Simple Quantitative PCR Approach to Reveal Naturally Occurring and Mutation-Induced Repetitive Sequence Variation on the *Drosophila Y* Chromosome

John C. Aldrich, Keith A. Maggert*

Department of Biology, Texas A&M University, College Station, Texas, United States of America

Abstract

Heterochromatin is a significant component of the human genome and the genomes of most model organisms. Although heterochromatin is thought to be largely non-coding, it is clear that it plays an important role in chromosome structure and gene regulation. Despite a growing awareness of its functional significance, the repetitive sequences underlying some heterochromatin remain relatively uncharacterized. We have developed a real-time quantitative PCR-based method for quantifying simple repetitive satellite sequences and have used this technique to characterize the heterochromatic *Y* chromosome of *Drosophila melanogaster*. In this report, we validate the approach, identify previously unknown satellite sequence copy number polymorphisms in *Y* chromosomes from different geographic sources, and show that a defect in heterochromatin formation can induce similar copy number polymorphisms in a laboratory strain. These findings provide a simple method to investigate the dynamic nature of repetitive sequences and characterize conditions which might give rise to long-lasting alterations in DNA sequence.

Introduction

A significant portion of most eukaryotic genomes is composed of repetitive DNA elements [1]. It is estimated that as much as 1/3 of the genome of *Drosophila melanogaster* is composed of such sequences [2,3]. This fraction is largely confined to centric and telomeric regions where it forms constitutive heterochromatin, cytologically distinct in its appearance and genetically distinct in its properties. Constitutive heterochromatic sequences are largely of two types: middle repetitive sequences such as transposable elements, and highly repetitive major- and micro-satellite sequences [3–8]. Although highly-repetitive heterochromatic satellite sequences (*e.g.*, AAGAG, AATAT, AAGAGAG) house a variety of biological phenomena including centromere function, chromosome cohesion and pairing, nuclear organization, control of recombination, species-compatibilities, replication rate, and gene regulatory variation [9–14], understanding their function mechanistically has lagged far behind sophisticated understanding of the function of euchromatic sequences. This is due in large part to the difficulty in handling these sequences with modern molecular biological approaches. Next-generation sequencing technology has increased the rate with which we have learned about the structure and variation of euchromatin, but the heterochromatic portion of the genome remains relatively ignored in its characterization [3,15], even very recently not rising to the level of notice in debate over the role of "junk" DNA [16,17].

The *Y* chromosome of *Drosophila melanogaster* is a useful tool for understanding the evolution of satellite sequences and their contribution to genome regulation [18,19]. The *Drosophila Y* chromosome is naturally variant, can be made supernumerary in males or females, is dispensable in males, has very few genes, is a component of numerous chromosome rearrangements, and its functional and sequence elements have been roughly mapped. Apart from genes necessary for male fertility and a small set of non-essential genes, the *Y* chromosome is almost entirely composed of repetitive DNA such as megabase-long blocks of satellite repeats –variously called alphoid repeats, alpha-hetero-chromatic repeats, satellite repeats, simple repeats, simple satellite repeats (SSRs), highly-repetitive DNAs, repetitious DNAs, *etc*. –as well as interspersed or clustered transposable elements, the repetitive Ribosomal RNA genes (*rDNA*), and other genetic elements [5,6,20]. *Y* chromosomes isolated from diverse populations affect a number of phenotypes including temperature sensitivity, sex ratio, heterochromatin formation, male fitness, innate immunity, and others [21–25] and may do so by differentially influencing genome-wide transcription. Although some of these effects can be attributed to *rDNA* copy number polymorphisms [26,27], it is likely that the balance of unmapped variation lies within satellite sequence [28].

"Complex" euchromatin contains ample sequence variation to analyze for function, while the sequence variation of satellites has fewer parameters in which it can vary. Blocks of satellite repeats can vary in their length (*i.e.*, copy number), homogeneity (*i.e.*, polymorphisms in the consensus repeat unit), punctuation (*i.e.*, location, type, and copy number of transposable elements or transposable element remnants), orientation (*e.g.*, AAGAG or

CTCTT in relation to the centromere), juxtapositions (*e.g.*, the types or arrangements of satellite repeats at junctions), or linkage (to specific chromosomal locations). There have been some attempts to explore these features, but it is difficult to apply standard molecular tools to understand the architecture of the heterochromatin. Currently, studies to address variation have chiefly measured linkage and copy number using fluorescence *in situ* hybridization or Southern blot analysis.

Acknowledging that no approach is perfect, and following on our recent experiments [29,30] demonstrating the importance of *rDNA* copy number variation in heterochromatin formation and *Y*-linked Regulatory Variation, we wished to develop a similar method to quantify the copy number of satellite repeats that is (i) simple, (ii) robust, (iii) sensitive, (iv) quantifiable, (v) inexpensive, (vi) fast, and (vii) can be integrated with other approaches to provide an understanding of the arrangements of satellite DNAs.

"Real-Time" or "Quantitative" Polymerase Chain Reaction (qPCR) has been successfully used to accurately quantify *rDNA* copy-number variation in numerous studies [31–35], and is theoretically directly applicable to any repetitive sequence element whose repeat unit is longer than the typical approximately 100 base pair product of qPCR. The absence of unique primer binding sites in blocks of short (*e.g.*, pentameric or heptameric) satellites makes avoidance of primer-primer annealing the chief difficulty. An assay that circumvented this problem and allowed the amplification and quantification of simple telomeric repeats has been developed [36,37]. We thought this assay could in principle be adapted for heterochromatic satellites, which in many regards pose the same problems as telomeric DNA: short, homogenous, high copy number. In this study, we show that we successfully adapted this qPCR technique for the quantification of pentameric satellites. We validated precision using a dilution series and *Y* chromosome aneuploids, and found that geographically diverse *Y* chromosomes harbor previously uncharacterized satellite copy-number polymorphisms. Furthermore, we applied the approach to discover that long-term exposure to a mutation affecting heterochromatin formation and genome stability, the *Su(var)205* locus which encodes the HP1a gene product, results in measurable changes in satellite copy number, suggesting that much like *rDNA* [38–40], satellite copy number stability is regulated by chromatin factors.

Results

Design of Real-Time-Based Quantitative PCR Approach

Large blocks of simple pentameric repeats AACAC and AAGAC are constituents of the *Drosophila Y* chromosome [7,41], accounting for less than about 2% and about 20%, respectively, of the *Y*; the remaining balance largely resides in the pericentric heterochromatin of chromosome 2. In order to investigate copy number variation of these repeats, we adapted a Real-Time Quantitative PCR (qPCR) assay, originally designed for quantifying telomeric repeat copy number by Cawthon [36,37], that would allow us to quantify their relative copy number. The reaction used primers with designed self-incompatibilities to disfavor primer-dimer formation and instead heavily-favor template-dependent and product-dependent priming. The products of template-dependent synthesis created self-compatible products, which were preferentially amplified exponentially as is normal in PCR reactions.

