Published in final edited form as:

Insect Mol Biol. 2013 August ; 22(4): 433–441. doi:10.1111/imb.12034.

A unique Y gene in the Asian malaria mosquito *Anopheles stephensi* encodes a small lysine-rich protein and is transcribed at the onset of embryonic development

Frank Criscione, Yumin Qi, Randy Saunders, Brantley Hall, and Zhijian Tu

Department of Biochemistry, Virginia Tech, Blacksburg, VA

Abstract

In many organisms the Y chromosome initiates sex-determination and regulates male fertility and mating behavior. However, molecular characterization of Y genes is rare outside of a few model species because it is difficult to clone and analyze repeat-rich heterochromatic Y sequences. In insects, Y genes are only well characterized in a small number of *Drosophila* species. Here we report the discovery of *GUY1*, a gene unique to the Y in the Asian malaria mosquito, *Anopheles stephensi*, using an approach that compares Illumina sequences separately obtained from male and female genomic DNA. Experimental evidence confirmed that *GUY1* is a single copy gene found only on the Y chromosome. *GUY1* is transcribed at the very onset of zygotic transcription and encodes a small lysine-rich protein that forms two alpha helices and shows DNA-binding properties. Interestingly, three helix-loop-helix proteins are key factors that determine sex in the early embryo in *D. melanogaster*. Single embryo analysis indicated that *GUY1* is only transcribed in male embryos and the *GUY1* promoter is functional in the early embryos. *GUY1* may be used as a paternally inherited molecular marker. Further investigation of *GUY1* will contribute to the genetic approaches to control mosquito-borne diseases.

Keywords

male; maternal-to-zygotic transition; vector-borne infectious disease; Y chromosome

Introduction

Anopheles mosquitoes use the XX/XY sex-determination system. Females possess two X chromosomes and males possess one X and one Y. There is evidence that a dominant male-determining factor(s) on the *Anopheles* Y chromosome initiates sexual differentiation (Baker and Sakai, 1979; Marin and Baker, 1998), which is similar to humans and several species of flies (e.g., Marin and Baker, 1998; Kaiser and Bachtrog, 2010). The common fruit fly *Drosophila melanogaster* also has the XX/XY sex chromosomes. However, it is the X to autosome ratio, not the mere presence of the Y chromosome that determines the sexual fate of the *D. melanogaster* embryo (Marin and Baker, 1998). Genetic studies show that the *D. melanogaster* Y chromosome is involved in the regulation of male fertility, mating behavior, and global gene expression (e.g., Carvalho *et al.*, 2001; Lemos *et al.*, 2008; Zhou *et al.*, 2012). Similarly, the *Anopheles* Y chromosome has also been shown to control the stenogamy-eurygamy mating behavior (Fraccaro *et al.*, 1977). Therefore, it is clear that the Y chromosome is critical to the sexual dimorphism in dipteran insects regardless of its involvement in the initiation of sex determination.

Molecular characterizations of Y genes are rare outside of a few model species, mostly due to the repetitive nature of the Y chromosome and the difficulty of cloning repeat-rich heterochromatic sequences (Carvalho *et al.*, 2001). In insects, Y genes are only reported from a few *Drosophila* species (Carvalho *et al.*, 2001, 2009). Only a dozen or so Y genes have been reported in *Drosophila*, most of which were discovered through comparison of unmapped sequences to the non-redundant protein database. This method was applied to the African malaria mosquito *An. gambiae* but failed to identify any Y genes (Krzywinski *et al.*, 2006). In this study, we report the discovery of a unique Y chromosome gene in the Asian malaria mosquito, *An. stephensi*, using an approach that compares Illumina sequencing results of RNA samples and separately isolated male and female genomic DNAs. We name this gene *GUY1* for Gene Unique to the Y. *GUY1* encodes a small lysine-rich protein and it is expressed at the very beginning of the maternal-to-zygotic transition. The possible function of *GUY1* and the implications to mosquito control by manipulation of sexual dimorphism are discussed. Y genes such as *GUY1* can also provide a novel set of markers for phylogenetic, population and mating behavioral studies. Such markers offer perspectives fundamentally different from the other types of markers as their inheritance is exclusively paternal (e.g., Kopp *et al.*, 2006; Underhill and Kivisild, 2007).

Results

Discovery and molecular characterization of GUY1

Male and female *An. stephensi* genomic DNA sequences were generated via Illumina sequencing (SRA accession SRP013838). The male and female data were compared using blastn with an e-value threshold of $1e^{-3}$. The male data was used as the query and the female data was used as the database. All of the male sequences that matched any female sequences were removed, leaving only male-specific sequences. The male-specific sequences were then compared to the RNA-seq databases (SRA accession SRA054605) of all developmental stages to identify transcribed genes. Y genes should neither be expressed in adult females nor the 0–1 hr old embryos. This is because females do not have the Y chromosome and the 0–1 hr embryos are transcriptionally inactive and mainly contain maternally deposited transcripts. Using this strategy, we obtained 45 male-specific Illumina reads (Supplemental file 1) that had no matches to the transcriptomes of the adult female or 0–1 hr embryos but had at least 3 total matches to the transcriptomes from the adult male, other embryonic time points, larvae or pupae of mixed sexes. These 45 sequences were further analyzed as candidate Y genes. Many of these are similar to reverse transcriptase from retrotransposons.

