Dosage Compensation Via Transposable Element Mediated Rewiring of a Regulatory Network

Christopher E. Ellison and **Doris Bachtrog**
Department of Integrative Biology, University of California Berkeley, Berkeley, CA 94720, USA

Abstract

Transposable elements (TEs) may contribute to evolutionary innovations through the rewiring of networks by supplying ready-to-use *cis* regulatory elements. Genes on the *Drosophila* X chromosome are coordinately regulated by the Male Specific Lethal (MSL) complex to achieve dosage compensation in males. We show that the acquisition of dozens of MSL binding sites on evolutionarily new X chromosomes was facilitated by the independent co-option of a mutant helitron TE that attracts the MSL complex (i.e. TE domestication). The recently formed neo X recruits helitrons that provide dozens of functional, but suboptimal, MSL binding sites, while the older XR chromosome has ceased acquisition and appears to have fine-tuned the binding affinities of more ancient elements for the MSL complex. Thus, TE mediated rewiring of networks through domestication and amplification may be followed by fine-tuning of the *cis*-regulatory element supplied by the TE and erosion of non-functional regions.

Active transposable elements (TEs) impose a significant mutational burden upon the host genome (*1-4*). However, there is growing evidence implicating TEs as drivers of key evolutionary innovations, by creating or re wiring regulatory networks (*5-11*). Many TEs harbor a variety of regulatory motifs and TE amplification may allow for the rapid accumulation of a specific motif throughout the genome, thus recruiting multiple genes into a single regulatory network (*12*).

In *Drosophila miranda*, multiple sex chromosome/autosome fusions have created a series of X chromosomes of differing ages (Fig. 1). The ancestral X chromosome, XL, is homologous to the *D. melanogaster* X and is at least 60 MY old (*13*). Chromosome XR became a sex chromosome about 15 MY ago and is shared among members of the *affinis* and *pseudoobscura* subgroups, while the neo X chromosome is specific to *D. miranda* and originated only 1 MY ago (*14, 15*). The Male Specific Lethal (MSL) complex coordinates gene expression on the Drosophila male X to achieve dosage compensation (*16*). This complex is recruited to the X chromosome in males to high affinity chromatin entry sites (CES) containing a conserved, roughly 21 bp long GA rich sequence motif termed the MSL

Recognition Element (MRE) (*17*). Once bound, the MSL complex spreads from the CES in *cis* to actively transcribed genes where it catalyzes the deposition of the activating histone modification H4K16ac, which ultimately results in a chromosome wide two fold increase in gene expression levels (*16*). *D. miranda* males show MSL binding specific to the X chromosomes, associated with full dosage compensation of chromosomes XL and XR. In contrast, the neo X shows incomplete dosage compensation (*18*).

The evolution of dosage compensation on XR and the neo X involved co option of the MSL machinery (*19*) and the creation of CES capable of recruiting this machinery, via MRE sequence motifs at a few hundred locations along the two X chromosomes. We used ChIP-seq profiling of MSL binding to conservatively define 132 CES on chromosome XL, 215 on XR, and 68 on the neo X (*18*), and a more realistic estimate identifies 219 CES on XL, 383 on XR, and 175 on the neo X [Fig. S1, (*20*)], and we refer to these two groups as our 'strict' versus 'broad' set of CES. The CES on XR and the neo X likely arose within the past 15 and 1 MY, respectively, after these chromosomes became X linked in an ancestor of *D. miranda*.