Five design elements were incorporated into primer design. First, a "Forward" primer matching the repeat (*e.g.*, AACAC) contained a base-pair change (therefore a mismatch with the repeat) every 5 nucleotides. Second, the "Reverse" primer (*e.g.*, GTGTT) did so as well, but the mismatch was not the same as that on the "Forward" primer. Third, the primer set (Forward and Reverse) converged at a position in the repeat that was not complementary (*i.e.*, they did not overlap at their 3′ ends). Fourth, the primers each contained five nucleotides at their 5′ends that were not homologous to the repeat. Fifth, the primers had nucleotides at their 3′ends such that the best primer-primer annealing configurations had minimally two 3′mismatches [37].

This design balanced qPCR primers (i) effectively binding to and priming from the genomic satellite DNA repeat, (ii) exponentially amplifying from products of previous cycles of the "chain reaction" amplification, and (iii) avoiding primer-dimers forming between primers both directed at the same repetitive DNA sequence. Key to this end, introduced base pair mismatches ("First" and "Second" design elements above) were out of phase with each other and compromised the binding between primers and target genomic DNA, but more egregiously compromised binding with each other. This is clarified in Figure 1A, which shows the sequence of primers directed at AACAC repeats. A homogenous block of AACAC binds "AACAC Forward" ("sense," homologous to AACAC) and "AACAC Reverse" ("antisense," homologous to GTGTT), each with multiple single mismatches spread throughout the primer length.

Computational Justification of Approach

Computational analyses of the separate AACAC Forward and AACAC Reverse primers show that the designs retain the preference for annealing to a defined phase of the pentanucleotide repeats (Figure 1B–C, and as shown in Figure 1A). In this case, AACAC Forward anneals best when the 5′end of the primer overlaps with the second nucleotide of AACAC (aAcac) and ends at the fourth nucleotide in the repeat unit (aacAc). Annealed primers in this phase pair perfectly with genomic sequence at 28 bases (Figure 1B, offset 1 *vs*. AACAC), including five clusters of five consecutive pairs. There is no simple method to accurately predict pairing energy or melting temperature of intentionally mismatched primers that can compete for multiple sequences, but using a thumb-rule (2°C for each A/T and 4°C for each G/C) is suggestive of the predilection for this phase of binding. By this calculation, the melting temperature is 80°C for this phase and less than 50°C for other phases. AACAC Reverse also has a preference for 5′alignment with the second nucleotide of the pentanucleotide repeat (gTgtt) (Figure 1C, offset 1 *vs*. GTGTT), and also ends on the fourth nucleotide (gtgTt). Pairing in this phase has a similar distribution of five consecutive paired bases, and a similar thumb-rule melting temperature of 78°C.

Primer-dimers are a constant concern in primer design, and the repetitious nature of the target sequences makes avoidance difficult because there are multiple pairing arrangements that are a function of the repeat-length. We analyzed the number of possible base pairs forming given every degree of overlap between AACAC Forward and AACAC Reverse (Figure 1D–E). The repeat unit length is clear as a local maximum every five nucleotides, flanked by two far-sub-optimal arrangements around each local maximum (*i.e.*, offset by 4–5–6 nucleotides, 9–10–11, *etc*.). AACAC Forward and Reverse best pair with an offset of five nucleotides (as shown in Figure 1D) which creates eight internal mismatches, disrupts pairing of more than 3 consecutive bases, and leaves 3′mismatches on both ends, which significantly inhibits polymerase elongation. This total number of base pairs (22/38) is lower than reactions primed on genomic DNA (28/38 for AACAC Forward or AACAC Reverse), and the thumb-rule approximate melting temperature (62°C) is less favored than either primer annealing to genomic DNA. Additionally, this "best" match allows only ten
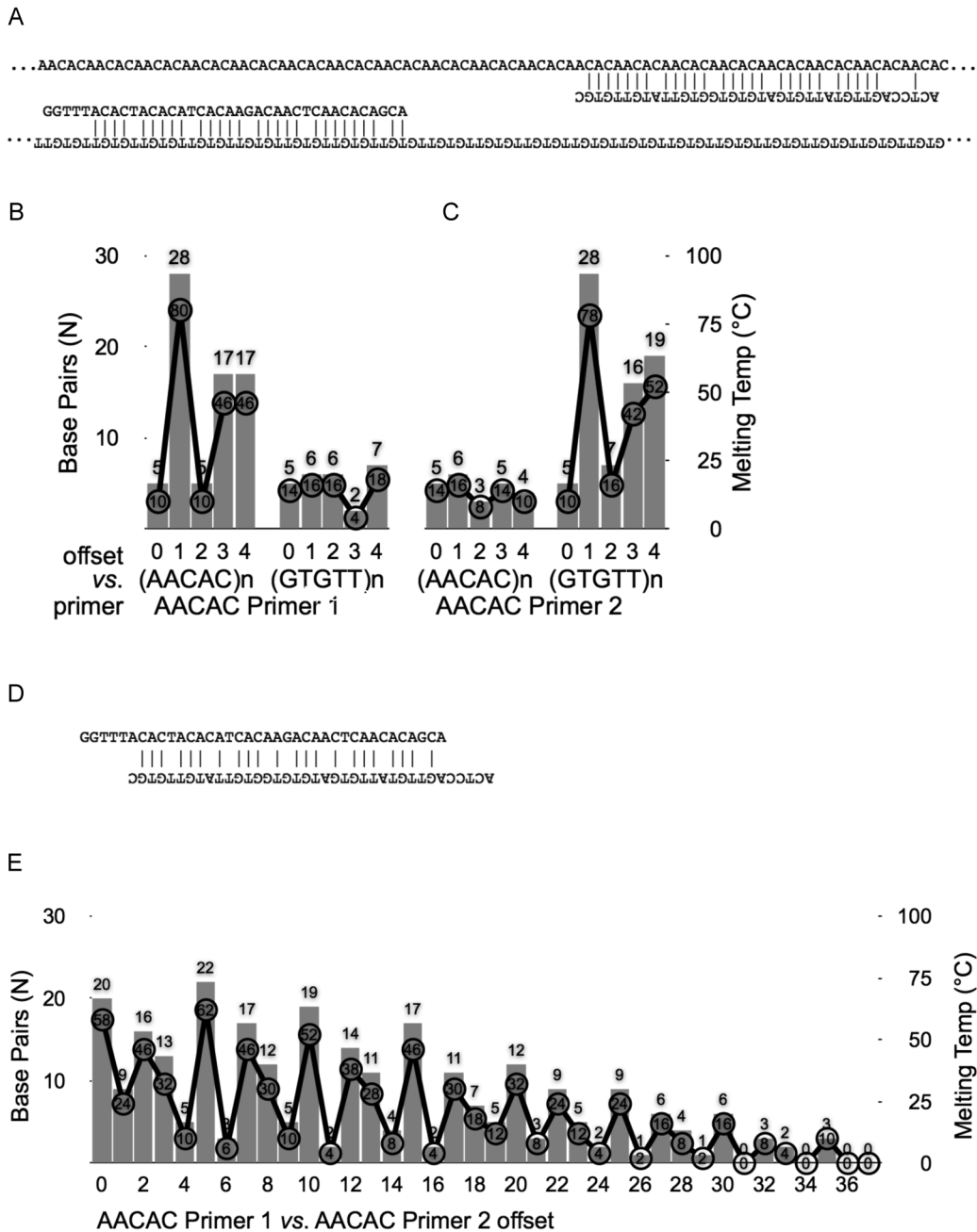
Figure 1. Computational validation of qPCR primers directed at AACAC satellite. (**A**) Representation of homogenous block of AACAC repeat denatured and annealing to AACAC Forward and AACAC Reverse. Vertical lines indicate base pairs. (**B**) AACAC Forward hypothetically annealing to each of the five "phases"of AACAC (*i.e.,* AACAC, ACACA, CACAA, ACAAC, and CAACA) and it's reverse-complement. Bars indicate the number of base pairs possible in that phase, and line indicates *ad hoc*"melting"analysis. AACAC Forward has a strong preference for one phase and no affinity for the reverse-complement strand of the repeat. (**C**) as in (B), but with AACAC Reverse. (**D**) Shows the best possible pairing between AACAC Forward and Reverse. (**E**) AACAC Forward and AACAC Reverse pairing with every possible degree of overlap, from completely (at "0 offset") to single 3′-most nucleotides pairing (at "37 offset"). Bars and lines are as in (B–C). In no case is AACAC Forward/Reverse dimer preferred over annealing to genomic targets (A–C) or the product of previous amplification.
doi:10.1371/journal.pone.0109906.g001