We focused on a set of paired-end sequences (Supplemental file 1) that showed embryonic transcription and were not of retrotransposon origin. These sequences were used to design primers (Table 1) for PCR and RT-PCR, which subsequently led to the discovery and characterization of the full-length *GUY1*cDNA (Genbank JX174417) and a 5.5 kb genomic fragment that contains the *GUY1* gene (GenBank JX154128). The *GUY1* transcription start site and polyA addition site were determined by 5' and 3' RACE. Figure 1A depicts key features of the *GUY1* gene, which includes a short ORF, 5' UTR, 3' UTR, and the 5' flanking sequences. *GUY1* is intronless, a feature shared with many early zygotic genes (Biedler *et al.*, 2012), and it has a short ORF encoding a novel protein of 56 amino acid residues. As shown in Figure 1B, this protein is rich in lysine and could form two alpha helices. The second helix contains a stretch of basic lysine residues, a structural feature that has been shown to mediate sequence-specific DNA binding in many transcription factors. The first helix is amphipathic with hydrophobic residues on one side (Figure 1C), which could facilitate dimerization through hydrophobic interactions.

Male-specific GUY1 amplification in different strains of *An. stephensi*

Although some Y genes have highly similar paralogues on the autosome or X chromosome (e.g., Carvalho et al., 2009), our *in silico* subtractive approach focused on the identification of sequences unique to the Y chromosome. Such Y-specific sequences, like *GUY1*, should be found exclusively in the males. Male-specific amplification was demonstrated for all strains tested when *GUY1* primers (Table 1) were used to amplify genomic DNA. There are three biological forms or races of *An. stephensi*, namely the “type” form, the *mysorensis* form and the intermediate form. Shown in Figure 2A are PCR results indicating male-specific amplification of *GUY1* in three strains (VT, UCI, and JHU) of the “type” biological form. Male-specific amplification of *GUY1* is also observed in the IRN and CHB strains of the *mysorensis* form and the Beech strain of the intermediate form (data not shown). As positive controls, *RPS4* (ribosomal protein subunit 4) was successfully amplified from both male and female genomic DNA sources in all experiments described above.

GUY1 containing DNA is transferred to females during mating

To further establish the Y chromosome origin of the novel sequences we performed PCR using genomic DNA isolated from females before and after mating (Figure 2B). Mature sperm are a mixture of X-containing and Y-containing haploid cells. During mating the sperm are transferred to the female and stored in the spermatheca in the female abdomen. Virgin and mated females were dissected into two parts: head and thorax (together) and abdomen. *GUY1* amplification, after 40 cycles, was detected only with genomic DNA isolated from the abdomens of mated females, the location where sperm are stored.

GUY1 is a single copy sequence in males

Copy number of *GUY1* was determined as described by Hindson *et al.* (2011) using QX100 Droplet Digital PCR System from Bio-Rad. Setting the copy number of a known single-copy autosomal reference to 2 for the diploid alleles, the copy number of *GUY1* in males and females was determined. The primers and Taqman probes used to compare *GUY1* and the autosomal reference are provided in Table 1. Shown in Figure 2C are representatives of the QX100 Droplet Digital PCR experiments. Male copy number values (0.96, 0.85, 0.95) and female copy number values (0.01, 0.02, 0.01) are consistent with *GUY1* having a single copy in males and no copies in females. This copy number determination experiment was repeated on 18 male and 18 female individuals and produced similar results.

GUY1 expression is detected at the onset of maternal-to-zygotic transition

As shown in Figure 3, *GUY1* expression appears to be restricted to the embryonic stage as there was no amplification of *GUY1* from larval, pupal, adult male or adult female cDNA. *GUY1* transcripts start to accumulate at 2 hrs after egg deposition and tapers off after 12 hrs (Figure 3). RT-PCR results are consistent with RNA-seq analysis and both suggest that *GUY1* is mainly transcribed 2–8 hrs after egg deposition and not detectable in other stages (Table 2). RT-PCR on early embryonic samples with tighter time intervals in early embryos showed that *GUY1* is transcribed as early as 2–3 hrs post egg deposition (data not shown). DAPI nuclei staining was used to estimate the timing of the early embryonic development in *An. stephensi*. Under the experimental conditions described in this paper, most of the embryos had 16 to 32 nuclei at 1–2 hrs after egg deposition, representing mitotic cycles 5 and 6. By 2 to 3 hrs after oviposition, most embryos were at cycles 8 and 9. The syncytial blastoderm stage (mitotic cycles 10–13) occurs approximately 3–4 hrs after egg deposition in *An. stephensi* (Supplemental Figure S1), similar to *An. gambiae* (Juhn and James, 2006). Thus, *GUY1* transcription is initiated at the onset of maternal-to-zygotic transition, prior to syncytial blastoderm formation.