Comparison of the genomic regions at strict neo X CES sequences to their homologous regions in *D. pseudoobscura*, which are not X linked and do not recruit the MSL complex, identified the mutational paths responsible for the novel formation of a MRE at 41 CES on the neo X (*21*). In half of these sites, point mutations and short indels at pre binding sites created a stronger MRE. For the remaining half, however, the novel MREs appeared to have been gained via a relatively large (~1kb), *D. miranda* specific insertion. Sanger re-sequencing and manual curation of the genome assembly at these sites allowed us to determine that these insertions are derived from a transposable element [homologous to the ISY element (*22*)], that is highly abundant in the genome of *D. miranda* and its relatives (>1000 copies in *D. miranda* and *D. pseudoobscura*, Fig. 2A). The ISY element (~1150 bp) is a non autonomous helitron [Fig. S2, S3, (*20*)], a class of DNA transposable elements that replicate through a rolling circle mechanism (*23, 24*). All 21 elements found at strict CES on the neo X share a 10 bp deletion relative to the consensus ISY element, and we refer to the ISY sequence containing this deletion as ISX (Figs. 3A, S4). ISX is also found at 24 of our broad CES, and it is present at 43% of strict CES and 30% of broad CES on the neo-X [Fig. S5, S6, (*20*)]. Importantly, this 10 bp deletion creates a sequence motif more similar to the consensus MRE motif inferred from XL relative to the consensus ISY sequence (Fig. 2B), and thus might create a strong recruitment signal for the MSL complex. The ISX element – but not ISY-is unique to *D. miranda* and highly enriched on the neo X relative to other chromosomes (Figs. 3C, S7), and strongly bound by the MSL complex *in vivo* (Fig. 2C). Additionally,the sequence similarity among ISX elements found at CES on the neo X (Fig. 3A,B) is consistent with their recent acquisition on the neo X, after the formation of the neo sex chromosomes (*20*). Together, these results suggest that within the past 1 MY the *D. miranda* lineage was invaded by a domesticated helitron that recruits hundreds of genes into the MSL regulatory network on the neo-X. This process involved the formation of a high affinity MRE sequence motif via a 10 bp deletion, followed by amplification and fixation of this element at dozens of sites along the neo X chromosome [Fig. S8, S9, (*20*)].

We used a transgenic assay in *D. melanogaster* to functionally verify that the ISX element attracts the MSL complex and functions as a CES. We targeted our construct to the

previously characterized autosomal landing site 37B7 in *D. melanogaster* (*25*). Immunostaining of male polytene chromosomes shows that the ISX element can recruit the MSL complex of *D. melanogaster*, but no staining was detected with the ISY element (Figs. 2D,E, S10, S11). A higher affinity of the MSL complex to ISX vs. ISY was also confirmed by ChIP qPCR (Fig. S12). We also used mutagenesis assays to convert this ISX element into ISY by inserting the 10-bp sequence (ISX → ISY), and deleted the 10-bp fragment from the ISY element to create ISX (ISY → ISX). Immunostaining confirmed that the ISX → ISY construct could no longer recruit the MSL-complex to an autosomal location, while the ISY → ISX transgene was now able to attract MSL to an autosomal landing site in *D. melanogaster* (Fig. S13). Thus, the ISX element alone is able and sufficient to attract the MSL complex, and the 10-bp deletion creates a functional MSL recruitment site. This experimentally confirms that the amplification of this TE along the neo X chromosome may have resulted in the rapid wiring of neo X linked genes into the dosage compensation network. Dosage compensation of neo-X genes is advantageous since about 40% of homologous neo-Y genes are pseudogenized (*26*); however, due to its ability to recruit the MSL complex and induce dosage compensation, the ISX element should be selected against from autosomal locations. Indeed, out of a total of 82 copies of the ISX element, only two exist on an autosome, within repeat-rich and supposedly silenced regions on the dot chromosome [Fig. S14, (*20*)].