nascent nucleotides (both 5-base pair 3′-overhangs) to be incorporated during a PCR extension, much less signal than would included by even the worst priming of a previously-amplified primer (40 nucleotides if the primers annealed at juxtaposed genomic repeats; 38 for the primer plus the 2 of non-overlap in the repeat unit, …aacaCAacac…) or a valid genomic DNA-primed event. After the second successive cycle of priming

and elongation, there are no longer any mismatches between primer and PCR-produced template, thus normal qPCR conditions are established.

Experimental Validation of Quantitative PCR-Based Satellite Quantification

To confirm the robustness of our assay, we performed qPCR reactions on isolated genomic DNA over approximately a 100-fold dilution range (1.23 ng–100 ng per reaction) surrounding optimal conditions determined empirically in other studies ([31], Aldrich and Maggert, submitted). Over an intermediate range (3.7 ng–33.3 ng), we observed a very high correlation ($R^2 = 0.99$) between template concentration and quantification cycle (C_q, [42]) using primers directed at the copy number stable multicopy tRNA^{K-CTT} gene [31], AACAC, or AAGAC (Figure 2A), which matched our experience with amplification of the middle-repetitive *35S/45S* ribosomal RNA gene [31] and others'experience with simple telomeric repeats [36]. In practice to assure robustness, we routinely perform reactions using DNA concentrations falling within the middle of this range (about 4–10 nanograms). We recommend this concentration, however our results indicate that fluctuations in the DNA concentration due to variation in extraction or errors in preparation will have negligible influence over the result.

We analyzed the raw data from the qPCR reaction of each target for efficiency using the LinRegPCR software package [43]. This analysis ascertains closeness-to-doubling (efficiency) with each

qPCR cycle, thus a score of 2.0 is theoretically ideal. The efficiency values for tRNA^{K-CTT}, AACAC, AAGAC, and AAGAG, respectively, are 1.86 ± 0.003, 1.87 ± 0.004, 1.87 ± 0.002, and 1.86 ± 0.004 (each calculated from 12 reactions, errors are standard errors of the mean). Although these values are below theoretical maximal efficiency, they are all similar, thus any correction that would be applied to the data to account for sub-ideal efficiencies would be applied equally to all values and are effectively canceled out when reporting relative values. These efficiency values are within generally accepted guidelines (90%–110%) despite the intentional mismatches in satellite-directed primer sets. *Post-hoc* melt-curve analysis confirmed that only single melting peaks were observed from these reactions, indicating single PCR products were amplified during qPCR (Figure S1), supporting our computational justification.

We next used the $\Delta\Delta C_q$ method of analyzing qPCR results to quantify repeat copy-number of AACAC and AAGAC relative to that of the tRNA^{K-CTT} gene [31]. Although these satellite repeats have been cytologically mapped, little information about their overall abundance in the genome is available. They are found on the Y chromosome, which can be removed or made supernumerary without defects in viability, allowing us to manipulate Y chromosome copy number to monitor the sensitivity of our assay. We collected infrequent (frequency$\approx 10^{-4}$) spontaneous primary nondisjunctional exceptional progeny from a *yellow¹ white⁶⁷ᶜ²³/Y, 10B y⁺* stock [44], or created secondary nondisjunctional progeny (see Materials and Methods). Yellow + females were crossed to euploid brothers and *y¹ w⁶⁷ᶜ²³/Y, 10B y⁺/Y, 10B y⁺* progeny were identified by their duskier bodies, a consequence of the Y-terminal duplication of the *yellow⁺* gene translocation. We determined copy number of pentameric AACAC and AAGAC in sibling *X/Y* and *X/Y/Y* males (Figure 2B); data are shown as %AANAN (indicating either AACAC or AAGAC) with the values for *y¹ w⁶⁷ᶜ²³/Y, 10B y⁺* (our reference chromosomes) defined as 100%. AACAC and AAGAC are thus treated separately because we cannot support an *a priori* expectation that the AACAC and AAGAC primers sets should prime qPCR reactions with the same metrics (annealing temperature, elongation rate, fluorescence, efficiency, *etc.*). Similarly, determining absolute copy number (using known tRNA copy number as a multiplier) is not valid.

As expected, males with an extra Y chromosome possessed elevated Y-linked AACAC and AAGAC repeats. By pooling siblings during DNA extraction, we lost data on standard deviations between individuals, hence the error bars report standard errors of the means. Based on these averages, we estimate that Y-linked blocks of AACAC and AAGAC contribute approximately 29% and 44% to the total amounts of those respective satellites to the euploid *y¹ w⁶⁷ᶜ²³/Y, 10B y⁺* genome.