Single-embryo analysis showed that *GUY1* is only expressed in males and the *GUY1* promoter is able to direct reporter gene expression in both male and female embryos

By dividing the lysate from a single embryo for RNA and genomic DNA isolation, we were able to sex the single embryo and detect the transcription within the same embryo. In addition to *GUY1*, other Y-chromosome sequences including a Y-specific repeat (Table 1) were used as markers to sex the embryo prior to RT-PCR. As expected, *GUY1* is only expressed in the male embryo (Figure 4A). The 850 bp upstream promoter of *GUY1* contains an arthropod initiator TCAGT, an upstream TATAA box (GenBank JX154128), and 6 copies of a TAGteam-like motif that was shown to activate early zygotic transcription in *Ae. aegypti* (Biedler *et al.*, 2012) and *Drosophila* (Erickson *et al.*, 1998; Sefton *et al.*, 2000; Liang *et al.*, 2008). This promoter region was used in a dual luciferase reporter assay and shown to be active in *An. stephensi* early embryos (Figure 4B). As expected for a promoter active at the very onset of zygotic transcription, the *GUY1* promoter functions in both male and female embryos (Figure 4C). However, native *GUY1* transcription was not detected in female embryos (Figure 4A) because there is no endogenous *GUY1* gene in the XX embryo.

Discussion

We have discovered a Y-specific gene, *GUY1*, in *An. stephensi* using a bioinformatic approach comparing high throughput sequences of male and female genomic DNA and RNA samples. Florescent *in situ* hybridization (FISH) can provide direct evidence if *GUY1* signal is observed on the Y chromosome. However, Y-chromosome FISH is challenging in *Anopheles* mosquitoes as we cannot use polytene chromosomes from ovarian nurse cells. In salivary glands, the Y chromosome does not polytenize but becomes a part of the chromocenter where all chromosomes contact each other (Sharakhova *et al.*, 2006). Mitotic chromosome FISH is an alternative but the sensitivity is low for mapping short single copy sequences such as *GUY1* onto highly heterochromatic Y. Male-specific amplification is accepted as strong evidence for location on the Y (e.g., Kryzinski *et al.*, 2004; Ng'habi *et al.*, 2007). In addition to male-specific amplification of *GUY1* in seven strains of all three biological forms of *An. stephensi*, we have also shown other evidence including the transfer of *GUY1*-containing DNA to the abdomen of mated females, *GUY1* being a single copy (single allele) gene in males, and the lack of expression of *GUY1* in females or 0–1 hr embryos. Therefore we have described one of the strongest cases of Y genes in non-*Drosophila* insects.

GUY1 is transcribed at the very onset of zygotic transcription and encodes a protein that contains two alpha helices and has the potential to bind DNA. The *GUY1* protein may belong to either one of the two major families of transcription factors, namely the helix-loop-helix or the helix-turn-helix proteins (Massari and Murre, 2000; Aravind *et al.*, 2005). The strength of the predicted structure connecting the two helices in *GUY1* does not allow us to distinguish between these two possibilities with confidence. Interestingly, three helix-loop-helix proteins (*daughterless*, *deadpan*, and *sisterless-b/scute*) are key transcription factors that trigger the sex determination pathway the early embryos in *D. melanogaster* (flybase.org). The expression pattern of *GUY1* suggests it has a role in very early embryonic development in males. The fact that the *GUY1* promoter functions in XX embryos (Figure 4C) suggests that *GUY1* transcription does not require other Y factor(s), which is consistent with the hypothesis that *GUY1* may serve as an initial signal in sex-determination.

GUY1 has no apparent paralogues on the X or autosomes as it is unique to males. Genes on the *Drosophila* Y chromosome often originate from the autosomes through translocation (Carvalho *et al.*, 2009). It is not clear whether *GUY1* is a Y-specific innovation or there existed an X- or autosomal- *GUY1* paralogue, which either was lost or took a divergent

evolutionary path than *GUY1*. There is no apparent *GUY1* homolog in *An. gambiae*, which is not surprising as there is very low gene content conservation in the Y chromosome of the 12 sequenced *Drosophila* species (Koerich *et al.*, 2008).

Y-specific genes such as *GUY1* can provide a novel set of markers for phylogenetic, population and mating behavioral studies. Y chromosome markers lack recombination and offer advantages such as rapid coalescence and reduced interspecific introgression (e.g., Kopp *et al.*, 2006; Underhill and Kivisild, 2007). They offer perspectives fundamentally different from the other types of markers as their inheritance is exclusively paternal. These new markers will help resolve the often intricate phylogenetic relationships of *Anopheles* mosquitoes which often comprise cryptic species complexes. As mentioned above, there are three biological forms of *An. stephensi* that vary in vector capacity. Y-specific markers provide tools for the study of population structure within and between these forms which may offer new insights into the genetic basis of vectorial capacity and disease transmission. *GUY1* is also a reliable marker to sex *An. stephensi* embryos and early instar larvae, which will be very useful but until now was not feasible. For example, we may now analyze the transcriptomes of male and female embryos and investigate dosage compensation in the early embryos (e.g., Baker and Russell, 2011).

Approximately one-third of the world's population is at risk of infection with *Plasmodium* parasites which cause malaria. In the Middle East and Asia a major malaria vector is *An. stephensi*. Only female mosquitoes feed on blood. Hence, only females transmit disease pathogens. Current genetic strategies for controlling mosquito-borne diseases include the release of sterile males (Sterile Insect Technique, SIT), the release of insects carrying a dominant lethal gene (RIDL), or the replacement or conversion of a pathogen-susceptible population into a pathogen-resistant one (e.g., Benedict and Robinson, 2003; Wise de Valdez *et al.*, 2011). A better understanding of the possible functions of Y genes in sexual differentiation will facilitate the production of sterile males for improved SIT. It may also enable the production of highly competitive males to improve SIT and RIDL and to help spread pathogen resistance genes. Genetic manipulation of sex ratios could greatly reduce or crash a local mosquito population (Schliekelman *et al.*, 2005) and subsequently control mosquito-borne infectious diseases.