In the ancestor of the *affinis* and *pseudoobscura* subgroups (~15 MYA), Muller element D became incorporated into the dosage compensation network after it fused to the ancestral X to form chromosome XR (Fig. 1). We compared all CES sequences on XR to determine if they were enriched for sequence elements besides the MRE motif that would be indicative of a TE burst. Three repeat elements were present in ~22% of strict (and in 14.4% of broad) XR CES sequences, but not in the homologous regions from *D. subobscura*, where this chromosome is an autosome (Fig. 1). Furthermore, these elements were all determined to be conserved fragments from a single TE (hereafter referred to as ISXR), which is derived from the same helitron family as the ISY/ISX elements (Figs. 3A, S15). Individual ISXR copies are less similar to each other than the ISX elements, and sequence divergence among the different copies of this TE is consistent with a burst of transposition activity coinciding with the formation of chromosome XR (Fig. 3B). Additionally, ISXR is enriched on chromosome XR (Fig. 3C), and similar to ISX/ISY, its autosomal homologs show less sequence similarity to the MRE consensus motif and cannot recruit the MSL-complex *in vivo* (Fig. S16). ISXR contains a ~350 bp region that is not present in any of the ISY or ISX elements, and this unique region to ISXR contains an additional MRE motif in close proximity to the MRE whose location is conserved between the ISX and ISXR elements (Figs. 3A, D, S17). In addition, while the location of the 3′ ISXR MRE is conserved with ISX, there is no evidence of the 10-bp deletion seen in ISX. The presence of this unique sequence region suggests that, although ISX and ISXR evolved from a similar helitron progenitor TE, they represent independent TE domestications and chromosomal expansions at different time points [Fig. 3A, S18, (*20*)]. Consistent with the more ancient expansion of ISXR, non-functional parts of the TE are severely eroded (Fig. 3A, S15).

Similarity-based clustering of the MRE consensus motifs from each helitron subtype reveal that both ISXR MRE motifs are more similar to the canonical XL MRE motif, compared to the ISX MRE motif (Fig. 3D). This suggests that MSL binding motifs supplied by ISX may be suboptimal, while ISXR binding affinity is optimized. A large number of substitutions observed at MRE motifs among ISXR copies across the genome [Fig. S19, (*20*)], and elevated rate of evolution at homologous ISXR MRE sites relative to XL MREs across species (Fig. S20) suggests that the ISXR element initially may have also harbored a suboptimal MRE motif (*20*). Over time, mutation and selection may have fine-tuned the nucleotide composition at ISXR independently across elements and species, to maximize MSL recruitment by increasing their similarity to the canonical XL MRE motif (Fig. 3D). In agreement with this observation, the TE derived XR CES show a higher affinity for MSL complex *in vivo* compared to those on the neo X (Fig. 3E).

The recently formed sex chromosomes of *D. miranda* provide insights into the role of TEs in rewiring regulatory networks. The evolutionary pressure driving the acquisition of dosage compensation as well as the molecular mechanism of MSL function and targeting provide clear expectations of which genes should be recruited into the dosage compensation network, and when and how. Additionally, the comparison of XR and the neo X allow us to study the dynamic process of TE mediated wiring of chromosomal segments into the dosage compensation network at two different evolutionary stages; both the initial incorporation of the neo-X chromosome by amplification of a domesticated TE and possible subsequent fine-tuning of the regulatory element supplied by the TE on XR together with the erosion of TE sequence not required for MSL-binding. Our data support a 3 step model for TE mediated rewiring of regulatory networks (domestication, amplification and potential fine tuning) followed by erosion of non functional parts of the transposon (Fig. 4). Eventually, the footprints left behind by TE mediated rewiring will completely vanish, and many ancient bursts of domesticated TEs that rewired regulatorynetworks are likely to go undetected. Indeed, we do not observe any TE relics within the CES of chromosome XL which acquired MSL-mediated dosage compensation over 60 MY ago, either because they evolved via a different mechanism, or deletions and substitutions have degraded the signal of TE involvement to the point where they are no longer recognizable.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Charlesworth B, Charlesworth D. Genetical Research. 1983; 42:1.