Quantitative PCR Satellite Analysis Reveals Strain Differences in Satellite Repeat Copy Number

It is of note that our estimate of Y-linked AAGAC levels differs from a previously published estimate of 69% [7]. While this discrepancy might simply reflect the differing sensitivities of qPCR and radiolabelled or fluorescence *in situ* hybridization, it might also represent variation between different laboratory stocks. Repetitive sequence variation is of course not without precedence [28]. Examples include the expansion and contraction of ribosomal DNA in yeast and flies [39,45], as well as interspersed satellite copy-number polymorphisms in humans and plants [46,47]. Indeed, it is hypothesized that such variation may underlie the differential gene-regulatory effects of geographically divergent *Drosophila* Y chromosomes [23–25,27].
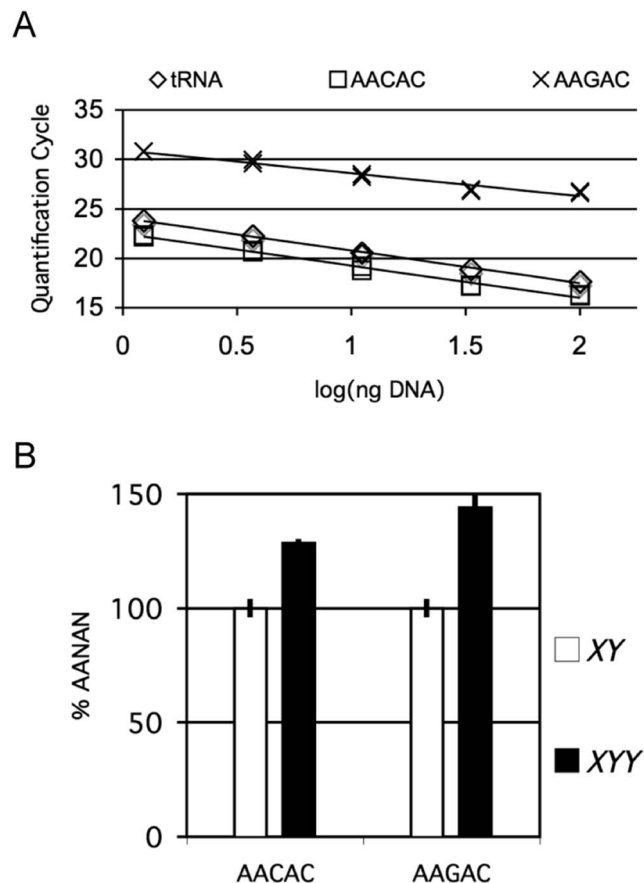
Figure 2. Experimental validation of qPCR assay to measure satellite copy number. (**A**) Quantification cycle (C_q) of duplicate qPCR reactions plotted as a function of template DNA per reaction. X-axis represents $\log_{10}$ of an approximately 100-fold dilution series. (**B**) Quantification of satellite copy number in *X/Y* males and to *X/Y/Y* males, relative to *X/Y* (defined as 100%). Error bars represent standard error of the mean (S.E.M.) derived from triplicate reactions.
doi:10.1371/journal.pone.0109906.g002

To address this possibility, we asked if we could detect satellite copy number differences on three of the Y chromosomes used in studies of unidentified Y-linked regulatory variation, referred to as Y, Ohio, Y, Congo, and Y, Zimbabwe. We introduced each of these Y chromosomes into otherwise-isogenic backgrounds by multiple patrilineal backcross to strains bearing homozygous recessive mutations on the X and autosomes [19], effectively replacing all non-Y nuclear and cytoplasmic DNA (so are thus y^1/Y; bw^1; e^1; ci^1 ey^1). In this way, we ensured that any observed satellite copy number differences were linked to the Y chromosome. Compared to Y, Ohio (our reference genotype for this experiment), AACAC levels were significantly higher (~130%) on both Y, Congo and Y, Zimbabwe (P = 0.033 and 0.008, respectively, using Student's t-test), while Y, Congo possessed relatively fewer copies of both AAGAC (~79%) and AAGAG (~75%) (P = 0.038 and 0.037, respectively). No significant difference was observed in Y, Zimbabwe AAGAC or AAGAG copy numbers compared to Y, Ohio (p = 0.098 and 0.862, respectively) (Figure 3A).

To support these findings and compare our approach to alternative techniques, we used fluorescence *in situ* hybridization to detect AACAC sequences in larval neuroblast nuclei (Figure 3B). Integration of data from ninety nuclei (thirty nuclei each from three separate brains dissected from sibling males) were largely consistent with our qPCR results: we confirmed significantly more AACAC in Y, Congo and Y, Zimbabwe compared to Y, Ohio (p = 0.008 and 0.036). The error bars in Figure 3C report standard deviation of integrated fluorescence from each nucleus and highlight the difficulty in quantification using fluorescence hybridization, which is prone to vagaries in hybridization, photobleaching, and chromosome spread quality. It is therefore impossible to say whether the differences in intensity are due to differences in cell-specific loss of AACAC copies or due to errors introduced during the procedure.

Mutations Can Alter Repetitive DNA Copy Number

Several models exist to explain repetitive sequence copy number variation of the type that we see in wild-caught Y chromosomes. Polymerase slippage during replication is thought to be responsible for the changes in of simple sequence tracts while interchromosomal and intrachromosomal recombination events account for the gain or loss of larger portions of repetitive sequence [48–50]. Aberrant recombination in particular may be a common mechanism linking copy number variation to the type of genomic instability observed at other repetitive arrays [51]. In *Drosophila*, rDNA stability is regulated by a variety of chromatin factors (*e.g.* Histone H3 Lysine-9 methyltransferase, HP1a, DCR-2, CTCF) [26,40,52]. Removal of these factors by mutation results in genomic instability, increased damage, and repair defects in heterochromatin and copy number changes [40,52,53].