Experimental procedures

Mosquito rearing and mating

There are three biological forms or races of *An. stephensi* which differ in egg morphology (Subbarao *et al.*, 1987; Oshaghi *et al.*, 2006). *An. stephensi stephensi* is the “type” form and is an efficient malaria vector in urban areas. *An. stephensi mysorensis* is a rural species and a poor vector. The third form is the intermediate. The Indian “type” form of *An. stephensi stephensi* maintained at Virginia Tech (the VT strain) was used for Illumina sequencing and PCR analysis of genomic DNA and cDNA templates. The VT strain was reared in incubators at 27°C and 60% relative humidity on a 16 hr light/8 hr dark photoperiod. Larvae were fed Sera Micron Fry Food with brewer's yeast, and Purina Game Fish Chow. Adult mosquitoes were fed on a 10% sucrose soaked cotton pad. Mosquitoes were blood-fed on female Hsd:ICR (CD-1®) mice (Harlan Laboratories, <http://www.harlan.com>). Egg collections occurred approximately 3 days post blood-feeding. Six strains that belong to the three *An. stephensi* forms were included in a PCR survey using genomic DNA as template. They are: two additional *An. stephensi stephensi* strains from the laboratory of Anthony James at the University of California, Irvine (the UCI strain) and the laboratory of George Dimopolous at the Johns Hopkins University (the JHU strain); two *An. stephensi mysorensis* strains collected from Iran (the IRN and the CHB strains provided by Maryam Kamali and Igor Sharakhov at Virginia Tech); and one intermediate form from the London School of

Hygiene and Tropical Medicine (the Beech strain). For mating experiments, individual pupae were separated into 50 mL tubes. After emergence, the mosquitoes were sexed based on antennae dimorphism and double caged to mitigate mating through screen. Virgin female mosquitoes were flash frozen 3 days post emergence. Mated female mosquitoes were obtained by placing males and females in a 3:1 ratio in cages 3 days post emergence. Mating was allowed to occur for 5 days, at which point males and females were separated and flash frozen.

Genomic DNA (gDNA) and RNA sequencing by Illumina

All Illumina sequencing was performed using genomic DNA and RNA isolated from the VT strain of *An. stephensi*. Illumina sequencing was performed separately on *An. stephensi* DNA extracted from virgin adult males and adult females. DNA extraction was performed using DNAeasy from Qiagen (Valencia, CA). The male and female genomic sequence data, each consisting of over 35 million 84 bp paired-end reads, are deposited in the NCBI *Sequence* Read Archive (SRA) with the accession number of SRP013838. Library preparation and Illumina sequencing were performed at the University of California Davis Genome Center. RNA-seq was performed on polyA RNAs across the following embryonic time points and developmental stages: 0–1 hr, 2–4 hr, 4–8 hr, and 8–12 hr embryos, mixed-instar larva, pupa, adult females, and adult males. RNA-seq data of all samples, each consisting of more than 7 million reads of 38–41 bp sequences, are also deposited to SRA and the accession number is SRA054605. RNA-seq library preparation and Illumina sequencing were performed at the DNA facility of the Iowa State University.

gDNA PCR to confirm male-specific amplification

gDNA was isolated from individual adult males and virgin females of various strains. gDNA was also isolated from dissected body parts for the purposes of comparing virgin and mated females. The heads and thoraces of 5 virgin females were pooled together as a sample and the remaining abdomens were pooled together as another sample. The same procedure was performed for the mated females. All gDNA samples were collected in triplicate and processed using DNAzol® (Molecular Research Center, Inc., Cincinnati, OH). The manufacturer's recommendations were followed for gDNA isolation. After air drying the DNA pellets were re-suspended in 30 uL of nuclease-free water. PCR for the *GUY1* and *RPS4* control genes was performed using the primers listed in Table 1. The cycling conditions for all reactions were as follows: 95°C for 5 minutes; 30 cycles of (95°C for 30 seconds, 60°C for 30 seconds, and 72°C for 3 minutes); and final extension at 72 °C for 5 minutes. Amplification of *GUY1* from virgin and mated mosquito parts was performed for 40 cycles.