2. Doolittle WF, Sapienza C. Nature. 1980; 284:601. [PubMed: 6245369]

3. Hickey DA. Genetics. 1982; 101:519. [PubMed: 6293914]

4. Orgel LE, Crick FH. Nature. 1980; 284:604. [PubMed: 7366731]

5. Lynch VJ, Leclerc RD, May G, Wagner GP. Nature Genetics. 2011; 43:1154. [PubMed: 21946353]

6. Bourque G, et al. Genome research. 2008; 18:1752. [PubMed: 18682548]

7. Kunarso G, et al. Nature genetics. 2010; 42:631. [PubMed: 20526341]

8. Wang T, et al. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104:18613. [PubMed: 18003932]

9. Johnson R, et al. Nucleic acids research. 2006; 34:3862. [PubMed: 16899447]

10. Bringaud F, et al. PLoS Pathogens. 2007; 3:1291. [PubMed: 17907803]

11. Cowley M, Oakey RJ. PLoS Genetics. 2013; 9:e1003234. [PubMed: 23358118]

12. Feschotte C. Nature reviews. Genetics. 2008; 9:397.

13. Richards S, et al. Genome Research. 2005; 15:1. [PubMed: 15632085]

14. Carvalho AB, Clark AG. Science. 2005; 307:108. [PubMed: 15528405]

15. Bachtrog D, Charlesworth B. Nature. 2002; 416:323. [PubMed: 11907578]

16. Conrad T, Akhtar A. Nature reviews. Genetics. 2011; 13:123.

17. Alekseyenko AA, et al. Cell. 2008; 134:599. [PubMed: 18724933]

18. Alekseyenko AA, et al. Genes Dev. 2013; 27:853. [PubMed: 23630075]

19. Marin I, Franke A, Bashaw GJ, Baker BS. Nature (London). 1996; 383:160. [PubMed: 8774878]

20. Supplementary Materials.

21. Zhou Q, et al. PLoS Biology. 2013 in press.

22. Steinemann M, Steinemann S. Proceedings of the National Academy of Sciences of the United States of America. 1992; 89:7591. [PubMed: 1323846]

23. Jurka J. Repbase Reports. 2012; 12:1376.

24. Kapitonov VV, Jurka J. Trends in Genetics. 2007; 23:521. [PubMed: 17850916]

25. Bateman JR, Lee AM, Wu CT. Genetics. 2006; 173:769. [PubMed: 16547094]

26. Zhou Q, Bachtrog D. Science. 2012; 337:341. [PubMed: 22822149]

27. Grant CE, Bailey TL, Noble WS. Bioinformatics. 2011; 27:1017. [PubMed: 21330290]
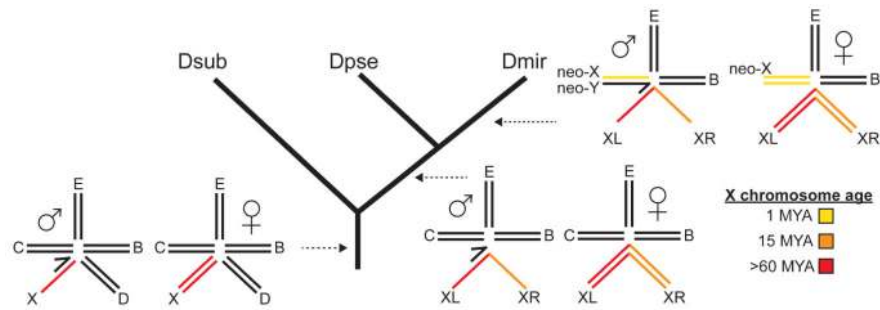
Fig. 1. Evolutionary history of *Drosophila miranda* sex chromosomes

