Available online: https://agris.fao.org/agris-search/search.do?recordID=XF2000393347 (accessed on 1 July 2022).

74. Dossey, A.T. Chemical defenses of insects: A rich resource for chemical biology in the tropics. In *Chemical Biology of the Tropics: An Interdisciplinary Approach*; Vivanco, J., Weir, T., Eds.; Springer: New York, NY, USA, 2011; pp. 27–57.

75. Stork, N.E. How many species of insects and other terrestrial arthropods are there on Earth? *Annu. Rev. Entomol.* **2018**, *63*, 31–45. [CrossRef] [PubMed]

76. Horch, W.H.; Mito, T.; Popadić, A.; Ohuchi, H.; Noji, S. (Eds.) *Cricket as a Model Organism*; Springer: Tokyo, Japan, 2017; p. 376.

77. Duffield, K.R.; Hunt, J.; Sadd, B.M.; Sakaluk, S.K.; Oppert, B.; Rosario, K.; Behle, R.W.; Ramirez, J.L. Active and covert infections of cricket iridovirus and *Acheta domesticus* Densovirus in reared *Gryllodes sigillatus* crickets. *Front. Microbiol.* **2021**, *12*, 780796. [CrossRef] [PubMed]

78. Hedges, L.M.; Brownlie, J.C.; O'Neill, S.L.; Johnson, K.N. *Wolbachia* and virus protection in insects. *Science* **2008**, *322*, 702. [CrossRef]

79. Fernandez-Cassi, X.; Söderqvist, K.; Bakeeva, A.; Vaga, M.; Dicksved, J.; Vagsholm, I.; Jansson, A.; Boqvist, S. Microbial communities and food safety aspects of crickets (*Acheta domesticus*) reared under controlled conditions. *J. Insects Food Feed.* **2020**, *6*, 429–440. [CrossRef]

80. Burse, A.; Boland, W. Deciphering the route to cyclic monoterpenes in Chrysomelina leaf beetles: Source of new biocatalysts for industrial application? *Z. Naturforsch. C* **2017**, *72*, 417–427. [CrossRef]

81. Yi, H.Y.; Chowdhury, M.; Huang, Y.D.; Yu, X.Q. Insect antimicrobial peptides and their applications. *Appl. Microbiol. Biotechnol.* **2014**, *98*, 5807–5822. [CrossRef]

82. Warr, E.; Das, S.; Dong, Y.; Dimopoulos, G. The Gram-Negative Bacteria-Binding Protein gene family: Its role in the innate immune system of *Anopheles gambiae* and in anti-Plasmodium defence. *Insect Mol. Biol.* **2008**, *17*, 39–51. [CrossRef]

83. Feng, M.; Fei, S.; Xia, J.; Labropoulou, V.; Swevers, L.; Sun, J. Antimicrobial peptides as potential antiviral factors in insect antiviral immune response. *Front. Immunol.* **2020**, *11*, 2030. [CrossRef] [PubMed]

84. Wu, Q.; Patočka, J.; Kuča, K. Insect antimicrobial peptides, a mini review. *Toxins* **2018**, *10*, 461. [CrossRef] [PubMed]

85. Adamo, S.A. Estimating disease resistance in insects: Phenoloxidase and lysozyme-like activity and disease resistance in the cricket *Gryllus texensis*. *J. Insect Physiol.* **2004**, *50*, 209–216. [CrossRef] [PubMed]

86. Sorrell, M.R.; Killian, K.A. Innate immune system function following systemic RNA-interference of the Fragile X Mental Retardation 1 gene in the cricket Acheta domesticus. *J. Insect Physiol.* **2020**, *126*, 104097. [CrossRef] [PubMed]

87. Marshall, J.L.; Huestis, D.L.; Hiromasa, Y.; Wheeler, S.; Oppert, C.; Marshall, S.A.; Tomich, J.M.; Oppert, B. Identification, RNAi knockdown, and functional analysis of an ejaculate protein that mediates a postmating, prezygotic phenotype in a cricket. *PLoS ONE* **2009**, *4*, e7537. [CrossRef]

88. Koren, S.; Walenz, B.P.; Berlin, K.; Miller, J.R.; Bergman, N.H.; Phillippy, A.M. Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **2017**, *27*, 722–736. [CrossRef]

89. Putnam, N.H.; O'Connell, B.L.; Stites, J.C.; Rice, B.J.; Blanchette, M.; Calef, R.; Troll, C.J.; Fields, A.; Hartley, P.D.; Sugnet, C.W.; et al. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* **2016**, *26*, 342–350. [CrossRef]

90. Lieberman-Aiden, E.; van Berkum, N.L.; Williams, L.; Imakaev, M.; Ragoczy, T.; Telling, A.; Amit, I.; Lajoie, B.R.; Sabo, P.J.; Dorschner, M.O.; et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **2009**, *326*, 289–293. [CrossRef]

91. Johnson, A.D.; Handsaker, R.E.; Pulit, S.L.; Nizzari, M.M.; O'Donnell, C.J.; de Bakker, P.I.W. SNAP: A web-based tool for identification and annotation of proxy snps using hapmap. *Bioinformatics* **2008**, *24*, 2938–2939. [CrossRef]

92. Wood, D.E.; Lu, J.; Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **2019**, *20*, 1–3. [CrossRef]

93. Flynn, J.M.; Hubley, R.; Goubert, C.; Rosen, J.; Clark, A.G.; Feschotte, C.; Smit, A.F. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 9451–9457. [CrossRef] [PubMed]

94. Smit, A.; Hubley, R.; Grenn, P. RepeatMasker Open-4.0. 2015. Available online: www.repeatmasker.org (accessed on 3 June 2021).

95. Stanke, M.; Steinkamp, R.; Waack, S.; Morgenstern, B. AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Res.* **2004**, *32*, W309–W312. [CrossRef] [PubMed]

96. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [CrossRef] [PubMed]

97. Cantarel, B.L.; Korf, I.; Robb, S.M.C.; Parra, G.; Ross, E.; Moore, B.; Holt, C.; Alvarado, A.S.; Yandell, M. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **2007**, *18*, 188–196. [CrossRef] [PubMed]

98. Chan, P.P.; Lowe, T.M. TRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. In *Gene Prediction [Internet]*; Kollmar, M., Ed.; Springer: New York, NY, USA, 2019; pp. 1–14, (Methods in Molecular Biology; vol. 1962); Available online: http://link.springer.com/10.1007/978-1-4939-9173-0_1 (accessed on 26 September 2021).

99. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [CrossRef]

100. Lawrence, M.; Huber, W.; Pagès, H.; Aboyoun, P.; Carlson, M.; Gentleman, R.; Morgan, M.; Carey, V.J. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **2013**, *9*, e1003118. [CrossRef]