RNA isolation, cDNA synthesis, RT-PCR, and RACE

An. stephensi embryos, larvae, pupae, adult males and adult stage females were collected and flash frozen. RNA was extracted from approximately 100 ug of each of the above mentioned samples following the total RNA extraction protocol for the mirVana RNA Isolation kit (Ambion). Total RNA was eluted into 100 uL of elution buffer and treated with Turbo DNA-free (Applied Biosystems Inc.) stored at –80°C. cDNA synthesis was performed using SuperScript III RT first strand cDNA synthesis kit (Invitrogen). The amount of starting RNA in each reverse transcription reaction was 5 ug. The final cDNA product was diluted 1:5 with nuclease-free water and stored at –20°C. RT-PCR for *GUY1* and *RPS4* control was performed using the primers listed in Table 1. The cycling conditions for all reactions were as follows: 95°C for 5 minutes; 30 cycles of 95°C for 30 seconds, 60°C for 30 seconds, and 72°C for 3 minutes; and final extension at 72°C for 5 minutes. For 5' and 3' RACE to amplify the *GUY1* cDNA ends, cDNA was made using 1 ug of RNA isolated from 2–8 hr old *An. stephensi* embryos. 5' and 3' RACE cDNA synthesis was

carried out using the Firstchoice RLM-RACE kit (Life Technologies). The manufacturer's recommendations were followed for cDNA synthesis. cDNA synthesis and RACE amplification primers are listed in Table 1. Amplification of target template utilized the following PCR cycling conditions: 94°C for 3 minutes, and 35 cycles of 95°C for 30 seconds, 60°C for 30 seconds, and 72°C for 30 seconds. The reaction was finished with 72°C for 7 minutes.

DAPI staining of *An. stephensi* embryos

We used DAPI staining to estimate the timing of the early embryonic development in *An. stephensi*. Gravid *An. stephensi* females were allowed to lay eggs over a period of 1 hr. Collected eggs/embryos were either processed immediately (for the 0–1 hr time point) or allowed to develop to the desired time period at 27°C. Embryos were treated with 1% sodium hypochlorite briefly and fixed for 1.5 hrs with 9% paraformaldehyde. The chorion was further cracked and peeled according to Goltsev *et al.* (2004). A final concentration of 0.5 µg/ml of DAPI (4',6-diamidino-2-phenylindole, MP Biomedicals) in PBS was used to stain the embryos for 10 minutes. Embryos were washed twice with PBS containing 0.1% Tween 20 and photographed at 100x or 200x magnification. For some of the early time points, we counted nuclei by observation at different focal planes within the embryo.

Single embryo gDNA and RNA analysis

Individual 7–8 hr old embryos were placed into 5 µL of Lysis Buffer (0.05M DTT and 10U RNase OUT from Invitrogen) and homogenized. The lysate was flash frozen and stored at –80°C until gDNA preparation or RNA extraction was required. For gDNA isolation 2.5 µL of the lysate was treated with 0.3 mAU of proteinase K for 30 minutes at 28°C, and then inactivated at 95°C for 2 minutes. The reaction was then diluted with 6.5 µL of ddH₂O. For RNA isolation 2.5 µL of the lysate was treated with 130U of DNase I (Invitrogen) for 1 hr at 25°C. The reaction was stopped with the addition 1 µL of 25 mM EDTA and incubation at 65°C for 10 minutes. The resulting 4.5 µL reaction was carried over to cDNA synthesis utilizing the previously mentioned SuperScript III reverse transcriptase protocol. The cycling conditions for all reactions were as follows: 95°C for 5 minutes; 40 cycles at 95°C for 30 seconds, 60°C for 30 seconds, and 72°C for 3 minutes; and final extension at 72°C for 5 minutes.

Single embryo dual-luciferase assay and sexing of the single embryo

GUY1 promoter activity was measured using the pGL3 vector containing firefly luciferase. The 847 bp upstream region of *GUY1* was cloned into upstream of the firefly luciferase ORF. The pRL actin plasmid (*Drosophila* actin promoter driving *Renilla* luciferase) was used as a control. The two plasmids were mixed to reach a final concentration of 150 ng/µl and co-injected into 30–35 minute old embryos. Dual-luciferase assay was performed on individual embryos 5 hrs after the injection. The manufacturer's suggestions for the Dual Luciferase Assay (Promega) were followed with a few modifications: 20 µL of Passive Lysis Buffer was used to homogenize the embryo, 20 µL of Luciferase Assay Reagent II was added during the first reading, and 20 µL of Stop and Glow was added during the second reading. This assay was modified from Biedler and Tu (2010). Phenol:chloroform extraction was performed on the single embryo luciferase assay mixture using the Quick-gDNA MicroPrep kit (Zymo Research, Irvine, CA). This gDNA was used as template to determine the sex of the single embryo through PCR targeting *GUY1* and a Y-specific repeat sequences (Table 1). As in other PCR assays, *RPS4* was used as positive control to ensure template quality. Genomic DNA of male embryos should produce positive PCR products for *GUY1*, the Y-specific repeat, and *RPS4*. Genomic DNA of female embryos should only produce positive PCR products for *RPS4*.

Digital Droplet PCR

Copy number of *GUY1* was determined using QX100 Droplet Digital PCR System from Bio-Rad (Hindson *et al.*, 2011) with known single copy autosome genes as references. The Bio-Rad Digital Droplet PCR uses microfluidics to convert a single Taqman (Applied Biosystems Inc.) PCR mixture into 20,000 droplets, thus performing and analyzing 20,000 TaqMan assays simultaneously. Results of the Taqman assays are then computed and the copy number of a gene of interest can be determined by comparing to a reference (Hindson *et al.*, 2011). Genomic DNA obtained via the previously mentioned protocol from male and female adult individuals was digested with AvrII (Fermentas, Waltham, MA). The digested DNA was purified with the GFX Gel band purification kit (GE Healthcare, Piscataway, NJ) and used as template in the copy number variation (CNV) experiment. Two probe and primer mixes were used to carry out TaqMan assays: one targeting *GUY1* and the other targeting an autosomal reference sequence (Table 1). The two probes are labeled with different fluorescent dyes to allow detection of both in one reaction droplet. Droplet generation and reading was also performed in accordance with the manufacturer's copy number variation protocol.