The ancestral X chromosome shared by all members of the Drosophila genus (in red) fused to an autosome roughly 15 MY ago, creating chromosome XR (in orange). Another autosome fused to the Y chromosome about 1 MY ago, creating the neo-X chromosome (in yellow). *D. miranda* thus harbors three X chromosomes of different ages. Species abbreviations are as follows: *D. subobscura* (Dsub), *D. pseudoobscura* (Dpse), and *D. miranda* (Dmir).
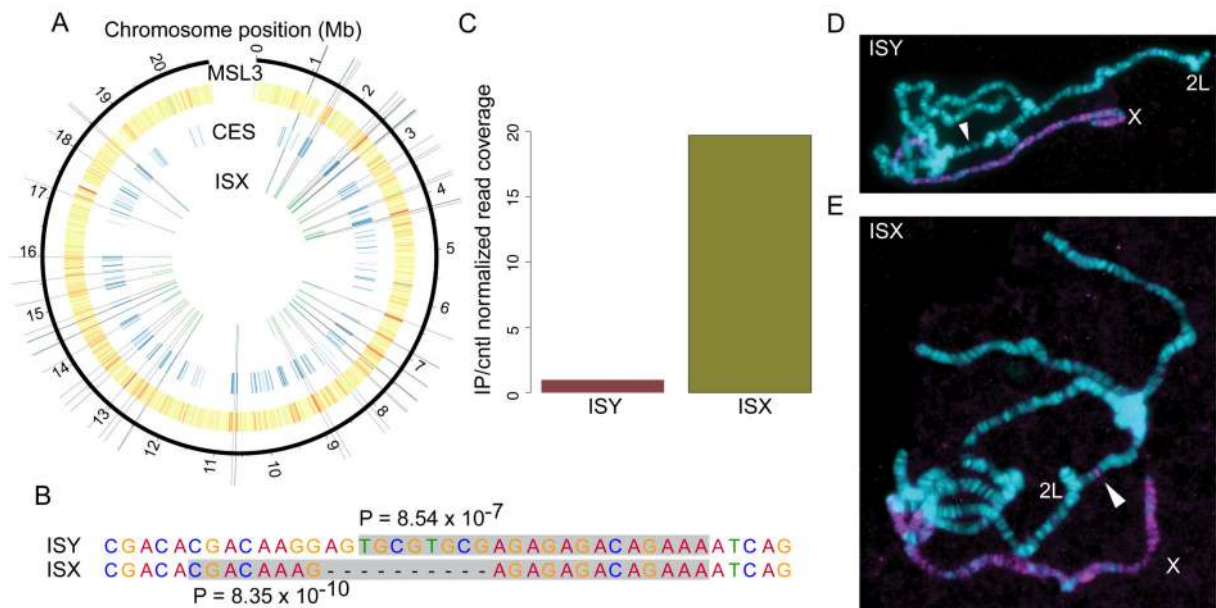
Fig. 2. ISX is a domesticated helitron TE that is associated with chromatin entry sites (CES) on the neo-X chromosome and recruits the MSL complex in transgenic assays