To determine if mutations that alter heterochromatin-induced position effect variegation, rDNA expression, and rDNA stability also affect other satellite DNA copy numbers, we exposed our standard Y chromosome (Y, $10B$ y^+) to a mutation hypothesized to destabilize heterochromatic repeats. The $Su(var)205$ gene encodes Heterochromatin Protein 1a (HP1a), which is enriched at sites of heterochromatin and is required for heterochromatic silencing [54,55]. Notably, it is also required to maintain genomic stability in heterochromatin, and is involved in DNA repair of those sites [40,56]. Given these properties, we hypothesized that the $Su(var)205$ mutation might act dominantly and induce satellite DNA copy number changes on Y, $10B$ y^+. We placed a Y, $10B$ y^+ into a $Su(var)205^{05}/+$ mutant background and maintained it without selection for approximately 150 generations (approximately 6 years). In parallel we maintained a control Y, $10B$ y^+ in a wild-type y^1 w^{67c23} background. After this, we moved the control Y, $10B$ y^+ and the six-year $Su(var)205$ "tempered" counterpart (Y, $10B^{t205}$) into the same isogenic background as above (y^1; bw^1; e^1;
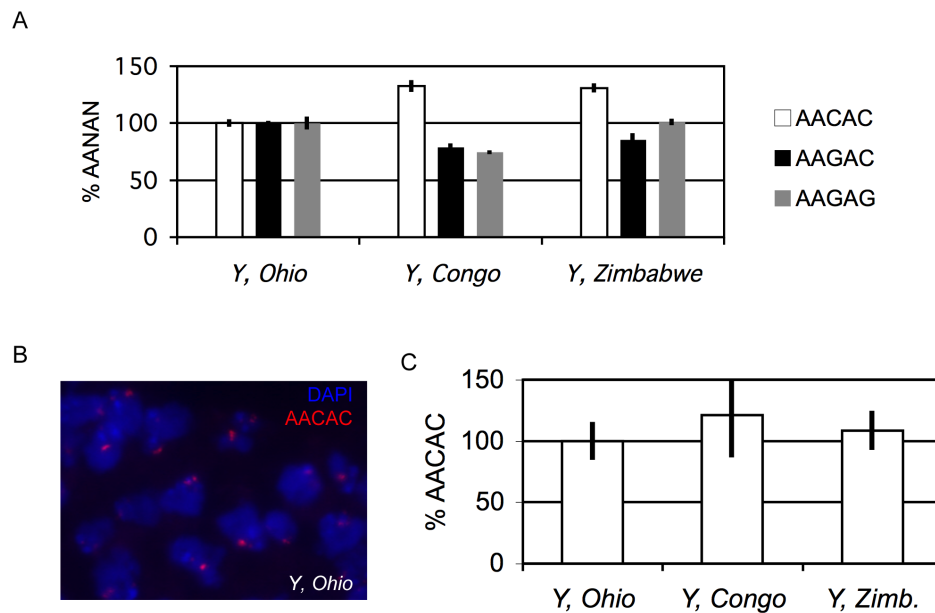
Figure 3. Quantification of Y-linked satellite copy number variation in geographically divergent lines. DNA and tissue samples obtained from males bearing Y chromosomes originally isolated from wild-caught flies. The genetic background of these males was otherwise isogenic. (**A**) Relative quantification of satellite copy number using qPCR. Percentages are relative to Y, Ohio (defined as 100%). Error bars represent standard error of the mean (S.E.M.) of triplicate qPCR reactions. (**B**) Fluorescence *in situ* hybridization to detect AACAC repeats (red) in squashed neuroblast cells derived from Y, Ohio larvae. DAPI stains DNA blue. (**C**) Quantification of *in situ* hybridization signals. Percentages calculated relative to Y, Ohio (defined as 100%). Error bars represent standard deviation (S.D.) of nuclei from thirty neuroblasts from each of three separate preparations per genotype (N = 90).
doi:10.1371/journal.pone.0109906.g003
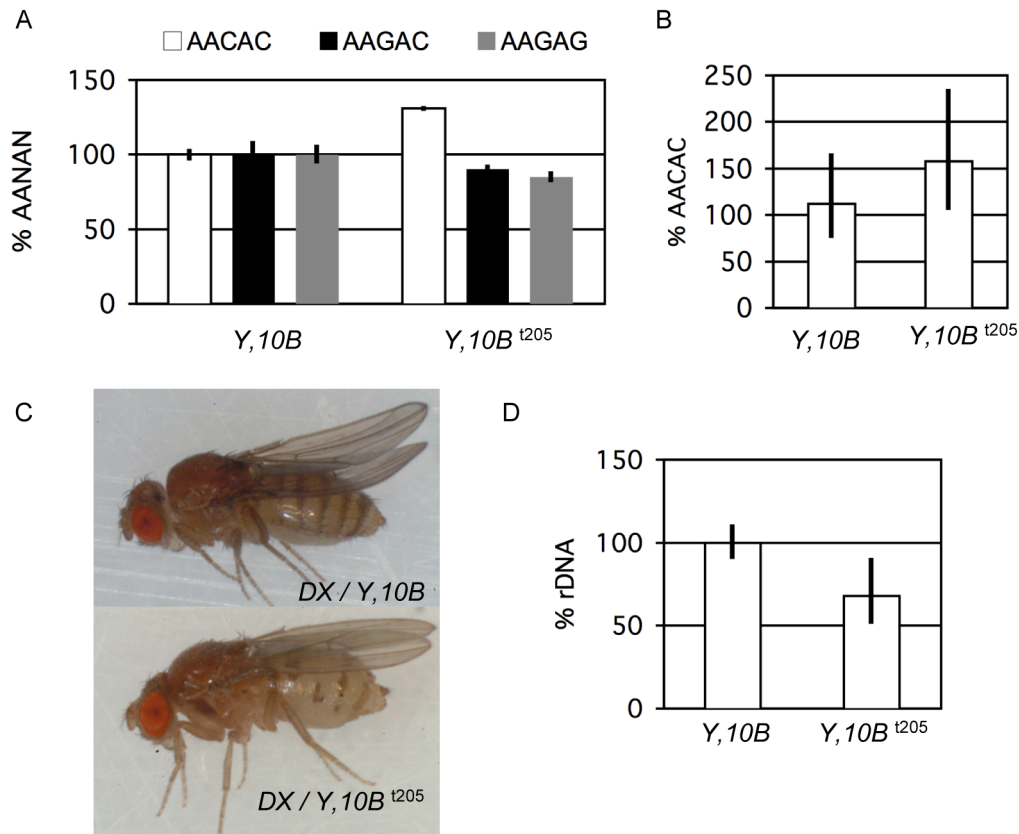
Figure 4. Quantification of satellite copy number variation after a exposure to *Su(var)205*[05]/+ mutation. (A) Relative satellite copy number on *Y, 10B*[t205] compared to *Y, 10B* (defined as 100%). The chromosomes are originally from a single progenitor, but the former was maintained for 6 years in a *Su(var)205*[5]/*CyO* mutant background. Error bars represent standard error of the mean (S.E.M.) of quadruplicate qPCR reactions. (B) DNA from four individual males bearing *Y, 10B*[t205] were separately prepared and used as template for qPCR. Graphs show population average, error bars represent standard deviation (S.D.) of individuals within populations pooled with standard error of the mean (S.E.M.) of replicate reactions. Average AACAC copy number (in A) are comparable to the average of the population (in B), and population distributions remain detectably different (10B *vs*. 10B[t205] in (B), P = 0.0336). (C) Images of female flies of genotype *C(1)DX/Y, 10B* (top) and *C(1)DX/Y, 10B*[t205] (bottom). The fly with *10B*[t205] as sole source of *rDNA* exhibits a strong bobbed phenotype, indicating significant *rDNA* loss. (D) qPCR determination of *rDNA* copy number in the flies from (C). Error bars report standard error of the mean of replicate qPCR reactions from pooled siblings.
doi:10.1371/journal.pone.0109906.g004

ci[1] *ey*[1]) and quantified satellite copy number of AACAC, AAGAC, and AAGAG. We discovered that *Y, 10B*[t205] had ~31% more AACAC compared to *Y, 10B* (p = 0.007) and apparent but nonsignificant decreases in AAGAC and AAGAG (p = 0.300 and 0.168, respectively) (Figure 4A).

To address the difference between individual chromosomes tempered by the *Su(var)205* mutation, we isolated DNA from four individual *y*[1]/*Y, 10B*[t205]; *bw*[1]; *e*[1]; *ci*[1] *ey*[1] males and performed qPCR for AACAC quantification on each. In each case, standard errors of the mean for replicate AACAC and tRNA reactions were pooled along with standard deviation calculations, and averages and pooled standard deviations of the four individuals are shown in Figure 4B. The variance around the mean is higher than when using pooled populations of flies (Figure 4A), however the averages of both *Y, 10B* and *Y, 10B*[t205] are comparable whether using pools (Figure 4A) or individuals (Figure 4B), and in both cases the difference in AACAC copy number between *Y, 10B* and *Y, 10B*[t205] is statistically robust (Figure 4B, P = 0.0036).