Bioinformatic analysis of nucleic acid and protein sequences

Illumina sequences were analyzed mainly using BLAST (Altschul *et al.*, 1997) on a 64-bit Thinkmate Linux workstation with 48 processors and 256 GB RAM. For male and female comparisons, male Illumina sequences were used as query and female reads were used as the database to perform BLASTN with an e-value cutoff of $1e-3$. We chose BLAST over faster short-read analysis programs such as Bowtie (Langmead, 2010) because we are interested in identifying male and female matches that may have insertion or deletions or partial sequence matches. To speed up the BLASTN search, only top match was reported for any given query. Candidate Y sequences were also compared to RNA-seq databases using BLASTN to infer transcription of these candidate sequences. Up to 1 million hits can be reported for each query and an evalue cutoff of $1e-7$ was used to minimize hits from paralogous sequences. To search for *GUY1* homologues, BLASTX, BLASTP, and TBLASTN were performed against the NCBI non-redundant databases and the *An. gambiae* PEST genome as well as all trace reads (vectorbase.org). Secondary structure of the deduced amino acid sequence of *GUY1* protein was predicted using Jpred3 (<http://www.compbio.dundee.ac.uk/www-jpred/>). Helical wheel analysis was performed using the Helical Wheel Projections program (<http://rzlab.ucr.edu>).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Jim Biedler for help with the dual-luciferase assay. We thank Maryam Kamali, Igor Sharakhov, Anthony James, and George Dimopolous for providing various *Anopheles stephensi* strains. We thank Yann Jouvenot and Jack Regan at Bio-Rad for help with digital-droplet PCR. This work is supported by NIH Grants AI077680 and AI105575, the Fralin Life Science Institute, and the Virginia Tech Experimental Station.

References

- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*. 1997; 25:3389–3402. [PubMed: 9254694]

- Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM. The many faces of the helix-turn-helix domain: Transcription regulation and beyond. *FEMS microbiology reviews*. 2005; 29:231–262. [PubMed: 15808743]
- Baker DA, Russell S. Role of testis-specific gene expression in sex-chromosome evolution of *Anopheles gambiae*. *Genetics*. 2011; 189:1117–1120. [PubMed: 21890740]
- Baker RH, Sakai RK. Triploids and male determination in the mosquito, *Anopheles culicifacies*. *Journal of Heredity*. 1979; 70:345–346. [PubMed: 43343]
- Benedict MQ, Robinson AS. The first releases of transgenic mosquitoes: an argument for the sterile insect technique. *Trends in Parasitology*. 2003; 19:349–355. [PubMed: 12901936]
- Biedler J, Tu Z. Evolutionary analysis of the kinesin light chain genes in the yellow fever mosquito *Aedes aegypti*: gene duplication as a source for novel early zygotic genes. *BMC Evolutionary Biology*. 2010; 10:206. [PubMed: 20615250]
- Biedler JK, Hu W, Tae H, Tu Z. Identification of early zygotic genes in the yellow fever mosquito *aedes aegypti* and discovery of a motif involved in early zygotic genome activation. *PLoS ONE*. 2012; 7:e33933. [PubMed: 22457801]
- Carvalho AB, Dobo BA, Vibranovski MD, Clark AG. Identification of five new genes on the Y chromosome of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*. 2001; 98:13225–13230.
- Carvalho AB, Koerich LB, Clark AG. Origin and evolution of Y chromosomes: *Drosophila* tales. *Trends in Genetics*. 2009; 25:270–277. [PubMed: 19443075]
- Carvalho AB, Vibranovski M, Carlson J, Celniker S, Hoskins R, Rubin G, Sutton G, Adams M, Myers E, Clark A. Y chromosome and other heterochromatic sequences of the *Drosophila melanogaster* genome: How far can we go? *Genetica*. 2003; 117:227–237. [PubMed: 12723702]
- Erickson JW, Cline TW. Key aspects of the primary sex determination mechanism are conserved across the genus *Drosophila*. *Development*. 1998; 125:3259–3268. [PubMed: 9671597]
- Fraccaro M, Tiepolo L, Laudani U, Marchi A, Jayakar SD. Y chromosome controls mating behaviour on *Anopheles mosquitoes*. *Nature*. 1977; 265:326–328. [PubMed: 834279]
- Goltsev Y, Hsiong W, Lanzaro G, Levine M. Different combinations of gap repressors for common stripes in *Anopheles* and *Drosophila* embryos. *Developmental Biology*. 2004; 275:435–446. [PubMed: 15501229]
- Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, Bright IJ, Lucero MY, Hiddessen AL, Legler TC, et al. High-throughput droplet digital pcr system for absolute quantitation of DNA copy number. *Analytical Chemistry*. 2011; 83:8604–8610. [PubMed: 22035192]
- Juhn J, James AA. *oskar* gene expression in the vector mosquitoes, *Anopheles gambiae* and *Aedes aegypti*. *Insect Molecular Biology*. 2006; 15:363–372. [PubMed: 16756555]
- Kaiser VB, Bachtrog D. Evolution of sex chromosomes in insects. *Annual Review of Genetics*. 2010; 44:91–112.
- Koerich LB, Wang X, Clark AG, Carvalho AB. Low conservation of gene content in the *Drosophila* Y chromosome. *Nature*. 2008; 456:949–951. [PubMed: 19011613]
- Kopp A, Frank A, Fu J. Historical biogeography of *Drosophila simulans* based on Y-chromosomal sequences. *Molecular Phylogenetics and Evolution*. 2006; 38:355–362. [PubMed: 16051503]
- Krzywinski J, Chrystal M, Besansky N. Gene finding on the Y: Fruitful strategy in *Drosophila* does not deliver in *Anopheles*. *Genetica*. 2006; 126:369–375. [PubMed: 16636930]
- Krzywinski J, Nusskern DR, Kern MK, Besansky NJ. Isolation and characterization of y chromosome sequences from the african malaria mosquito *Anopheles gambiae*. *Genetics*. 2004; 166:1291–1302. [PubMed: 15082548]
- Langmead, B. *Current Protocols in Bioinformatics*. Vol. Chapter 11. John Wiley & Sons, Inc; 2002. Aligning short sequencing reads with Bowtie.
- Lemos B, Araripe LO, Hartl DL. Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. *Science*. 2008; 319:91–93. [PubMed: 18174442]
- Liang HL, Nien CY, Liu HY, Metzstein MM, Kirov N, Rushlow C. The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*. *Nature*. 2008; 456:400–403. [PubMed: 18931655]