(A) 21 of 68 strict MSL complex CES on the neo-X chromosome overlap *D. miranda* specific insertions of the ISX element, and 69 (out of 77 total) neo-X linked ISX elements lie within broad CES. **(B)** A derived 10 bp deletion differentiates ISX from the related ISY element and creates a stronger match to the MRE consensus motif identified from chromosome XL. P-values are from *FIMO* (*27*) and are based on log-likelihood ratio scores between the *D. melanogaster* canonical MRE consensus motif and the sequence highlighted in grey. **(C)** MSL3 ChIP-seq data show that MSL complex binds ISX but not ISY elements. **(D, E)** Ectopic MSL targeting by the ISX element from *D. miranda*, and lack of activity from the corresponding ISY element. Transgenic polytene chromosomes stained with anti-MSL2 (magenta) to identify regions targeted by the MSL-complex, and DAPI to identify all chromosome arms (cyan). An ISY and ISX element were each targeted to cytosite 37B7 (location denoted by white arrow) on chromosome 2L in *D. melanogaster*. **(D)** No staining is detected at 37B7 when the insertion contains ISY **(E)** but we find robust MSL immunostaining at the location when the insertion contains ISX.
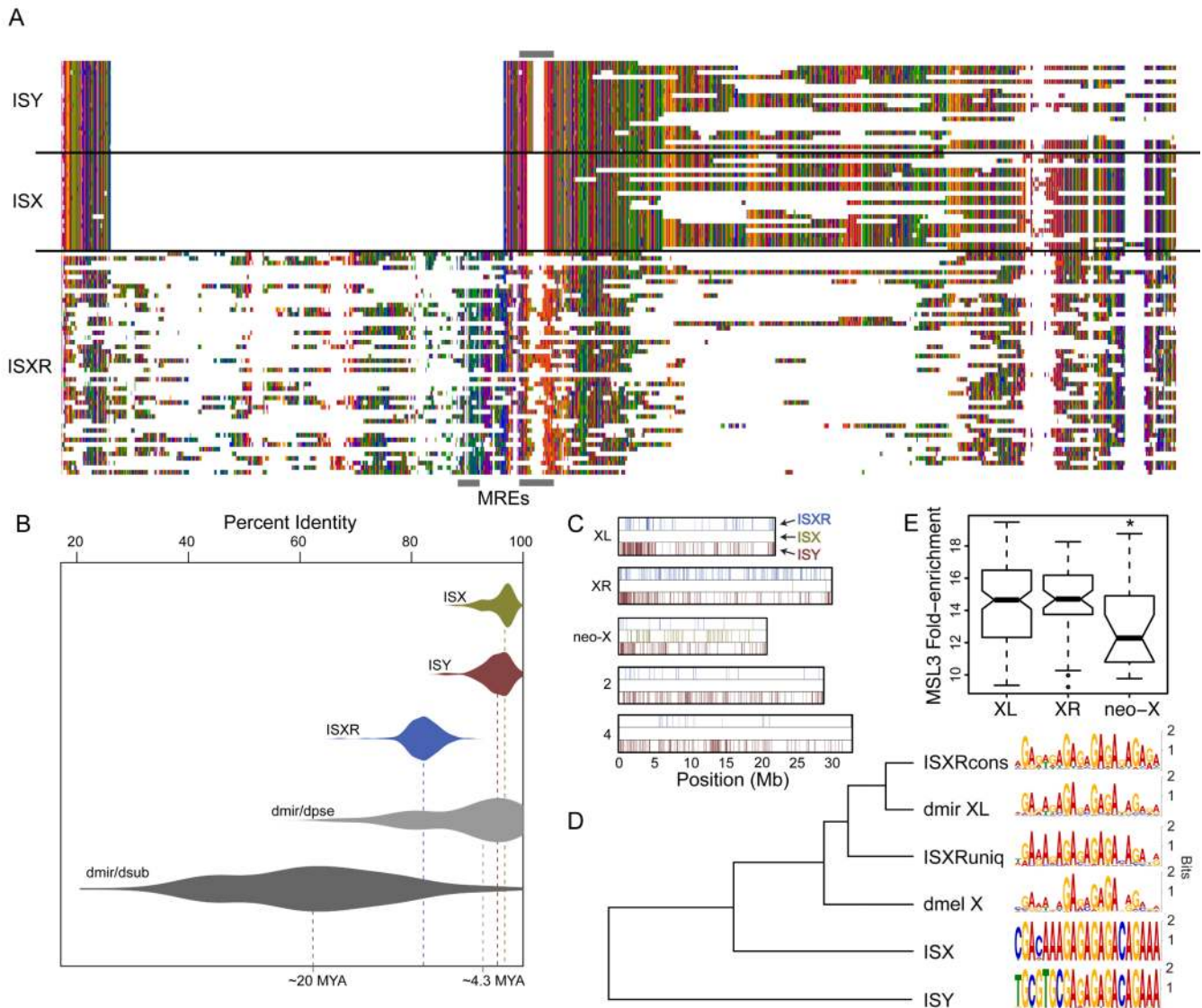
Fig. 3. The ISXR helitron is associated with CES on chromosome XR