Peng and Karpen have previously observed that mutations in *Su(var)205* destabilize the *rDNA* [40], and they, Greil and Ahmad [33], and we [26] have shown that *Drosophila* strains with mutations in the methyltransferase responsible for creating the histone modification to which HP1a binds (*Su(var)3–9*) have few

rDNA. We therefore expected that in addition to destabilizing AACAC, and potentially AAGAC and AAGAG, *rDNA* copy number would be different between *Y, 10B y*[+] and *Y, 10B*[t205]. Crossing these two chromosomes to females bearing a compound X chromosome devoid of *rDNA* (*C(1)DX, y*[1] *f*[1] *bb*[0]) revealed that the latter expressed a bobbed phenotype of etched and herniated abdominal dorsal cuticle, the manifestation of reduced translational capacity from reduced *rDNA* copy number (Figure 4C). *rDNA* copy number quantification using qPCR confirmed a loss of *rDNA* in the *Y, 10B*[t205] chromosome (Figure 4D). Hence, exposure to *Su(var)205* mutation affects other repetitive DNAs of the *Y* chromosome.

Discussion

A number of methods currently exist for determining the copy number of satellite DNAs –the repetitive simple sequences that comprise nearly half of most eukaryotic genomes. These methods include quantification using fluorescence *in situ* hybridization [57], hybridization blots [7], and next-generation sequencing [58]. Each has benefits and drawbacks, therefore none are ideal, but all are useful depending on the specific investigation and limitations. The Real-Time Quantitative PCR (qPCR) technique adapted for

this study is simple in that it requires only routine DNA purification, two specially-designed satellite-specific primers, two "denominator" comparison primers, and is mathematically simple to calculate relative amplifications. With the growing awareness that repetitive satellite DNA in centric constitutive heterochromatin may be linked to ecological variation or disease proclivity, this technique fills a large and growing need. The approach we describe here is simple to perform, robust to fluctuations in DNA concentration or preparation, sensitive to small changes (we estimate ~5% based on standard error) in satellite repeat copy number, very low-cost and rapid.

The total time from living organism to data is less than one day, making it rapid and useful for most purposes. The ability to perform analyses using as little as one nanogram of genomic DNA also allows independent assessment of satellite copy number in old samples, individuals, or dissected tissues, far below the useful detection limits of Southern blot analyses. The molecular nature allows satellite quantification even in cell types or organisms without established cytology. The rapidity, flexibility, and cost-effective nature of this assay makes it useful to a large number of investigators, even without resources for more expensive approaches (*e.g.*, next-generation sequencing).

The design of primers should be broadly amenable to any satellite repeat. Although we only validate it here for pentameric repeat satellites, design of the mismatches are expected to be easier as the repeat length increases. Provided some foreknowledge of the repeat identities, use of this technique will allow investigators to begin to investigate questions about natural variation in copy number, or mutation- or treatment- induced changes to satellite copy number. To that end, between-satellite comparisons are not valid, nor are determinations of absolute copy number, using this technique. This is evident from the different C_q values in Figure 2A, which we believe to be a function of the parameters of binding, priming, and elongation of different repeat sequences, or other factors that cannot be normalized across different primer or target sequences. However, between-organism comparisons of satellite copy numbers are valid, allowing investigators to determine if mutations or treatments results in copy number variability.

We used both natural ecological variation and mutant analyses to validate our approach. Using qPCR, we noted heretofore undiscovered variation in satellite copy number in natural populations from wild-caught *Y* chromosomes from three different geographical sources. These polymorphisms, and others like them, likely contribute to phenomena such as *Y*-linked Regulatory Variation or the ability of different chromosomes to variably suppress epigenetic heterochromatin-induced position effect variegation *in trans* [23–27].

We also discovered that a mutation in the *Su(var)205* gene, which encodes HP1a, results in satellite instability of a subset of repeat types. Previous cytological work showed that *Su(var)205* mutation, and a histone methyltransferase in the same chromatin modification pathway (*Su(var)3–9*) both act dominantly to cause nucleolar (*rDNA*) instability [26,33,40,52]. Moreover, the amount of damage (judged by repair foci in interphase cells) suggested that the damage was more widespread than just the *rDNA* [52]. Since it had not been mapped, it was undetermined if damage induced by *Su(var)205* heterozygosity was limited to the soma or could affect germ cells, and thus be a source of satellite variability in natural populations.

We showed that a chromosome maintained long-term in a mutant of *Su(var)205* was induced to alter satellite copy number (Figure 4). This finding was striking because it shows that mutations thought to act "epigenetically" may also act by altering chromosome structures at places that have not yet been investigated. HP1 appears to bind to all cytological heterochromatin, so discovery that AACAC was increased in copy number while AAGAC and AAGAG were reduced was not predicted. We imagine three possible explanations. First, it is possible that HP1 acts to stabilize some satellite sequences while destabilizing others. A mechanism for the former is apparent from the role of HP1 in establishing a chromatin structure conducive to silencing. The latter, while it has no obvious mechanism, is nonetheless logically consistent with our observation. Second, it is possible that *Su(var)205* stabilizes all satellite sequences, but loss of destabilized sequence is not the only consequence of instability. For example, while destabilizing may frequently lead to loss of DNA through intrachromosomal or interchromosomal recombination or damage/repair, it may also lead to increases. Mechanistically, this could be from one segregation product of any interchromosomal recombination event, but additionally increases in copy number could be accomplished by replication-coupled polymerase slippage, rolling-circle replication, re-replication, or some unknown programmed event. Third, it is possible that only a subset of satellites are affected by mutation in *Su(var)205* (*e.g.*, the *rDNA*), but there exists communication between different types of satellite. For example, a decrease in *rDNA* copy number alters heterochromatin formation, which results in selective pressure to enlarge other heterochromatic components to compensate. This idea, unencumbered by data, imagines that different forms of satellite are in balance in the genome and perturbations of one type will cause new equilibria to be reestablished by expansion or contraction of other interacting types.

Our anecdotal experience has been that stocks of some mutations – *Su(var)205* and *Su(var)3–9* among others – become stronger in their abilities to suppress variegation (their eponymous phenotype) after being established. While others have noted this, it has been informally accepted to be by selection as a consequence of the small populations and conditions of fly stock maintenance, we can now suggest that these mutations may also (or instead) induce copy number changes to unlinked satellite sequences, which themselves permanently alter phenotypes. The extent and consequence of such changes are unknown, but with qPCR copy number determination they may now be pursued.