- Marin I, Baker BS. The evolutionary dynamics of sex determination. *Science*. 1998; 281:1990–1994. [PubMed: 9748152]
- Massari ME, Murre C. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Molecular and Cellular Biology*. 2000; 20:429–440. [PubMed: 10611221]
- Ng'habi KR, Horton A, Knols BGJ, Lanzaro GC. A new robust diagnostic polymerase chain reaction for determining the mating status of female *Anopheles gambiae* mosquitoes. *The American Journal of Tropical Medicine and Hygiene*. 2007; 77:485–487. [PubMed: 17827364]
- Oshaghi MA, Yaaghoobi F, Abaie MR. Pattern of mitochondrial DNA variation between and within *Anopheles stephensi* (Diptera: Culicidae) biological forms suggests extensive gene flow. *Acta Tropica*. 2006; 99:226–233. [PubMed: 16989757]
- Schliekelman P, Ellner S, Gould F. Pest control by genetic manipulation of sex ratio. *Journal of Economic Entomology*. 2005; 98:18–34. [PubMed: 15765662]
- Sefton L, Timmer JR, Zhang Y, Beranger F, Cline TW. An extracellular activator of the *Drosophila* JAK/STAT pathway is a sex-determination signal element. *Nature*. 2000; 405:970–973. [PubMed: 10879541]
- Sharakhova MV, Xia A, McAlister SI, Sharakhov IV. A standard cytogenetic photomap for the mosquito *Anopheles stephensi* (Diptera: Culicidae): Application for physical mapping. *Journal of Medical Entomology*. 2006; 43:861–866. [PubMed: 17017220]
- Subbarao SK, Vasantha K, Adak T, Sharma VP, Curtis CF. Egg-float ridge number in *Anopheles stephensi*: ecological variation and genetic analysis. *Medical and Veterinary Entomology*. 1987; 1:265–271. [PubMed: 2979540]
- Underhill PA, Kivisild T. Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annual Review of Genetics*. 2007; 41:539–564.
- Wise de Valdez MR, Nimmo D, Betz J, Gong HF, James AA, Alphey L, Black WC. Genetic elimination of dengue vector mosquitoes. *Proceedings of the National Academy of Sciences*. 2011; 108:4772–4775.
- Zhou J, Sackton TB, Martinsen L, Lemos B, Eickbush TH, Hartl DL. Y chromosome mediates ribosomal DNA silencing and modulates the chromatin state in *Drosophila*. *Proceedings of the National Academy of Sciences*. 2012; 109:9941–9946.

Figure 1.

Structure of the *GUY1* gene and its predicted protein. A) *GUY1* gene schematic. The structure is drawn according to the cDNA and genomic DNA sequences that have been submitted to GenBank (JX174417 and JX154128). 5' and 3' RACE were performed to experimentally obtain the full length cDNA sequence. The promoter region shown is the region that was investigated by a dual-luciferase reporter assay as shown in Figure 4. B) Predicted secondary structure of the *GUY1* protein. The top line is the sequence and the bottom indicates the secondary structure. H, alpha helix; E, extended. C) The helical wheel depiction of the first alpha helix. Note that hydrophobic residues from V4 to Y10 are located on one side of the helix.

Figure 2.

GUY1 is a male-specific and single copy sequence in *An. stephensi*. A) Amplification of *GUY1* from adults of various strains of *An. stephensi*. VT, UCI and JHU are strain names. *RPS4* is amplified from each template as a positive control. B) Transfer of *GUY1*-containing DNA to female abdomen during mating. Pools of virgin and mated females were dissected. DNA was isolated from the collective heads and thoraces as well as the abdomens. *GUY1* was only amplified from the abdomens of mated females. *RPS4* was successfully amplified from all samples. C) Determination of *GUY1* copy number by Digital Droplet PCR (ddPCR). ddPCR was performed on genomic DNA isolated from individual adult male and female mosquitoes. Copy numbers are reported for each individual with Poisson correction.