(**A**) Multiple sequence alignment showing helitrons from within neo-X CES (ISX), XR CES (ISXR), and the related ISY element. Grey boxes indicate approximate location of the MRE motifs in the alignment. (**B**) Divergence of elements from their consensus sequence places the burst of ISXR amplification after the divergence of *D. subobscura* from the *miranda/pseudoobscura* ancestor (> 15 MYA) and the ISX burst after the divergence of *D. pseudoobscura* and *D. miranda* (~4 MYA). See Supplemental Methods for details. (**C**) The ISY element is distributed evenly among *D. miranda* chromosomes (permutation test p > 0.08 in all cases, except for chromosome 4, where it is depleted [permutation test p<0.0001]) while ISXR is enriched on XR and ISX is enriched on the neo-X (permutation test p<0.0001 in both cases)(see Table S2 for additional details). (**D**) Clustering of the canonical *D. miranda* and *D. melanogaster* MRE motifs inferred from the ancient X chromosome along with the consensus from each helitron TE. ISXRcons refers to the ISXR MRE that is conserved with ISX while ISXRuniq refers to the MRE that is unique to ISXR [see (*20*) for

details]. **(E)** MSL3 ChIP-seq data show that XR CES overlapping ISXR elements have a higher affinity for the MSL complex *in vivo* compared to the neo-X CES that overlap ISX elements (comparing the distribution of MSL-enrichment from 21 strict CES created by ISX, vs. 47 strict CES created by ISXR, Wilcoxon test *P*=0.01).
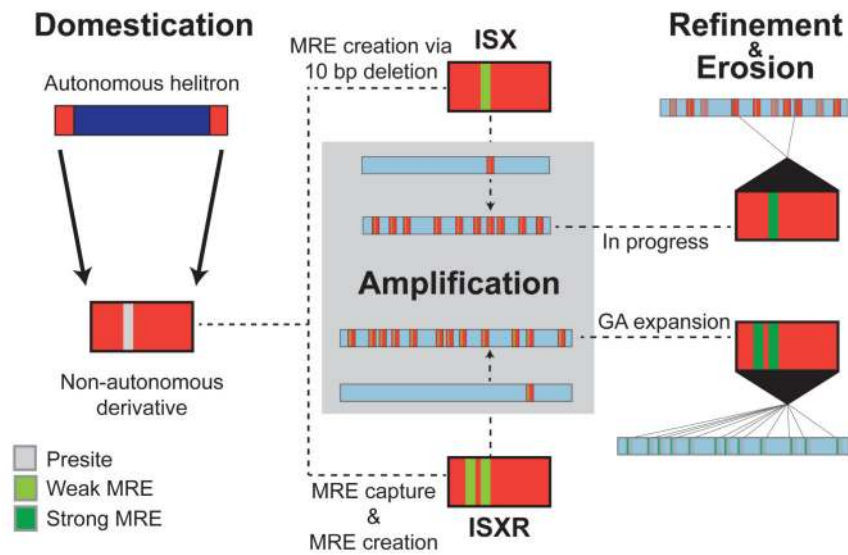
Fig. 4. TE-mediated evolution of MSL complex chromatin entry sites (CES)
Comparison of the two evolutionary timepoints of acquiring chromatin entry sites on XR and the neo-X suggest a three-step model for the TE-mediated wiring of a newly evolved X chromosome into the dosage compensation network followed by erosion of non-functional elements of the TE. The first step, domestication, involves the acquisition of a MRE sequence motif capable of acting as a CES for the MSL complex. The domesticated TE is amplified across the genome and beneficial on a newly formed X chromosome but selected against on autosomal locations. This results in the accumulation of the domesticated TE, along with the MRE motif that it carries, on the X. The amplified MRE motif may initially be suboptimal, as seen with the younger ISX elements on the neo-X, but over time, secondary fine-tuning mutations within each MRE can refine the ability to recruit optimal levels of MSL complex, as seem to have occurred with the older ISXR elements on XR. This is accompanied by erosion of TE sequences that are not required for MSL-binding, eventually degrading the signature of TE involvement for supplying CES.