Materials and Methods

Fly husbandry and stocks

All crosses and stocks were maintained on standard cornmeal molasses media at 25°C. *X/Y/Y* males were generated by crossing spontaneously occurring $y^1\ w^{67c23}/y^1\ w^{67c23}/Y,\ 10B\ y^+$ or $y^1\ w^{67c23}/y^1\ w^{67c23}/Y,\ B^S$ female primary nondisjunctants to $y^1\ w^{67c23}/Y,\ B^S$ or $y^1\ w^{67c23}/Y,\ 10B\ y^+$ males, respectively. For the former, $y^1\ w^{67c23}/y^1\ w^{67c23}/Y,\ B^S$ virgins were crossed to $y^1\ w^{67c23}/Y,\ 10B\ y^+$, then $y^1\ w^{67c23}/Y,\ 10B\ y^+/Y,\ B^S$ male offspring backcrossed to $y^1\ w^{67c23}$ to create and maintain secondary nondisjunctional strains which produce large numbers of *X/Y/Y* males. *X/Y/Y* were distinguished from their *X/Y* siblings by the severity of the Bar-stone or yellow+ phenotypes. Geographically diverse *Y* chromosomes (*Y, Ohio, Y, Congo*, and *Y, Zimbabwe*) were obtained from Bernardo Lemos [23]. Chromosomes were placed in an isogenic background by crossing males to $y^1;\ bw^1;\ e^4;\ ey^R$ females and backcrossing to the maternal genotype until all four recessive markers were made homozygous. $y^1\ w^{67c23}/Dp(1;Y)\ y+,\ P\{w = RSw\}10B\ (Y,\ 10B\ y^+)$ is described in [59]. $Y,\ 10B^{t205}$ was generated by maintaining *Y, 10B* in a $y^1\ w^{67c23};\ Su(var)205^{05}/CyO$ background for approximately six years. Prior to quantifica-

tion, *Y, 10B* and *Y, 10B*[t205] were placed in an isogenic *y¹; bw¹; e⁴; ey^R* background as above.

Quantitative Polymerase Chain Reaction

Quantitative PCR (qPCR) analysis was performed with 12 μL reactions as described in [31]. DNA was extracted from adult flies homogenized in pools of ten and quantified using a NanoDrop ND-1000. 10 nanograms was used for each reaction (except where indicated). DNA from individual flies did not perform well in our reactions; to circumvent this problem the reactions in Figure 4B used 0.1 ng template. qPCR was performed with a StepOne Real-time PCR system and Power SYBR Green reagents (Applied Biosciences). The following conditions were used for 40 cycles: 95°C for 3 s; 50°C for 15 s; 60°C for 30 s. Relative differences were calculated using the "ΔΔCT"method. Each data point represents an average obtained from three or four qPCR reactions. P-values were calculated from untransformed ΔCq values using Student's t-test.

Satellites were amplified using primers designed according to [37]. AACAC: GGTTTACACTACACATCACAAGACAACT-CAACACAGCA and ACTCCAGTTG- TATTGT-GATGTGTGGTGTTATGTTGTGC; AAGAC: GGTTTTAGCCAAGAGAA-GACCAGACACGACAACA-CAAGACTA and ACTCCATCTTGCCTTGTTTTGTC-CTGTCTCGTCTTTTCTTGCCTTGTCTA; AAGAG: GGTTTTAGAAGTGAAGAT-AAGAGTAGAGATGAGAA-GACAA and ACTCCATCTCTACTCTCTTGTCTTCA-CTTCTGTTCTCTT. The endogenous control, tRNA[K-CTT], was amplified using CTAGCTCAGTCGGTAGAGCATGA and CCAACGTGGGGCTCGAAC. Primers were used at a concentration of 0.5 μM.

Fluorescent *in situ* hybridization (FISH) and Microscopy

Fluorescence probe was made by end-labeling oligonucleotides of the respective satellite repeat (*e.g.*, AACACAACACAACA-CAACACAACACAACAC) with digoxigenin-conjugated dUTP, and visualized with a mouse anti-digoxigenin antibody conjugated to rhodamine. Dissections, tissue preparation, and hybridizations were performed as described by Larracuente and Ferree (submitted). DNA was counterstained with 1 ng/mL DAPI (MP Biomedicals). All images were obtained using a Zeiss Axioskop 2 epifluorescence microscope running AxioVision (v. 4.6.3.0) with a 20X objective (numerical aperature = 0.5). Sequential excitation was performed at 543 nm (for Rhodamine) and 405 nm (for DAPI).

Supporting Information

Figure S1   Melt Curve Analysis of qPCR Primer Sets. First derivative with respect to temperature of Relative Fluorescence Units (RFU) through the indicated temperature range. Derivative was calculated by ΔY/ΔX for each temperature interval after maximal fluorescence was set at 100%. Single major peaks indicate monophasic melting, indicative of single qPCR products with relatively-homogenous melting profiles. (TIFF)

References

1. Yunis JJ, Yasmineh WG (1971) Heterochromatin, satellite DNA, and cell function. Structural DNA of eucaryotes may support and protect genes and aid in speciation. Science 174: 1200–1209.
2. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, et al. (2000) The genome sequence of Drosophila melanogaster. Science 287: 2185–2195.
3. Hoskins RA, Carlson JW, Kennedy C, Acevedo D, Evans-Holm M, et al. (2007) Sequence finishing and mapping of Drosophila melanogaster heterochromatin. Science 316: 1625–1628.
4. Peacock WJ, Brutlag D, Goldring E, Appels R, Hinton CW, et al. (1974) The organization of highly repeated DNA sequences in Drosophila melanogaster chromosomes. Cold Spring Harb Symp Quant Biol 38: 405–416.
5. Long EO, Dawid IB (1980) Repeated genes in eukaryotes. Annu Rev Biochem 49: 727–764.
6. Gatti M, Pimpinelli S (1992) Functional elements in Drosophila melanogaster heterochromatin. Annu Rev Genet 26: 239–275.
7. Lohe AR, Hilliker AJ, Roberts PA (1993) Mapping simple repeated DNA sequences in heterochromatin of Drosophila melanogaster. Genetics 134: 1149–1174.
8. Krassovsky K, Henikoff S (2014) Distinct chromatin features characterize different classes of repeat sequences in Drosophila melanogaster. BMC Genomics 15: 105.
9. Dernburg AF, Sedat JW, Hawley RS (1996) Direct evidence of a role for heterochromatin in meiotic chromosome segregation. Cell 86: 135–146.
10. Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. Science 293: 1098–1102.
11. Sage BT, Csink AK (2003) Heterochromatic self-association, a determinant of nuclear organization, does not require sequence homology in Drosophila. Genetics 165: 1183–1193.
12. Sun X, Le HD, Wahlstrom JM, Karpen GH (2003) Sequence analysis of a functional Drosophila centromere. Genome Res 13: 182–194.
13. Ferree PM, Barbash DA (2009) Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in Drosophila. PLoS Biol 7: e1000234.
14. Ferree PM (2014) Mitotic misbehavior of a Drosophila melanogaster satellite in ring chromosomes: insights into intragenomic conflict among heterochromatic sequences. Fly (Austin) 8.
15. Eichler EE, Clark RA, She X (2004) An assessment of the sequence gaps: unfinished business in a finished human genome. Nat Rev Genet 5: 345–354.
16. Kapranov P, St Laurent G (2012) Dark Matter RNA: Existence, Function, and Controversy. Front Genet 3: 60.
17. Graur D, Zheng Y, Price N, Azevedo RB, Zufall RA, et al. (2013) On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. Genome Biol Evol 5: 578–590.
18. Dimitri P, Pisano C (1989) Position effect variegation in Drosophila melanogaster: relationship between suppression effect and the amount of Y chromosome. Genetics 122: 793–800.
19. Carvalho AB (2002) Origin and evolution of the Drosophila Y chromosome. Curr Opin Genet Dev 12: 664–668.
20. Bonaccorsi S, Lohe A (1991) Fine mapping of satellite DNA sequences along the Y chromosome of Drosophila melanogaster: relationships between satellite sequences and fertility factors. Genetics 129: 177–189.
21. Chippindale AK, Rice WR (2001) Y chromosome polymorphism is a strong determinant of male fitness in Drosophila melanogaster. Proc Natl Acad Sci U S A 98: 5677–5682.
22. Rohmer C, David JR, Moreteau B, Joly D (2004) Heat induced male sterility in Drosophila melanogaster: adaptive genetic variations among geographic populations and role of the Y chromosome. J Exp Biol 207: 2735–2743.
23. Lemos B, Araripe LO, Hartl DL (2008) Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. Science 319: 91–93.
24. Jiang PP, Hartl DL, Lemos B (2010) Y Not a Dead End: Epistatic Interactions Between Y-linked Regulatory Polymorphisms and Genetic Background Affect Global Gene Expression in Drosophila melanogaster. Genetics.
25. Lemos B, Branco AT, Hartl DL (2010) Epigenetic modulation of gene expression by polymorphic Y chromosomes: Sexual conflict, chromatin components, and immune response. Proc Natl Acad Sci U S A.
26. Paredes S, Maggert KA (2009) Ribosomal DNA contributes to global chromatin regulation. Proc Natl Acad Sci U S A 106: 17829–17834.