Figure 3.

Expression profile of *GUY1*. The cDNA samples included the following embryonic time points: 0–2 hr, 2–4 hr, 4–40 hr in 4 hr increments, and 40–42 hr. Larvae, pupae, whole adult male and female, and adult male tissue or body parts (head, thorax, abdomen, and testis) were also used. *RPS4* was amplified from each template as a positive control.

Figure 4.

Sexing the single embryo and single embryo luciferase assays. A) gDNA and cDNA derived from the same embryo, at 7–8 hrs post oviposition, was used as template for PCR. Male embryos were shown to express *GUY1* whereas female embryos do not. B) *GUY1* promoter activity measured by a dual-luciferase assay. Embryos were injected with two plasmids: *GUY1* promoter driving the firefly luciferase and the *Drosophila* acting promoter driving the *Renilla* luciferase. Individual embryos were assayed for both luciferase activities and the *Renilla* luciferase activity is used for normalization and the ratio of the relative fluorescent unit (RFU) is reported. The averages and standard errors are shown (p-value=0.001 for comparison between the *GUY1* promoter and the pGL3-null). C) *GUY1* promoter activity in male and female embryos. The assays are the same as described in panel B. Here, embryos of known sex were grouped according to gender. The *GUY1* promoter activity measured as the ratio of RFUs is plotted for each sex. The *GUY1* promoter activity is not significantly different between male and female embryos (p-value=0.39).

Table 1

Primer and probe sequences.

<i>GUY1_F</i> ¹	TTTACTCGTCAAAGCTGCCA
<i>GUY1_R</i> ¹	GATCCGTAAAAATTGACACCA
<i>RPS4_F</i> ²	GAGTCCATCAAAGGAGAAAGTCTAC
<i>RPS4_R</i> ²	TAGCTGGCGCATCAGGTAC
<i>GUY1_FullcDNA_F</i> ³	CCTGAAATGATGCTCTGGAAA
<i>GUY1_FullcDNA_R</i> ³	GATAAATAACCCAGACTATGTGAAGG
<i>GUY1_RACE_F</i> ⁴	CCGGTATAGTCTAATCTCGTGAGT
<i>GUY1_RACE_R</i> ⁴	GTGGTCTTATGCTATTGTCTGTCA
<i>GUY1_RACE_Fnest</i> ⁴	GGGAAATTTACTCGTCAAAGC
<i>GUY1_RACE_Rnest</i> ⁴	CCAATGTCACAGCAGAGTGTT
<i>GUY1_promoter_F</i> ⁵	TTGAAAAACACACGGTAAACG
<i>GUY1_promoter_R</i> ⁵	AATTTGCCTATTTTACAGTGACGC
<i>AutoRef_ddP_F</i> ⁶	ATCACCCTCGTCGTCCTT
<i>AutoRef_ddP_R</i> ⁶	CGAACGAACTCGATTGACCC
<i>AutoRef_ddP_P</i> ⁶	HEX-GCAAACACCACAACAGCAGC-BHQ1
<i>GUY1_ddP_F</i> ⁷	GTCAAAGCTGCCACGGATCT
<i>GUY1_ddP_R</i> ⁷	TCCAATGTCACAGCAGAGTGTTT
<i>GUY1_ddP_P</i> ⁷	FAM-TCACAAAGTAGGCGATACAAAAACA-BHQ1
<i>Y-repeat_F</i> ⁸	CAAGCCCAAACAACTCCTC
<i>Y-repeat_R</i> ⁸	ATGCAGCATTGAGCTTCCTT

¹Primers used to demonstrate male-specific amplification of *GUY1* and to amplify *GUY1* cDNA.²Positive control primers used to perform both genomic DNA and RT-PCR, which produces products of different sizes due to splicing.³Primers used to perform RT-PCR to amplify full-length *GUY1* cDNA.⁴Initial and nested primers used to perform 5' RACE (R primers) and 3' RACE (F primers) for *GUY1*.⁵Primers used to clone the *GUY1* promoter to the pGL3 plasmid for the luciferase assay.⁶The autosomal reference primers and probe used for Digital Droplet PCR Taqman assays.⁷The *GUY1* primers and probe used for Digital Droplet PCR Taqman assays⁸Primers used to amplify the Y-specific repeat.

Table 2

Number of *GUY1* reads in 11 *An. stephensi* RNA-seq samples.

Sample	0-2 hr	2-4 hr	4-8 hr	8-12 hr	L	P	M	F1	F2	F3	F4
<i>GUY1</i> (Raw)	0	54	32	0	0	0	0	0	0	0	0
<i>GUY1</i> (Normalized)	0	4.15	3.86	0	0	0	0	0	0	0	0

The first four samples are embryos of different time points. L, larvae; P, pupae; M, adult males. F1-4, four adult female samples including ovaries and carcasses from sugar-fed and blood-meal females. Normalized *GUY1* reads are calculated as the number of *GUY1* reads per million of total reads of that sample.