27. Paredes S, Branco AT, Hartl DL, Maggert KA, Lemos B (2011) Ribosomal DNA deletions modulate genome-wide gene expression: "rDNA-sensitive" genes and natural variation. PLoS Genet 7: e1001376.

28. Bosco G, Campbell P, Leiva-Neto JT, Markow TA (2007) Analysis of Drosophila species genome size and satellite DNA content reveals significant differences among strains as well as between species. Genetics 177: 1277–1290.

29. Lindsley DL, Sandler L, Baker BS, Carpenter AT, Denell RE, et al. (1972) Segmental aneuploidy and the genetic gross structure of the Drosophila genome. Genetics 71: 157–184.

30. Spofford JB, DeSalle R (1991) Nucleolus organizer-suppressed position-effect variegation in Drosophila melanogaster. Genet Res 57: 245–255.

31. Paredes S, Maggert KA (2009) Expression of I-CreI Endonuclease Generates Deletions Within the rDNA of Drosophila. Genetics 181: 1661–1671.

32. Pietrzak M, Rempala G, Nelson PT, Zheng JJ, Hetman M (2011) Epigenetic silencing of nucleolar rRNA genes in Alzheimer's disease. PLoS One 6: e22585.

33. Greil F, Ahmad K (2012) Nucleolar Dominance of the Y Chromosome in Drosophila melanogaster. Genetics.

34. Zhou J ST, Martinsen L, Lemos B, Eickbush TH, Hartl DL (2012) Y chromosome mediates ribosomal DNA silencing and modulates the chromatin state in Drosophila. Proc Natl Acad Sci U S A 10.1073/pnas.1207367109.

35. Maggert KA (2014) Reduced rDNA Copy Number Does Not Affect "Competitive" Chromosome Pairing in XYY Males of Drosophila melanogaster. G3 (Bethesda).

36. Cawthon RM (2002) Telomere measurement by quantitative PCR. Nucleic Acids Res 30: e47.

37. Cawthon RM (2009) Telomere length measurement by a novel monochrome multiplex quantitative PCR method. Nucleic Acids Res 37: e21.

38. Ritossa FM (1968) Unstable redundancy of genes for ribosomal RNA. Proc Natl Acad Sci U S A 60: 509–516.

39. Lyckegaard EM, Clark AG (1989) Ribosomal DNA and Stellate gene copy number variation on the Y chromosome of Drosophila melanogaster. Proc Natl Acad Sci U S A 86: 1944–1948.

40. Peng JC, Karpen GH (2007) H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. Nat Cell Biol 9: 25–35.

41. Cuadrado A, Jouve N (2011) Novel simple sequence repeats (SSRs) detected by ND-FISH in heterochromatin of Drosophila melanogaster. BMC Genomics 12: 205.

42. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, et al. (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem 55: 611–622.

43. Ramakers C, Ruijter JM, Deprez RH, Moorman AF (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Lett 339: 62–66.

44. Bridges CB (1916) Non-Disjunction as Proof of the Chromosome Theory of Heredity. Genetics 1: 1–52.

45. Takeuchi Y, Horiuchi T, Kobayashi T (2003) Transcription-dependent recombination and the role of fork collision in yeast rDNA. Genes Dev 17: 1497–1506.

46. Akkaya MS, Bhagwat AA, Cregan PB (1992) Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132: 1131–1139.

47. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, et al. (2002) Genetic structure of human populations. Science 298: 2381–2385.

48. Tartof KD (1973) Regulation of ribosomal RNA gene multiplicity in Drosophila melanogaster. Genetics 73: 57–71.

49. Tartof KD (1974) Unequal mitotic sister chromatid exchange and disproportionate replication as mechanisms regulating ribosomal RNA gene redundancy. Cold Spring Harb Symp Quant Biol 38: 491–500.

50. Messier W, Li SH, Stewart CB (1996) The birth of microsatellites. Nature 381: 483.

51. Cohen S, Yacobi K, Segal D (2003) Extrachromosomal circular DNA of tandemly repeated genomic sequences in Drosophila. Genome Res 13: 1133–1145.

52. Peng JC, Karpen GH (2009) Heterochromatic genome stability requires regulators of histone H3 K9 methylation. PLoS Genet 5: e1000435.

53. Peng JC, Karpen GH (2008) Epigenetic regulation of heterochromatic DNA stability. Curr Opin Genet Dev 18: 204–211.

54. James TC, Elgin SC (1986) Identification of a nonhistone chromosomal protein associated with heterochromatin in Drosophila melanogaster and its gene. Mol Cell Biol 6: 3862–3872.

55. Vermaak D, Malik HS (2009) Multiple roles for heterochromatin protein 1 genes in Drosophila. Annu Rev Genet 43: 467–492.

56. Chiolo I, Minoda A, Colmenares SU, Polyzos A, Costes SV, et al. (2011) Double-strand breaks in heterochromatin move outside of a dynamic HP1a domain to complete recombinational repair. Cell 144: 732–744.

57. Lecuyer E, Parthasarathy N, Krause HM (2008) Fluorescent in situ hybridization protocols in Drosophila embryos and tissues. Methods Mol Biol 420: 289–302.

58. Lu Y, Soong TD, Elemento O (2013) A novel approach for characterizing microsatellite instability in cancer cells. PLoS One 8: e63056.

59. Maggert KA, Golic KG (2005) Highly efficient sex chromosome interchanges produced by I-CreI expression in Drosophila. Genetics 171: 1103–1114.